

Vincent Manet

Méthode des éléments finis

*Vulgarisation des aspects mathématiques
et illustration de la méthode*

Vincent Manet — 2015 (Ceci est la version « livre » de ce document)

Ce document est sous licence Creative Commons 3.0 France :

- paternité ;
- pas d'utilisation commerciale ;
- partage des conditions initiales à l'identique ;

<http://creativecommons.org/licenses/by-nc-sa/3.0/deed.fr>



Introduction

Dans ce (de moins en moins court) document, plutôt à destination d'ingénieurs mécaniciens connaissant déjà la méthode des éléments finis, nous allons essayer de faire une présentation un peu plus théorique que ce qui leur est généralement proposé (et qui est quand même souvent de type « preuve par les mains », ce qui occulte trop de points).

Nous ne ferons appel qu'à des notions mathématiques de bases généralement déjà vues pour la plupart en taupe (ou en tout début de cycle d'ingénieur)... bien que des compléments que l'on peut qualifier d'élémentaires nous aient été demandés et aient été inclus.

Nous espérons, grâce à cette présentation théorique montrer toute la souplesse et la puissance de la méthode, afin de permettre au lecteur d'envisager d'autres simulations que celles qu'il a pu déjà réaliser par le passé.

Pourquoi un ingénieur pratiquant déjà les éléments finis devrait-il s'intéresser plus en profondeur aux mathématiques derrière la méthode ?

Tout d'abord parce que c'est beau et intéressant : deux raisons parfaitement licites et suffisantes. Mais surtout parce que le monde change ! On souhaite des modélisations toujours plus proches du réel, toujours plus détaillées, toujours plus complexes, toujours plus couplées... Par ailleurs un constat s'impose : si la physique d'hier était essentiellement celle du continu, celle d'aujourd'hui est le règne du discontinu. Ainsi, connaître l'intégrale de Riemann et savoir intégrer par parties étaient autrefois des outils suffisants, alors qu'aujourd'hui il faut en passer par les dérivées au sens des distributions, les espaces de Sobolev, l'intégrale de Lebesgue...

Rester sur les outils d'hier c'est se condamner à résoudre les problèmes d'hier ! C'est pourquoi ce document a vu le jour, pour essayer de présenter et d'expliquer « simplement » (nous l'espérons) les mathématiques derrière la méthode, mais sans demander au lecteur de se transformer en mathématicien.

But du document

Le but initial était de *présenter brièvement la théorie mathématique* derrière les éléments finis afin que les ingénieurs utilisant cette méthode puisse en envisager toutes les applications, ainsi que de *couvrir les aspects qui, selon nous, devraient être connus de tout ingénieur mécanicien impliqué ou intéressé par le calcul numérique.*

Toutefois, il s'envisage comme support de référence à plusieurs cours, cours qui ne portent pas sur tous les aspects traités dans ce document, et pendant lesquels les aspects pratiques sont plus développés (avec mise en situation sur machine).

Même si nous avons voulu rester le plus succinct possible, l'introduction de notions de proche en proche a conduit à un document qui fait aujourd'hui une certaine taille (par exemple, nous avons besoins des espaces de Sobolev, mais comment les introduire sans parler des espaces de Lebesgue, mais comment les introduire sans parler...).

Aussi le document a-t-il finalement été découpé en plusieurs parties : un survol des notions mathématiques, puis le traitement du problème continu constituent l'ossature théorique nécessaire à assoir la méthode des éléments finis sur un socle solide. La discrétisation par éléments finis à proprement parler n'est aborder qu'ensuite, et d'ailleurs un seul chapitre suffirait à en faire le

tour... sauf à entrer plus dans le détail concernant « ce qui fâche » : homogénéisation, non linéarité, dynamique, ce qui est fait dans des chapitres séparés.

Enfin, d'autres méthodes sont abordées car également très employées aujourd'hui. Aussi est-il indispensable selon nous d'en avoir entendu parlé et d'en connaître les principales notions (BEM, FEEC...).

En annexes, se trouve un petit fourre-tout comprenant des choses censées être maîtrisées depuis la taupe (mais qui parfois nous sont demandées) et les compléments qui alourdiraient encore les propos précédents.

Certaines notions (essentiellement de topologie) ne sont pas présentées dans ce document. Il nous a semblé que le lecteur devait avoir quelques souvenirs de ce qu'est un ouvert, un fermé, l'adhérence, la densité... Par ailleurs, leur nom peut être suffisamment évocateur pour se passer d'une définition formelle dans le contexte de ce document.

Attention, ce document n'est pas un document de mathématiques, il ne contient d'ailleurs aucune preuve. C'est, dans ces deux premières parties, un document de vulgarisation de notions mathématiques nécessaires à une bonne compréhension de la méthode des éléments finis.

Nous avons voulu réaliser un survol des notions importantes, mais malgré tout, afin de ne pas être parfois trop laconique, nous avons un peu débordé.

En fin de document, un petit index des noms propres permettra au lecteur de replacer les divers développements mentionnés dans l'histoire... Il se peut qu'il subsistent quelques erreurs, notamment au niveau des nationalités mentionnées, car il n'est pas toujours aisé de déterminer rapidement cette information (et nous ne connaissons pas toutes les biographies des personnes citées).

Ce document a été réalisé très rapidement, et de manière extrêmement hachée. Il comporte forcément encore beaucoup de fautes : merci de m'en faire part.

Démarche de l'ingénieur numéricien

En préambule à ce document, nous tenions à synthétiser la démarche complète de l'ingénieur numéricien :

- Modélisation / mise en équations – Construction du problème continu (système d'équations aux dérivées partielles).
- Analyse mathématique du problème posé – Existence, unicité, propriétés des solutions.
- Conception d'une méthode numérique – Construction d'un problème discrétisé.
- Analyse numérique – Questions de stabilité, convergence, précision.
- Algorithmique – Choix de méthodes de résolution en dimension finie.
- Mise en œuvre sur ordinateur – Programmation.
- Pre et Post Traitement (maillages / visualisation) – Interpolation, extrapolation, outils de la CAO.

Tous ces points ne seront évidemment pas abordés dans ce document !

Remerciements :

Nous remercions Mathias Legrand pour ses conseils avisés et sa relecture pertinente.

Notre collaboration a permis une très nette amélioration de la qualité typographique générale du document, et a conduit à la coexistence de deux versions, issues du même code source : l'une que nous appelons « version cours » correspond à ce que nous proposons en cours (couleurs, notations) ; l'autre que nous nommons « version livre », plus classique et sage dans sa forme, est plus proche d'un ouvrage.

Table des matières

Introduction	3
But du document	3
Démarche de l'ingénieur numéricien	4
Table des matières	5

I	SURVOL MATHÉMATIQUE
1	Les espaces de base 17
1.1	Panorama non exhaustif des espaces 17
1.1.1	Point de vue topologique 18
1.1.2	Point de vue métrique 19
1.1.3	Point de vue algébrique 19
1.2	Tribu, mesure, espaces mesurable et mesuré 22
1.3	Tribu borélienne, mesures de Dirac et Lebesgue 23
1.4	Propriétés de la mesure de Lebesgue 25
1.5	Petit exemple amusant d'injection dans un Hilbert 26
2	Applications et morphismes 27
2.1	Fonction, application, injection, surjection, bijection 28
2.2	Morphismes 29
2.2.1	Présentation 29
2.2.2	Cas des espaces vectoriels : application et forme linéaires 30
2.2.3	Endo, iso, auto-morphismes 30
2.2.4	Espace dual d'un espace vectoriel 30
2.2.5	Noyau et image 31
2.3	Opérateur 32
3	Continuité et dérivabilité 33
3.1	Continuité et classe C^0 33
3.2	Continuité de Hölder et Lipschitz 35
3.3	Dérivée 35
3.4	Fonctions de classe C^k 37
3.5	Différentielle 37
3.6	Dérivées partielles 37
3.7	Retour sur les classes C^k pour une fonction de plusieurs variables 39

3.8	Nabla et comparses	40
3.8.1	Champs de vecteurs et de scalaires	40
3.8.2	Gradient, divergence, rotationnel, Laplacien et D'Alembertien	40
3.8.3	Normale et dérivée normale	41
3.8.4	Potentiel d'un champ vectoriel	42
3.8.5	Signification « physique »	42
3.9	Quelques théorèmes sur les intégrales	43
4	Espaces de Lebesgue	45
4.1	Présentation des espaces de Lebesgue L^p	48
4.2	Construction de L^p	49
4.3	Espace L^0	50
4.4	Espace L^∞ et dualité avec L^1	50
4.5	Espace L^2	51
4.6	Compléments et retour sur les fonctions continues et différentiables	51
5	Espaces de Sobolev	55
5.1	Distributions	56
5.2	Dérivées au sens des distributions	58
5.3	Espaces $W^{m,p}(\Omega)$	60
5.4	Espaces $H^m(\Omega)$, $H_0^m(\Omega)$ et $H^{-m}(\Omega)$	60
5.5	Prise en compte du contour du domaine	62
5.5.1	Trace	62
5.5.2	Espace trace	63
5.6	Espaces $H^1(\Omega)$, $H_0^1(\Omega)$ et $H^{-1}(\Omega)$	64
5.7	Espaces $H(\text{div})$ et $H(\text{rot})$	66
5.8	Inégalités utiles	66

II

PROBLÈME CONTINU

6	Problèmes physiques : équations différentielles et aux dérivées partielles	75
6.1	Introduction	75
6.2	Conditions aux limites	77
6.2.1	Dirichlet – valeurs aux bords	77
6.2.2	Neumann – gradients aux bords	77
6.2.3	Robin – relation gradient/valeurs sur le bord	78
6.2.4	Condition aux limites dynamique	78
6.2.5	Condition aux limites mêlée	78
6.3	Types d'équation aux dérivées partielles	78
6.4	Phénomènes de propagation et de diffusion	78
6.4.1	Équations de Laplace et Poisson	79
6.4.2	Équation d'onde, phénomènes vibratoires	80
6.4.3	Équation de la chaleur	80

6.5	Mécanique des fluides	83
6.5.1	Équation de Navier-Stokes	83
6.5.2	Équation de Stokes	85
6.5.3	Équation d'Euler	86
6.6	Équations de la mécanique des milieux continus des solides	86
6.6.1	Notions générales conduisant aux équations de la mécanique	86
6.6.2	Formulation générale	89
6.6.3	Dynamique / statique	89
6.6.4	Grands / petits déplacements	90
6.6.5	Loi de comportement	90
6.7	Équations de l'acoustique	92
6.8	Multiplicateurs de Lagrange	92
7	Formulations faible et variationnelle	95
7.1	Principe des formulations faible et variationnelle	95
7.2	Théorème de représentation de Riesz-Fréchet	98
7.2.1	Cas des formes linéaires	98
7.2.2	Extension aux formes bilinéaires	98
7.3	Théorème de Lax-Milgram	98
7.4	Théorème de Babuška et condition inf-sup	100
7.5	Théorèmes de Brezzi et condition BBL	100
7.6	Multiplicateurs de Lagrange	102
8	Problèmes physiques : formulations faibles et variationnelles	105
8.1	Phénomènes de propagation et de diffusion	105
8.1.1	Équations de Laplace et Poisson	105
8.1.2	Équation d'onde	108
8.1.3	Équation de la chaleur	108
8.2	Mécanique des fluides	109
8.2.1	Équation de Stokes	109
8.2.2	Équation de Navier-Stokes	110
8.2.3	Équation d'Euler	111
8.3	Équations de la mécanique des milieux continus des solides	111
8.3.1	Formulation générale	111
8.3.2	Choix des variables	112
8.3.3	Équation des plaques	114
8.4	Équations de l'acoustique	116
8.4.1	Équation de Helmholtz	117
8.4.2	Conditions aux limites en acoustique	117
8.4.3	Formulation faible	117
9	Exemple de formulation variationnelle d'un problème de Neumann	119
9.1	Étude directe de l'existence et l'unicité de la solution	119
9.2	Formulation variationnelle	119
9.3	Formulation mixte duale	121

10	La méthode des éléments finis	129
10.1	Introduction	129
10.2	Problèmes de la modélisation « réelle »	131
10.2.1	Problèmes géométriques	131
10.2.2	Problèmes d'échelle	131
10.2.3	Couplage géométrique	132
10.2.4	Couplage intrinsèque	133
10.3	Principe de la méthode : résolution d'un système matriciel	133
10.4	Approximation conforme et lemme de C��a	135
10.4.1	Cas Lax-Milgram	135
10.4.2	Cas Babu��ska	136
10.4.3	Cas Brezzi	137
10.5	Approximations non conformes et lemmes de Strang	137
10.5.1	Approximation interne	137
10.5.2	Approximation externe	138
10.6	Convergence de la m��thode des ��l��ments finis en approximation conforme interne	138
10.6.1	Calcul de la majoration d'erreur	139
10.6.2	Majoration de l'erreur	140
11	Choix d'un Mod��le	141
11.1	La m��canique, un probl��me �� plusieurs champs	142
11.2	Plusieurs mod��lisations d'un m��me probl��me	145
11.3	Exemple : retour sur le calcul de poutre du paragraphe 11.1 avec CAST3M	149
11.3.1	Mod��lisation 2D	149
11.3.2	Mod��le 3D	150
11.4	Interpolation des champs et de la g��om��trie	151
12	Formulation pratique d'��l��ments finis	153
12.1	��l��ments de Lagrange	153
12.1.1	Unisolvance	153
12.1.2	��l��ments finis de Lagrange	154
12.1.3	Famille affine d'��l��ments finis et ��l��ment de r��f��rence	156
12.1.4	Construction de la base globale	157
12.1.5	Exemple : ��l��ments finis \mathbb{P}_k en ��lasticit�� lin��aire	157
12.2	��l��ments d'Hermite	158
12.2.1	Classe d'un ��l��ment fini	158
12.2.2	��l��ments finis d'Hermite	159
12.2.3	��l��ments uni- et bidimensionnels	159
12.2.4	Exemple : ��l��ments finis d'Hermite pour l'��quation des plaques	160
12.3	Traitement de plusieurs champs	163
12.4	Validation pratique et indicateurs d'erreur	164
12.4.1	Modes rigides et parasites	164
12.4.2	Modes associ��s aux d��formations constantes	165
12.4.3	Patch-tests	165
12.4.4	Test de pr��cision d'un ��l��ment	165

12.5	Exemple : quelques variations sur le thème des éléments unidimensionnels	166
12.5.1	Élément de référence unidimensionnel linéaire à deux nœuds	166
12.5.2	Rappels sur la jacobienne et le jacobien d'une transformation	167
12.5.3	Éléments de référence unidimensionnels linéaires à n nœuds	168
12.5.4	Élément de référence unidimensionnel infini	169
12.5.5	Élément fini de barre unidimensionnel	169
12.5.6	Assemblage de trois éléments unidimensionnels linéaires à deux nœuds	170
12.5.7	Élément de barre unidimensionnel de type Hermite (élément subparamétrique)	171
12.5.8	Élément mixte unidimensionnel	172
12.6	Sur les déplacements imposés	173
12.6.1	Problème considéré	173
12.6.2	Retour sur la résolution de systèmes linéaires	173
12.6.3	Complément de Schur et déplacements imposés	174
12.6.4	Multiplicateurs de Lagrange et déplacements imposés	174
12.6.5	Actions extérieures et déplacements imposés	175
12.6.6	Retour sur notre exemple	175
12.6.7	Relations linéaires entre degrés de liberté	176
13	Calcul efficient : qualité des résultats et efforts de calcul	177
13.1	Amélioration d'un modèle : méthodes r, h et p	177
13.2	Post-traitement	177
13.3	Exemple d'implémentation d'un post-traitement dans ANSYS	178
13.3.1	Macro dans ANSYS	178
13.3.2	Poutre en U	181
13.3.3	Résultats	183
13.4	Sous-structuration et simulation multi-échelles	190
13.4.1	Condition de périodicité – méthodes multi-niveaux	192
13.4.2	Couplage des cellules microscopiques – méthodes de décomposition de domaine	193
13.5	Super-éléments	194
13.5.1	Condensation statique	195
13.5.2	Remonter aux degrés de liberté internes	196
13.6	Pseudo-inversion et réanalyse	196
13.6.1	Modification du chargement uniquement	196
13.6.2	Modification de la matrice	196
13.6.3	Modification des conditions cinématiques	197
13.6.4	Deux mots de statistiques	198
13.7	Dérivées d'ordre supérieur	198
13.7.1	Dérivées par rapport à la géométrie	199
13.7.2	Calcul des dérivées	200
14	Le maillage	201
14.1	Maillage de Delaunay	202
14.1.1	Maillage simplicial	202
14.1.2	Maillage de Delaunay-Voronoi	203
14.1.3	Remarques	205
14.2	Maillage par avancement de fronts	206
14.3	Maillage par transformation	206
14.4	Remarques sur le maillage quadrangulaire et hexaédrique	208

15	Homogénéisation	209
15.1	Méthodes d'homogénéisation	210
15.1.1	Méthode de développement régulier	212
15.1.2	Méthode de la couche limite	212
15.1.3	Méthode de développement asymptotique infini	213
15.1.4	Cas des coefficients discontinus	213
15.2	Homogénéisation simplifiée pour les matériaux composites	213
15.2.1	Introduction	213
15.2.2	Loi des mélanges, bornes de Voigt et de Reuss	214
15.3	Homogénéisation des matériaux poreux	215
15.4	Homogénéisation des problèmes non stationnaires	216
15.5	Changement de dimension et raccord de maillage	216
16	Optimisation	217
16.1	Théorie de l'optimisation	219
16.1.1	Existence et unicité d'un minimum	219
16.1.2	Différentiabilité et optimalité	220
16.1.3	Lagrangien	221
16.1.4	Dualité et point selle	223
16.1.5	Maximisation	224
16.1.6	Ensemble admissible	224
16.1.7	L'optimisation discrète	225
16.1.8	Remarques sur la mise en œuvre	225
16.2	Optimisation de forme	226
16.2.1	Optimisation paramétrique	226
16.2.2	Optimisation géométrique	226
16.2.3	Optimisation topologique par homogénéisation	228
16.3	Exemples d'optimisation d'une membrane	230
16.3.1	Optimisation du matériau élastique anisotrope en rigidité	230
16.3.2	Optimisation du matériau élastique anisotrope pour un déplacement cible	231
16.3.3	Optimisation de l'épaisseur en rigidité	232
16.3.4	Optimisation de l'épaisseur pour un déplacement cible	233
16.3.5	Optimisation de l'épaisseur pour une contrainte cible	234
16.3.6	Optimisation géométrique en rigidité sous tension prescrite	235
16.3.7	Optimisation de l'épaisseur pour la première fréquence propre	236
17	Problèmes non stationnaires	239
17.1	Équation non stationnaire de la dynamique	239
17.2	Schéma explicite : différences finies centrées	240
17.3	Schéma implicite : schéma de Newmark classique	240
17.4	Comparaison des méthodes explicite et implicite	241
17.5	Exemple : un calcul de propagation avec FREEFEM++	242
17.6	Décomposition modale	244
18	Les ondes	247
18.1	Introduction	247
18.2	Notions de valeur, vecteur, mode et fréquence propres	249
18.3	Vibration des structures	250
18.3.1	Vibrations libres non amorties	250
18.3.2	Vibrations libres amorties	252

18.3.3	Vibrations périodiques forcées	253
18.3.4	Régimes transitoires	254
18.3.5	Calcul des modes propres et méthodes de réduction modale	256
18.4	Remarques sur l'amortissement	257
18.5	Pour aller plus loin : cas des chocs large bande	259
18.5.1	Approches temporelles	260
18.5.2	Approches fréquentielles	261
18.5.3	Remarques	264
19	L'acoustique	265
19.1	Introduction à l'acoustique physique	266
19.1.1	Émission	266
19.1.2	Transmission	266
19.1.3	Réception	272
19.2	Calculs acoustiques par éléments finis	278
19.2.1	Modèles simplifiés pour valeurs de référence	278
19.2.2	Constitution d'un modèle éléments finis	280
19.2.3	Convergence	284
19.2.4	Vers l'infini...	284
19.2.5	... et au-delà	285
19.2.6	Post-Traitement	285
19.3	Quelques illustrations avec FREEFEM++	286
19.3.1	Un exemple en acoustique des salles	286
19.3.2	Un silencieux automobile	286
19.3.3	Deux mots de statistiques	287
19.3.4	Sur les conditions aux limites	288
19.3.5	Un peu d'amortissement	290
19.3.6	Un obstacle	291
19.3.7	Transmission entre deux milieux	292
19.3.8	Interaction fluide-structure semi-forte	293
20	Les non linéarités	297
20.1	Tenseurs, décomposition des tenseurs	297
20.1.1	Tenseur des contraintes	298
20.1.2	Tenseur des déformations	301
20.2	Non linéarité géométrique	302
20.3	Non linéarité matérielle	303
20.3.1	Modèles rhéologiques	303
20.3.2	Visoélasticité	305
20.3.3	Visoplasticité	305
20.3.4	Plasticité	306
20.3.5	Les élastomères	309
20.3.6	Les composites, l'anisotropie	311
20.4	Le contact	313
20.4.1	Lois de contact et de frottement	315
20.4.2	Algorithme local	316
20.4.3	Algorithme global	316
20.5	Exemple : une toute première approche du contact avec CAST3M	317
20.5.1	Contact sur une surface infiniment rigide	317
20.5.2	Contact entre deux solides	318
20.5.3	Résolution pas à pas	320

21	La rupture en mécanique	323
21.1	Approches globale et locale	323
21.2	Mécanique linéaire de la rupture	325
21.2.1	Concentrations de contraintes des défauts	325
21.2.2	Équilibre énergétique	325
21.2.3	Taux d'énergie libre G	327
21.2.4	Facteur d'intensité de contrainte K	328
21.2.5	Intégrale J	329
21.3	Mécanique élastoplastique de la rupture	331
21.3.1	Détermination de la zone plastique	331
21.3.2	Modèle d'Irwin	331
21.3.3	Autres modèles	332
21.4	Modélisation numérique de la rupture	332
21.4.1	Par la méthode des éléments finis	332
21.4.2	Par les méthodes sans maillage	334
21.4.3	Par les éléments étendus	334
21.5	Fatigue et durée de vie	335
21.5.1	Courbe et limite de fatigue	335
21.5.2	Cumul des dommages : principes de Miner	335
21.5.3	Propagation : loi de Paris	336
21.5.4	Prédiction de la durée de vie	337
21.5.5	Sur la fatigue des composites	338
22	Stochastique : prise en compte de la variabilité	339
22.1	Introduction	339
22.2	Représentation des processus stochastiques	341
22.2.1	Variable aléatoire	341
22.2.2	Espace de probabilité	341
22.2.3	Processus ou champ aléatoire	342
22.2.4	Discretisation de champs aléatoires	343
22.2.5	Développement en série de Karhunen-Loève	344
22.2.6	Chaos polynomial	345
22.3	Éléments finis stochastiques	346
22.3.1	Développement en séries de von Neumann	347
22.3.2	Un exemple en statique linéaire	347
22.3.3	Projection sur la base du chaos polynomial	348
22.3.4	Obtention des coefficients du système matriciel à résoudre	348
22.3.5	Synthèse et remarques	349
22.4	Post-traitement : moments statistiques	351
22.5	Fiabilité des structures	351
22.5.1	Définitions	351
22.5.2	Méthodes FORM/SORM	354
22.5.3	Tirages d'importance	355
23	Quelques méthodes dérivées	357
23.1	Méthode des éléments frontières	357
23.2	Méthodes particulières ou sans maillage	358
23.3	Partition de l'unité	359
23.4	Méthode des éléments finis étendue	359
23.5	Méthodes de treillis de Boltzmann	360

23.6	FEEC	361
23.7	Systèmes multi-corps	361
24	Quelques mots sur les singularités	363
24.1	Qu'est-ce qu'une singularité ?	363
24.2	Singularités et éléments finis	363
24.3	Quand les singularités se produisent-elles ?	364
24.4	Comment éviter les singularités	364
24.5	Singularités et pertinence d'un résultat	365
24.6	Conclusion	365
	Sur la fiabilité des résultats	367
	Quelques perspectives	368

IV

ANNEXES

A	Interpolation et approximation	371
A.1	Quelques bases polynomiales	373
A.1.1	Motivation	373
A.1.2	Orthogonalité	373
A.1.3	Base naturelle	374
A.1.4	Polynômes de Lagrange	374
A.1.5	Polynômes d'Hermite	374
A.1.6	Polynômes de Legendre	375
A.1.7	Polynômes de Tchebychev	376
A.1.8	Polynômes de Laguerre	378
A.1.9	Polynômes de Bernstein	379
A.2	Interpolation polynomiale	380
A.2.1	Interpolation de Lagrange	380
A.2.2	Interpolation par Spline	380
A.2.3	Interpolation d'Hermite	380
A.3	Méthodes d'approximation	381
A.3.1	Courbe de Bézier	381
A.3.2	B-Spline	382
A.3.3	B-splines rationnelles non uniformes	383
B	Intégration numérique	385
B.1	Méthodes de Newton-Cotes	385
B.2	Méthodes de quadrature de Gauß	387
C	Résolution des équations différentielles ordinaires	391
C.1	Résolution exacte des équations différentielles linéaires	391
C.1.1	Équation différentielle linéaire scalaire d'ordre 1	391
C.1.2	Équation différentielle du premier ordre à variables séparées	393
C.1.3	Équation différentielle linéaire d'ordre deux	394
C.2	Résolution numérique	396
C.2.1	Méthode d'Euler, Runge-Kutta d'ordre 1	396
C.2.2	Méthode de Runge-Kutta d'ordre 2	396
C.2.3	Méthode de Runge-Kutta d'ordre 4	397
C.2.4	Méthode de Crank-Nicolson	397

C.2.5	Méthode de Newmark	398
D	Méthode de Newton Raphson	399
D.1	Présentation	399
D.2	Algorithme	400
E	Schémas numériques	403
E.1	Les propriétés d'un schéma numérique	403
E.1.1	Problème bien posé	403
E.1.2	Conditionnement	404
E.1.3	Stabilité ou robustesse	404
E.1.4	Consistance	404
E.1.5	Convergence	405
E.2	Théorèmes de convergence	405
F	Index des noms propres	407
G	Index des concepts	413
H	Références bibliographiques	417

Chapitre 1

Les espaces de base

Résumé — Dans les problèmes que nous envisageons de traiter *in fine*, il n'est finalement besoin que d'espaces vectoriels normés finis sur le corps des réels. On pourrait rapidement en donner les principales propriétés qui sont sans doute encore en mémoire des lecteurs tant elles sont « naturelles ».

Mais en faisant cela, nous ne respecterions pas nos engagements de présenter un peu plus avant les fondements mathématiques.

Sans toutefois aller trop loin dans les notions topologiques générales, nous allons essayer dans ce chapitre de passer en revue la liste des espaces depuis les espaces topologiques jusqu'aux espaces de Hilbert, de manière succincte et didactique (nous l'espérons) en relevant les points importants essentiellement du point de vue de leur utilité.

1.1 Panorama non exhaustif des espaces

Histoire

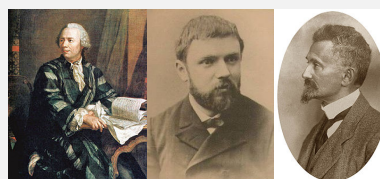
Le mot « topologie » vient de la contraction des noms grecs « topos » (lieu) et « logos » (étude), c'est donc l'étude du lieu. On a d'ailleurs commencé par l'appeler *Analysis Situs*, le terme « topologie » n'étant introduit qu'en 1847, en allemand, par Johann Benedict Listing dans *Vorstudien zur Topologie*.

La topologie vise à définir ce qu'est un lieu (i.e. un espace) et quelles peuvent être ses propriétés (je dirais uniquement en tant que tel, sans autre ajout). Elle s'intéresse plus précisément à ce que l'on appelle aujourd'hui espaces topologiques et aux applications, dites continues, qui les lient, ainsi qu'à leurs déformations (*"A topologist is one who doesn't know the difference between a doughnut and a coffee cup"*).

En analyse, elle fournit des informations sur l'espace considéré permettant d'obtenir un certain nombre de résultats (existence et/ou unicité de solutions d'équations au dérivées partielles, notamment). Les espaces métriques ainsi que les espaces vectoriels normés sont des exemples d'espaces topologiques. L'origine de la topologie est l'étude de la géométrie dans les cultures antiques. Le travail de Leonhard Euler datant de 1736 sur le problème des sept ponts de Königsberg est considéré comme l'un des premiers résultats de géométrie qui ne dépend d'aucune mesure, i.e. l'un des premiers résultats topologiques.

Henri Poincaré publia *Analysis Situs* en 1895, introduisant les concepts d'homotopie et d'homologie. Bien d'autres mathématiciens ont contribué au sujet parmi lesquels nous citerons : Cantor, Volterra, Arzelà, Hadamard, Ascoli, Fréchet, Hausdorff...

Finalement, une dernière généralisation en 1922, par Kuratowski, donna le concept actuel d'espace topologique.



Euler

Poincaré

Hausdorff

1.1.1 Point de vue topologique

E, un ensemble

Cela suffit déjà pour pouvoir s'intéresser par exemple à des fonctions, à des relations d'équivalence (donc au quotient)... décomposition canonique, permutation...

On peut ensuite définir un ensemble ordonné...

Une topologie T est un ensemble de parties de E que l'on définit comme les ouverts de (E, T) , vérifiant les propriétés suivantes :

- L'ensemble vide et E appartiennent à T .
- Toute réunion quelconque d'ouverts est un ouvert, i.e. si $(O_i)_{i \in I}$ est une famille d'éléments de T , indexée par un ensemble I quelconque (pas nécessairement fini ni même dénombrable) alors $\bigcup_{i \in I} O_i \in T$.
- Toute intersection finie d'ouverts est un ouvert, i.e. si O_1, \dots, O_n sont des éléments de T (avec $n > 0$), alors $O_1 \cap \dots \cap O_n \in T$.

Espace topologique

À partir des ouverts, on définit les fermés, l'adhérence, l'intérieur, l'extérieur, voisinage...

Séparation :

deux points distincts quelconques admettent toujours des voisinages disjoints.

Espace séparé (ou de Hausdorff)

Intérêts :

- *Unicité de la limite de tout filtre convergent.*
- *Une suite convergente a une limite unique.*
- *Une topologie plus fine qu'une topologie séparée est séparée.*
- *Tout sous-espace d'un espace séparé est séparé.*
- *Deux applications continues à valeurs dans un séparé qui coïncident sur une partie dense sont égales.*

Régularité :

Il est possible de séparer un point x et un fermé F ne contenant pas x par deux ouverts disjoints.

On peut même alors choisir ces deux ouverts de manière à ce que leurs adhérences respectives soient disjointes.

Espace régulier

Application :

- *Tout point admet une base de voisinages fermés.*
- *Tout fermé est l'intersection de ses voisinages fermés.*

Compacité (recouvrement fini) :

Un espace séparé est compact (vérifie la propriété de Borel-Lebesgue), si chaque fois qu'il est recouvert par des ouverts, il est recouvert par un nombre fini d'entre eux.

Espace compact

L'intérêt des compacts est de pouvoir étendre des propriétés trivialement vérifiées par des applications définies sur un ensemble fini à des applications définies sur des espaces topologiques infinis, à condition bien sûr qu'elles soient continues.

Tout produit de compacts est compact.

\mathbb{R} est compact

Un espace peut ne pas être compact, mais une de ses parties l'être :

- \mathbb{R} n'est pas fermé, mais tout $[a; b]$ fermé borné l'est.
- \mathbb{R}^n n'est pas compact, mais tout pavé fermé l'est (voir théorème 6).

Notons bien que rien n'est dit sur les éléments de l'ensemble E . Ce peuvent être des éléments discrets, des scalaires, des vecteurs, des fonctions...

1.1.2 Point de vue métrique

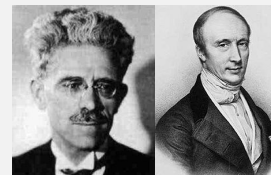
Comme nous le mentionnions au paragraphe précédent, lorsque l'on parle d'espace on a intuitivement envie de parler de « distance ».

Histoire

Unifiant les travaux de ses prédécesseurs sur les espaces de fonctions, c'est en 1906 que Maurice Fréchet introduit le concept d'espace métrique.

La métrique qui nous est la plus usuelle est évidemment la métrique euclidienne, qui est celle que nous utilisons en géométrie « classique » (euclidienne) : la distance entre deux points est égale à la longueur du segment les reliant. La structure métrique fournit beaucoup plus d'information sur la forme géométrique des objets que la structure topologique.

Enfin nous redonnerons le si important critère de Cauchy (qui est valable pour tout espace uniforme, dont notamment les espaces métriques) qui permet de définir la toute aussi importante notion de complétude.



Fréchet

Cauchy

E, un ensemble

Une distance ou métrique d est une application de $E \times E \rightarrow \mathbb{R}_+$ telle que :

- $d(x, y) = d(y, x)$ (symétrie) ;
- $d(x, y) > 0$ si $x \neq y$, et $d(x, x) = 0$ (positivité) ;
- $d(x, z) \leq d(x, y) + d(y, z)$ (inégalité triangulaire).

Espace métrique

On peut réexprimer les notions d'ouvert, fermé, adhérence... densité, continuité... avec la métrique (les ε ...).

Deux normes sont équivalentes si elles définissent la même topologie.

Deux normes $\|\cdot\|_1$ et $\|\cdot\|_2$ sont équivalentes s'il existe deux constantes strictement positives k' et k'' telles que $\forall x \in E, \|x\|_1 \leq k'\|x\|_2$ et $\|x\|_2 \leq k''\|x\|_1$.

Critère de Cauchy :

Soit E un espace métrique et soit $x_0, x_1, \dots, x_n, \dots$ une suite d'éléments de E . Cette suite est de Cauchy de E si :

$$\forall \varepsilon > 0, \exists N \in \mathbb{N}, (\forall n \in \mathbb{N}, \forall m \in \mathbb{N}, n \geq N, m \geq N) : d(x_m, x_n) \leq \varepsilon \quad (1.1)$$

ou encore : $d(x_m, x_n)$ tend vers 0 quand m et n tendent vers l'infini.

Espace métrique complet

Toute suite convergente est de Cauchy.

Réciproque : si $x_0, x_1, \dots, x_n, \dots$ est de Cauchy sur \mathbb{R} ou \mathbb{C} , alors elle converge.

Cette propriété est fondamentale car elle permet de reconnaître si une suite converge sans connaître sa limite.

Attention, la notion d'espace complet est une notion métrique et non topologique (elle peut donc être vraie pour une métrique et fausse pour une autre).

L'importance de la complétude tient à ce que la solution d'un problème passe souvent par une solution approchée. Si la suite des solutions approchées est de Cauchy d'un espace convenable, et si cet espace est complet, alors la convergence vers une solution du problème est assurée.

1.1.3 Point de vue algébrique

Jusqu'à présent, nous n'avons pas vraiment parlé d'opérations que nous pourrions effectuer à l'intérieur des espaces que nous avons définis, ou entre ces espaces. Pour une présentation des structures algébriques on se reportera au cours homonyme. Elles sont riches et nombreuses. Dans le cadre de ce document, nous ne nous intéresserons qu'au cas de la structure d'espace vectoriel, le but étant, comme mentionné en introduction, d'en arriver aux espaces de Hilbert, fondements de l'analyse fonctionnelle.

Lorsque l'on demande de citer l'un des grands mathématiciens du XX^e siècle, Henri Poincaré et David Hilbert se partagent souvent la première place, aussi bien pour l'éventail considérable des sujets qu'ils ont abordés que pour avoir fait émerger de nombreuses idées fondamentales.

Hilbert reste célèbre pour ses 23 problèmes (dits problèmes de Hilbert) qu'il présenta au deuxième congrès international des mathématiciens à Paris en 1900, qui tenaient jusqu'alors les mathématiciens en échec et devaient marquer le cours des mathématiques du XX^e siècle (et il avait raison ; tous ne sont pas résolus à ce jour). Notons que c'est von Neumann, reprenant les travaux de Hilbert, qui formalise et nomme ces espaces les espaces de Hilbert en 1927.



Hilbert Banach

Nous croiserons également un autre fondateur de l'analyse fonctionnelle, Stephan Banach qui a généralisé entre autre les travaux de Hilbert sur les équations intégrales, notamment en approfondissant la théorie des espaces vectoriels topologiques.

E, un ensemble

Structure d'espace vectoriel :

C'est une structure comportant une loi de composition interne et une loi de composition externe sur un corps \mathbb{K} permettant d'effectuer des combinaisons linéaires (voir cours sur les structures algébriques).

La loi de composition interne, notée $+$ en fait un groupe abélien, la loi de composition externe est la multiplication par un scalaire, scalaire pris sur le corps \mathbb{K} considéré.

Espace vectoriel (topologique)

Sur un espace vectoriel de dimension finie sur \mathbb{R} ou \mathbb{C} , deux normes quelconques sont équivalentes.

Une norme sur un espace vectoriel E est une fonction, $x \mapsto \|x\|$ possédant les propriétés :

- positivité : $\|x\| > 0$ pour $x \neq 0$, $\|0\| = 0$;
- transformation par les homothéties : $\|\lambda x\| = |\lambda| \|x\|$, $\lambda \in \mathbb{K}$;
- inégalité de convexité : $\|x + y\| \leq \|x\| + \|y\|$.

La distance issue de la norme est la distance définie par $d(x, y) = \|x - y\|$.

Espace vectoriel normé

Tout espace vectoriel normé est automatiquement un espace métrique (avec la distance issue de sa norme). De plus :

- la distance est invariante par translation : $d(x - a, y - a) = d(x, y)$.
- une homothétie de rapport λ multiplie la distance par $|\lambda|$: $d(\lambda x, \lambda y) = |\lambda| d(x, y)$.

Tout espace vectoriel normé de dimension finie est localement compact (i.e. tout point possède au moins un voisinage compact), car toute boule fermée est compacte.

L'intérêt des distances issues d'une norme est qu'elles rendent continues les opérations de l'espace vectoriel et qu'en particulier les translations sont des homéomorphismes.

Comme un espace vectoriel normé muni de la distance issue de sa norme est un espace métrique, on peut se demander si cet espace métrique vérifie le critère de Cauchy (voir paragraphe précédent) afin d'en faire un espace complet.

Espace de Banach

C'est donc finalement un espace vectoriel normé sur un sous-corps \mathbb{K} de \mathbb{C} (en général, $\mathbb{K} = \mathbb{R}$ ou \mathbb{C}), complet pour la distance issue de sa norme.

Comme la topologie induite par sa distance est compatible avec sa structure d'espace vectoriel, c'est un espace vectoriel topologique.

Une norme $\|\cdot\|$ découle d'un produit scalaire ou hermitien $\langle \cdot, \cdot \rangle$ si l'on a : $\|x\| = \sqrt{\langle x, x \rangle}$.

Espace de Hilbert

C'est donc un espace préhilbertien complet, i.e. un espace de Banach dont la norme découle d'un produit scalaire ou hermitien unique.

*Un espace de Hilbert est la généralisation en dimension quelconque d'un espace euclidien ou hermitien (Un **espace vectoriel euclidien** est un espace vectoriel réel de dimension finie muni d'un produit scalaire).*

Le plus important est de disposer d'un produit scalaire, car on va alors pouvoir déterminer des parties orthogonales de l'espace, donc en somme directe (voir théorème 3).

On rappelle, si besoin, qu'un produit scalaire est une forme bilinéaire (i.e. une application de $E \times E$ dans \mathbb{R} , linéaire pour chacune des deux variables) possédant les propriétés :

- symétrie : $\forall x, y \in E, \langle x, y \rangle = \langle y, x \rangle$
- positivité : $\forall x \in E, \langle x, x \rangle \geq 0$
- définie : $\forall x \in E, (\langle x, x \rangle = 0 \Rightarrow x = 0)$

ce qui se traduit par la définition : « Un produit scalaire sur un espace vectoriel réel E est une forme bilinéaire symétrique définie positive ».

Théorème 1 Tout sous-espace vectoriel fermé d'un espace de Hilbert est un espace de Hilbert.

Voici quelques espaces de Banach couramment utilisés :

- Les espaces euclidiens \mathbb{R}^n et les espaces hermitiens \mathbb{C}^n munis de la norme :

$$\|(x_1, \dots, x_n)\| = \sqrt{\sum_{i=1}^n x_i \overline{x_i}} \quad (1.2)$$

où $\overline{x_i}$ désigne le conjugué de x_i .

- L'espace des fonctions (dans \mathbb{R} ou \mathbb{C}) continues et bornées sur un espace topologique X , muni de la norme $\|f\| = \sup_{x \in X} (|f(x)|)$. En particulier, l'espace des fonctions continues sur un espace X compact, comme un intervalle réel $[a; b]$.
- Pour tout réel $p \geq 1$, l'espace L^p des classes de fonctions mesurables, dans \mathbb{R} ou \mathbb{C} , sur un espace mesuré X .

Voici quelques espaces de Hilbert classiques :

- L'espace euclidien \mathbb{R}^n muni du produit scalaire usuel est hilbertien,
- L'espace L^2 des fonctions de carré intégrable (voir chapitre 4).
- Certains espaces de Sobolev (voir chapitre 5) sont des espaces de Hilbert (ceux qui nous intéresseront en général, ça tombe bien).

Théorème 2 — Théorème de complétion sur un espace de Banach. Soit E un espace vectoriel normé incomplet. Alors E est isomorphe, en tant qu'espace vectoriel, et isométrique à un sous-espace vectoriel dense dans un espace de Banach \overline{E} .

Cet espace de Banach \overline{E} est unique à un isomorphisme isométrique près.

Ce théorème de complétion répond par l'affirmative à la question : si l'espace normé E n'est pas complet, existe-t-il un espace de Banach minimal \overline{E} le contenant ?

Théorème 3 — Théorème de projection dans un espace de Hilbert. Soit F un sous-espace vectoriel fermé d'un espace de Hilbert H (il est alors lui-même un espace de Hilbert d'après le théorème 1). Alors :

- son orthogonal F^\perp , qui est l'ensemble des vecteurs orthogonaux à chaque vecteur de F , est un sous-espace vectoriel fermé (donc est aussi un espace de Hilbert, toujours en vertu du théorème 1) ;
- et est également le supplémentaire de F : tout $h \in H$ s'écrit de façon unique $h = h' + h''$ où h' est le projeté orthogonal de h sur F , et h'' le projeté orthogonal de h sur F^\perp .

Ainsi, F^\perp est l'orthogonal et le supplémentaire de F , et ces deux sous-espaces sont donc en somme directe. On a donc $H = F \oplus F^\perp$ (on utilise parfois le terme de « somme hilbertienne »).

Ce résultat subsiste si l'on suppose seulement que H est un espace préhilbertien et que F est un sous-espace vectoriel complet de H .

1.2 Tribu, mesure, espaces mesurable et mesuré

En complément à l'aspect métrique (Fréchet 1906), nous allons maintenant aborder l'aspect mesure.

Le but, à travers la notion de mesure, est d'étendre la notion usuelle de longueur pour les ensembles de \mathbb{R} , ou de volume pour ceux de \mathbb{R}^n , et ceci de deux manières :

- on veut d'une part pouvoir considérer des espaces de base plus généraux (plus « abstraits » : espaces de dimension infinie, espaces sur lesquels on définit les probabilités...);
- et d'autre part on veut englober dans le même cadre mathématique les notions de longueurs, surface, volume, mais aussi de masses ou charges ponctuelles issues de la mécanique, de l'électricité... car toutes ces quantités possèdent une même propriété évidente : l'additivité (i.e. si l'on désigne par $\mu(A)$ le volume, la masse, la charge électrique... d'une partie « raisonnable » A , alors $\mu(A \cap B) = \mu(A) + \mu(B)$ dès que les parties A et B sont disjointes).

Afin de réaliser cela, il faut en passer par quelques complications mathématiques, et cela essentiellement pour deux raisons :

- Il nous faut tout d'abord définir ce qu'est une partie « raisonnable » d'un ensemble E (s'il est aisé de définir le volume d'un polyèdre par exemple, il existe des parties dont la « frontière » est si complexe qu'elles ne possèdent pas de notion de volume).
- Ensuite, la propriété d'additivité si dessus est un peu trop naïve et se révèle insuffisante pour avoir de bonnes propriétés pour les mesures.

En effet, la classe des mesures additives a une structure mathématique extrêmement pauvre, ne permettant pas, en particulier, de définir une notion satisfaisante d'intégrale par rapport à ces mesures additives. On est donc conduit à utiliser les mesures possédant la propriété de σ -additivité (voir définition d'une mesure ci-dessous), ce qui nous oblige à considérer comme classe d'ensembles « mesurables » une tribu (voir définition 2) au lieu de la notion plus simple d'algèbre.

Définition 1 — Algèbre. Une algèbre (de Boole) \mathcal{E} sur un ensemble X est un ensemble non vide de parties de X , stable par passage au complémentaire et par union finie (donc aussi par intersection finie), i.e. :

1. $\mathcal{E} \neq \emptyset$
2. $\forall A \in \mathcal{E}, {}^c A \in \mathcal{E}$, où ${}^c A$ désigne le complémentaire de A dans X . (donc $X \in \mathcal{A}$)
3. si $A_1, \dots, A_n \in \mathcal{E}$ alors $A_1 \cup \dots \cup A_n \in \mathcal{E}$

Définition 2 — Tribu. Une tribu \mathcal{A} sur un ensemble X est un ensemble non vide de parties de X , stable par passage au complémentaire et par union dénombrable (donc aussi par intersection dénombrable), i.e. :

1. $\mathcal{A} \neq \emptyset$
2. $\forall A \in \mathcal{A}, {}^c A \in \mathcal{A}$
3. si $\forall n \in \mathbb{N}, A_n \in \mathcal{A}$ alors $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$

Ces deux dernières définitions étant placées l'une sous l'autre, leur différence doit apparaître clairement : dans le cas d'une algèbre, on a à faire à un nombre fini d'intersections ou de réunions, dans celui d'une tribu, on prend en compte un ensemble dénombrable (donc fini ou non).

Il est donc évident que toute tribu est une algèbre, la réciproque étant fausse.

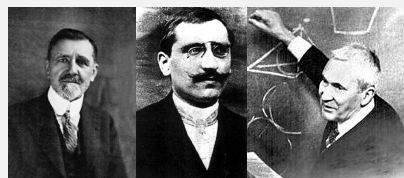
Définition 3 — Espace mesurable. Le couple (X, \mathcal{A}) est appelé espace mesurable ou espace probabilisable en fonction du contexte. Sur les espaces mesurables on définit des mesures (voir ci-après) ; sur les espaces probabilisables, on appelle ces mesures des probabilités.

La théorie de la mesure s'occupe de regarder un peu plus en détail ce qui se passe à l'intérieur des espaces dits mesurés (définis plus bas) : il s'agit de mesurer les différentes parties existant dans un tel espace, parties en lien avec la topologie, ce qui reboucle le sujet...

Lorsque l'on évoque le concept de mesure, on en arrive assez rapidement à se demander : Est-ce que l'on peut tout mesurer ? Est-ce que l'on doit tout mesurer ? Qu'est-ce qui est négligeable ?

En 1894, Émile Borel énonce la première définition d'ensemble négligeable. En 1897, il définit les ensembles mesurables. En 1901, Henri-Léon Lebesgue introduit la notion de mesure. La théorie se développe jusque dans les années 1950.

Andrei Kolmogorov proposera une axiomatisation du calcul des probabilités basée notamment sur l'intégrale définie à partir d'une mesure.



Borel Lebesgue Kolmogorov

Les parties de X qui appartiennent à la tribu \mathcal{A} sont appelées ensembles mesurables. Dans un contexte probabiliste, on les appelle événements (il suffit que $\mu(X) = 1$ pour que μ soit une probabilité).

Une mesure μ sur un ensemble X est une fonction qui associe à chaque élément d'une tribu d'un ensemble X un scalaire positif (une longueur, un volume, une probabilité...).

Définition 4 — Mesure. Soit (X, \mathcal{A}) , un espace mesurable. Une application μ définie sur \mathcal{A} , à valeurs dans $[0, +\infty]$ est appelée mesure lorsque les deux propriétés suivantes sont satisfaites :

1. L'ensemble vide a une mesure nulle : $\mu(\emptyset) = 0$
2. L'application μ est σ -additive : si E_1, E_2, \dots est une famille dénombrable de parties de X appartenant à \mathcal{A} et si ces parties sont deux à deux disjointes, alors la mesure $\mu(E)$ de leur réunion E est égale à la somme des mesures des parties :

$$\mu\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \mu(E_k). \quad (1.3)$$

On appelle espace mesuré un triplet (X, \mathcal{A}, μ) , où X est un ensemble, \mathcal{A} une tribu sur X et μ une mesure sur \mathcal{A} .

Complément sur le produit d'espaces mesurables. Soient (X, T) et (Y, U) deux espaces mesurables. On appelle rectangle mesurable du produit $\Omega = X \times Y$, toute partie de Ω de la forme $A \times B$ où A et B sont des éléments respectivement T et U -mesurables.

On appelle produit tensoriel des deux tribus T et U , la tribu engendrée par l'ensemble des rectangles mesurables. Cette tribu est notée $T \otimes U$, et est la plus petite tribu de Ω qui contient toutes les parties de la forme $A \times B$, $A \in T$, $B \in U$.

Le produit des espaces mesurables est $(X \times Y, T \otimes U)$ et est un espace mesurable :

$$\int f \, d(\mu \otimes \nu) = \int \int f(x, y) d\mu(x) d\nu(y) \quad (1.4)$$

1.3 Tribu borélienne, mesures de Dirac et Lebesgue

Définition 5 — Tribu borélienne. On appelle tribu borélienne sur un espace topologique donné la tribu engendrée par les ensembles ouverts. Dans le cas simple et fondamental de l'espace usuel à n dimensions, la tribu borélienne de \mathbb{R}^n est engendrée par une famille dénombrable de parties, les pavés, dont les sommets sont à coordonnées rationnelles.

On aura besoin de ces notions (dont la tribu borélienne) pour l'intégrale de Lebesgue et par extension toute la théorie de l'intégration qui est à la base de l'analyse numérique (et oui, formulation faible, quand tu nous tiens) ainsi notamment que pour le théorème de représentation de Riesz fondamental en éléments finis.

Notons que tout intervalle ouvert, fermé ou semi-ouvert appartient à la tribu borélienne de \mathbb{R} . Il en est de même de toute réunion finie ou dénombrable de ces intervalles.

Il n'est pas possible de donner une description plus concrète de la tribu borélienne de \mathbb{R} que ce qui a été fait. Toutes les réunions finies ou dénombrables d'intervalles sont des boréliens, mais certains boréliens ne sont pas de cette forme. En fait, toutes les parties de \mathbb{R} que l'on rencontre dans la pratique sont des boréliens. Il existe des parties de \mathbb{R} qui ne sont pas boréliennes, mais il faut un peu de sueur pour les construire.

Les tribus sont des familles de parties qui sont destinées à être mesurées. Pour pouvoir mesurer des parties suffisamment compliquées comme celles qui ne peuvent être définies que par des passages à la limite (comme l'ensemble triadique de Cantor), les tribus doivent être assez fines pour être stables par des opérations relativement générales comme le passage au complémentaire, les réunions et intersections dénombrables. Néanmoins, elles ne doivent pas être trop fines afin de ne pas contenir de parties non mesurables.

Rappelons également deux mesures simples et importantes :

- la mesure de comptage, qui donne le nombre d'éléments d'un ensemble.
- la mesure de Dirac ou masse de Dirac, qui est une mesure supportée par un singleton et de masse totale 1.

Définition 6 — Mesure de Dirac. Pour un espace mesurable (X, \mathcal{A}) et un point a de X , on appelle mesure de Dirac au point a la mesure notée δ_a sur (X, \mathcal{A}) telle que :

$$\forall A \in \mathcal{A}, \quad (\delta_a(A) = 1 \text{ si } a \in A \text{ et } \delta_a(A) = 0 \text{ si } a \notin A) \quad (1.5)$$

Le support de δ_a est réduit au singleton $\{a\}$.

Les masses de Dirac sont très importantes, notamment dans la pratique car elles permettent par exemple de construire des mesures par approximations successives.

Définition 7 — Mesure complète. Soit (X, \mathcal{A}, μ) un espace mesuré, on dit que μ est une mesure complète lorsque tout ensemble négligeable pour μ appartient à la tribu \mathcal{A} sur laquelle μ est définie, i.e. :

$$\forall M, N \in \mathcal{P}(X), \quad (N \subset M, M \in \mathcal{A} \text{ et } \mu(M) = 0) \Rightarrow N \in \mathcal{A} \quad (1.6)$$

La mesure de Lebesgue a permis de bâtir une théorie de l'intégration palliant les insuffisances de l'intégrale de Riemann (il suffit de vouloir intégrer $\mathbb{1}_{\mathbb{Q}}$ sur $[0; 1]$ avec Riemann pour être dans l'impasse). Nous détaillerons cela un peu plus au chapitre 4 sur les espaces de Lebesgue.

Parmi les définitions de cette mesure, nous présentons la plus intuitive, celle qui consiste à généraliser la notion de volume en gardant les mesures sur les pavés de \mathbb{R}^n .

Définition 8 — Mesure de Lebesgue. Il existe une plus petite mesure définie sur une tribu de parties de \mathbb{R}^n qui soit complète et coïncide sur les pavés avec leur volume (i.e. avec le produit des longueurs de leurs côtés).

Cette mesure est appelée la mesure de Lebesgue (notée λ_n) et sa tribu de définition la tribu de Lebesgue (notée \mathcal{L}_n et que nous ne définissons pas ici, et dont les éléments sont dits Lebesgue-mesurables).

Cette restriction aux boréliens de la mesure de Lebesgue est parfois dénommée mesure de Borel-Lebesgue.

Cela signifie par exemple qu'il existe une unique mesure de Lebesgue λ sur les boréliens de \mathbb{R} telle que $\forall a, b \in \mathbb{R}, \lambda([a; b]) = |b - a|$. Il en découle que pour un point $\{a\} = [a; a], \lambda(\{a\}) = 0$, puis par union dénombrable, que pour tout sous-ensemble X fini ou dénombrable de $\mathbb{R}, \lambda(X) = 0$. En particulier $\lambda(\mathbb{N}) = \lambda(\mathbb{Z}) = \lambda(\mathbb{Q}) = 0$. Un ensemble de mesure nulle est dit négligeable.

Si pour la mesure de Lebesgue λ sur \mathbb{R} , la notion de « presque partout » correspond bien à l'intuition, ce n'est pas vrai en général. Par exemple, pour $\mu = \delta_a$ sur la tribu de l'ensemble des parties de X , une propriété est vraie μ -pp simplement si elle est vérifiée en a .

Si la mesure μ est nulle, toute propriété est vérifiée pp (ainsi que sa négation).

« Construire une mesure », c'est montrer qu'il existe une unique mesure qui vérifie certaines propriétés. Pour cela, on utilise le théorème (ou lemme) de la classe monotone (dû à Waław Sierpiński et popularisé par Dynkin) pour montrer l'unicité et le théorème de Carathéodory pour montrer l'existence.

1.4 Propriétés de la mesure de Lebesgue

On appelle mesure de Lebesgue sur un espace euclidien E la mesure image de la mesure de Lebesgue sur \mathbb{R}^n par n'importe quelle isométrie de \mathbb{R}^n dans E .

Soit A une partie Lebesgue-mesurable d'un espace euclidien E . On appelle mesure de Lebesgue sur A la restriction à A de la mesure de Lebesgue de E .

Théorème 4 La mesure de Lebesgue est invariante sous toutes les isométries. Elle est en particulier invariante sous les translations : c'est une mesure de Haar du groupe topologique \mathbb{R}^n .

Oui, la mesure de Lebesgue peut être vue comme une mesure de Haar. Mais, historiquement, la mesure de Haar est définie plus tard (on parle souvent de *la* mesure de Haar, alors que l'on devrait parler d'*une* mesure de Haar). Elle généralise celle de Lebesgue.

Définition 9 — Mesure régulière. Une mesure (positive) μ définie sur une tribu contenant la tribu borélienne d'un espace topologique X est dite régulière lorsque elle est à la fois intérieurement régulière et extérieurement régulière :

1. $\forall A \subset X$ de la tribu, $\mu(A) = \sup\{\mu(K) \mid K \text{ compact contenu dans } A\}$;
2. $\forall A \subset X$ de la tribu, $\mu(A) = \inf\{\mu(O) \mid O \text{ ouvert contenant } A\}$.

Théorème 5 La mesure de Lebesgue est finie sur tout compact ; chaque compact, qui est borné, pouvant être enfermé dans un cube.

Elle est par voie de conséquence régulière, \mathbb{R}^n étant métrisable, localement compact et séparable.

- \mathbb{R}^n est métrisable : un espace topologique est un espace métrisable lorsqu'il est homéomorphe à un espace métrique (un homéomorphisme est une application bijective continue entre deux espaces topologiques dont la réciproque est continue) ;
- \mathbb{R}^n est localement compact : un espace localement compact est un espace séparé qui, sans être nécessairement compact lui-même, admet des voisinages compacts pour tous ses points (De plus, on a la propriété : Tout espace localement compact est régulier) ;
- \mathbb{R}^n est séparable : un espace séparable est un espace topologique contenant un sous-ensemble fini ou dénombrable et dense, i.e. contenant un ensemble fini ou dénombrable de points dont l'adhérence est égale à l'espace topologique tout entier.

Théorème 6 — Théorème de Borel-Lebesgue. On retiendra de ce théorème que dans \mathbb{R}^n , les compacts sont les ensembles fermés bornés.

En complément à la première note historique de ce chapitre, nous vous proposons une illustration très concrète de l'application de la topologie, que vous avez très certainement rencontrée : la carte du métro.



Carte du métro



Beck

La représentation schématique généralement utilisée pour représenter un réseau de métro a été mise au point, la première fois, en 1931 par Henry Beck, alors dessinateur industriel de 29 ans et engagé comme intérimaire par la société du métro londonien. Facile à comprendre et à utiliser, esthétique... elle est pourtant fautive à tous les égards sauf deux.

Elle n'est pas à l'échelle, donc toutes les distances sont fausses ; les lignes droites reliant les stations ne traduisent absolument pas le cheminement réel du métro sous les rues ; les orientations sont fausses (une ligne verticale ne signifie pas que le trajet s'effectue selon l'axe nord-sud).

Le premier aspect exact est que si une station de métro est représentée au nord de la Tamise, alors il en est de même pour la station réelle. Le second aspect exact est la description du réseau : l'ordre des stations sur chaque ligne et les interconnexions entre les lignes sont fidèles à la réalité. C'est d'ailleurs ce second aspect qui est finalement le seul dont les voyageurs ont effectivement besoin.

Notons enfin que cette illustration permet de comprendre aisément comment la notion de distance peut être appréhendée en topologie.

1.5 Petit exemple amusant d'injection dans un Hilbert

Soient V et H deux espaces de Hilbert sur \mathbb{R} et V inclus dans H avec injection continue et V dense dans H .

Comme V est inclus dans H , l'injection est tout bêtement $i : x \mapsto x$.

Dans ces conditions, H est un espace de Hilbert avec un produit scalaire $(\cdot | \cdot)_H$, et V muni du produit scalaire induit est un sous-espace dense, donc ne peut pas être un espace de Hilbert par restriction du produit scalaire $(\cdot | \cdot)_H$. La structure hilbertienne de V est définie par un produit scalaire $(\cdot | \cdot)_V$ propre à V .

Il y a donc deux topologies sur V : la topologie hilbertienne propre à V et la topologie héritée de la structure hilbertienne de H . La continuité de l'injection i impose donc une condition sur ces topologies : la topologie hilbertienne de V est plus fine que la trace sur V de la topologie hilbertienne de H .

... et maintenant que le terme « injection » a été prononcé, il est temps de passer au chapitre suivant...

Chapitre 2

Applications et morphismes

Résumé — Au chapitre précédent, des espaces ont été définis, mais on ne s'est pas intéressé beaucoup aux relations entre eux ou au sein d'eux.

Des distances, normes... ont été introduites, sans utiliser plus que ça le vocable de fonction. Le terme d'injection a été prononcé à la fin du chapitre précédent comme fil conducteur pour introduire celui-ci...

Dans ce chapitre, nous ne présenterons pour le coup que des choses extrêmement « rudimentaires » et toutes vues en taupe ou avant. Il s'agit uniquement d'un aide-mémoire.

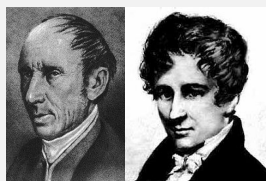
Histoire

L'univers mathématiques du début du XVIII^e siècle est dominé par Leonhard Euler et par ses apports tant sur les fonctions que sur la théorie des nombres, tandis que Joseph-Louis Lagrange éclairera la seconde moitié de ce siècle.

Euler a introduit et popularisé plusieurs conventions de notation par le biais de ses nombreux ouvrages largement diffusés. Plus particulièrement, il a introduit la notion de fonction (dans *L'Introductio in analysin infinitorum*, premier traité dans lequel le concept de fonction est à la base de la construction mathématique, et dont les premiers chapitres lui sont consacrés) et a été le premier à écrire $f(x)$ pour désigner la fonction f appliquée à l'argument x , en 1734 (bien que le terme de « fonction » apparaisse pour la première fois dans un manuscrit d'août 1673 de Leibniz, resté inédit, et intitulé *la Méthode inverse des tangentes ou à propos des fonctions*). Il a également introduit la notation moderne des fonctions trigonométriques, la lettre e pour la base du logarithme naturel (également connue sous le nom de nombre d'Euler) en 1727, la lettre grecque Σ pour désigner une somme en 1755 et la lettre i pour représenter l'unité imaginaire, en 1777. L'utilisation de la lettre grecque π pour désigner le rapport de la circonférence d'un cercle à son diamètre a également été popularisée par Euler, mais celui-ci n'est pas à l'origine de la notation.

Les différentes techniques mises au point (par exemple pour la résolution des équations différentielles, le développement en séries entières ou asymptotiques, applications aux réels négatifs, aux complexes...) conduisent à s'intéresser à la « fonction » en tant que sujet d'étude.

À la fin du XVIII^e siècle, les mathématiciens croient encore, pour peu de temps, que la somme infinie de fonctions continues est continue, et (pour plus longtemps) que toute fonction continue admet une dérivée... (sur ces notions, voir chapitre suivant).



Cauchy

Abel

C'est Cauchy qui met un peu d'ordre dans tout cela en montrant que la somme d'une série numérique n'est commutativement convergente que si la série est absolument convergente. Mais Cauchy, qui pourtant n'est qu'à un doigt de la notion de convergence uniforme, énonce un théorème faux de continuité d'une série de fonctions continues qu'Abel contredit par un contre-exemple le 16 janvier 1826 : Cauchy affirme en 1821 que la somme d'une série de fonctions continues est toujours continue. Cinq ans plus tard, Abel propose un contre-exemple en considérant la suite $(f_n)_{n \geq 1}$ d'applications continues de \mathbb{R} dans \mathbb{R} de terme général $f_n(x) = (-1)^n / n \cdot \sin(nx)$.

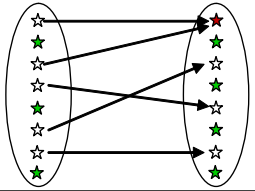
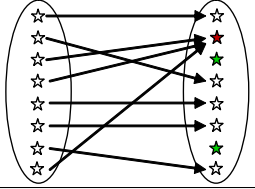
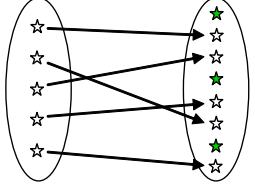
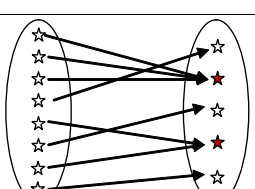
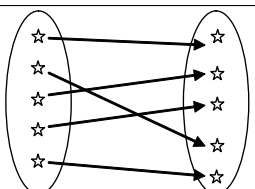


Euler

Lagrange

2.1 Fonction, application, injection, surjection, bijection

Si ça c'est pas du rappel de base, je ne m'y connais pas...

Fonction	Un « truc » qui met en relation certains éléments de E avec certains éléments de F.	
Application	Fonction définie partout : (plus d'étoile verte dans E) Une fonction f est une application si son ensemble de départ est égal à son ensemble de définition.	
Injection	Application telle que tout élément de F a au plus 1 antécédent. F a au moins autant d'éléments que E. (plus d'étoile verte dans E et plus d'étoile rouge dans F) Soit $\forall x, y \in E^2, x \neq y \Rightarrow f(x) \neq f(y)$, ou $\forall x, y \in E^2, f(x) = f(y) \Rightarrow x = y$	
Surjection	Application telle que tout élément de F a au moins 1 antécédent. F a au plus autant d'éléments que E. (plus d'étoile verte dans E et plus d'étoile verte dans F) Soit $\forall y \in F, \exists x \in E, f(x) = y$, ou f est surjective si son ensemble image est égal à son ensemble d'arrivée.	
Bijection	C'est une injection ET une surjection : chaque élément de l'ensemble de départ correspond à un seul élément de l'ensemble d'arrivée et vice-versa.	

* = élément n'ayant pas de relation ; * = élément ayant 1 relation ; * = élément ayant plus d'une relation.

Tableau 2.1: Types de fonctions

Une étoile rouge dans E n'a pas de sens, cela voudrait dire qu'un élément de E peut avoir plusieurs valeurs différentes par la relation considérée...

En fait, on sait donner un sens à cela. C'est ce que l'on appelle une fonction multivaluée ou fonction multiforme ou fonction multivoque ou multifonction. L'exemple le plus simple d'une fonction multiforme est la fonction réciproque d'une application non injective (penser simplement aux fonctions circulaires).

On trouve les fonctions multiformes en analyse complexe : lorsque l'on veut utiliser le théorème des résidus pour calculer une intégrale réelle, on peut être amené à considérer des restrictions (déterminations) qui font de ces fonctions multiformes des fonctions (univoques), par exemple en utilisant la théorie des revêtements qui considère des fonctions sur des surfaces de Riemann.

En restreignant une fonction à son domaine de définition, on en fait une application. En la restreignant en plus à son ensemble d'arrivée on en fait une surjection (une surjection, c'est un « truc » défini partout sur E et F).

Quand on a une surjection, on est sûr que tout élément de l'ensemble de départ a une image, et

que tout élément de l'ensemble d'arrivée a un antécédent (au moins un même).

Dans la pratique, on ne fait pas de distinction formelle entre fonction et application...

$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \sqrt{x}$ est une fonction, pas une application, car la racine carrée n'est pas définie sur \mathbb{R} mais sur \mathbb{R}_+ . $f : \mathbb{R}_+ \rightarrow \mathbb{R}, x \mapsto \sqrt{x}$ est une application ! Mais pourquoi s'intéresserait-on à f là où elle n'est pas définie... De plus, $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+, x \mapsto \sqrt{x}$ est une surjection par définition (puisque l'on considère f de son ensemble de définition jusqu'à son ensemble image). C'est évidemment une bijection (il ne reste que l'injectivité à prouver...).

Définition 10 — Support. On appelle support d'une fonction f l'adhérence (ou la fermeture, i.e. le plus petit fermé) du lieu où la fonction n'est pas nulle :

$$\text{supp}(f) = \overline{\{x \in \mathbb{R}^n, f(x) \neq 0\}} \quad (2.1)$$

2.2 Morphismes

Cette section est extraite du cours sur les structures algébriques. Nous l'avons toutefois sérieusement amputée pour coller à l'objectif de ce document.

2.2.1 Présentation

Histoire

Mort au cours d'un duel à l'âge de vingt ans (ce qui en fait un héros romantique), il laisse un manuscrit élaboré trois ans plus tôt, dans lequel il établit qu'une équation algébrique est résoluble par radicaux si et seulement si le groupe de permutation de ses racines a une certaine structure, qu'Emil Artin appellera justement résoluble.

Son Mémoire sur les conditions de résolubilité des équations par radicaux, publié par Joseph Liouville quatorze ans après sa mort, a été considéré par ses successeurs, en particulier Sophus Lie, comme le déclencheur du point de vue structural et méthodologique des mathématiques modernes.

Toutefois, pour être tout à fait exact, Lagrange, reprenant une idée de d'Alembert en vue de prouver qu'un polynôme de degré n possède n racines (i.e. que \mathbb{C} est algébriquement clos), utilise des résultats sur les fonctions semblables des racines, i.e. des fonctions rationnelles des racines qui restent invariantes par les mêmes permutations. Ce résultat, qu'il a établi dans son fameux mémoire de 1771 *Réflexions sur la résolution algébrique*, inspirera Abel et Galois et peut être considéré comme le tout premier de la théorie des groupes.



Galois

Soient deux ensembles G et G' munis d'un même type de structure (topologique, groupe, anneau, espace vectoriel, algèbre...). Un morphisme (ou homomorphisme) de $G \rightarrow G'$ est une application f qui respecte cette structure.

Pour ce faire, cette application doit vérifier certaines conditions, notamment une certaine « linéarité » vis-à-vis des lois des G et G' (on pourrait également remplacer le terme linéarité par « capacité à faire sortir de la fonction »).

Un morphisme entre deux espaces topologiques est tout simplement une application continue (voir chapitre suivant sur la continuité). C'est d'ailleurs ce dernier terme qui est utilisé en topologie, pas celui de morphisme (mais cela revient bien au même).

Un morphisme de groupe entre $(G, *)$ et (G', \star) satisfait à l'égalité suivante qui est bien une « condition de linéarité par rapport à la loi » : $\forall (x, y) \in G, f(x * y) = f(x) \star f(y)$. En particulier, si e et e' sont les éléments neutres de G et G' , alors : $f(e) = e'$. Une autre conséquence directe est que : $\forall x \in G, f(x^{-1}) = [f(x)]^{-1}$.

2.2.2 Cas des espaces vectoriels : application et forme linéaires

Définition 11 — Morphisme d'espace vectoriel. Soient $(E, +, \cdot)$ et $(F, +, \cdot)$ deux espaces vectoriels sur un corps \mathbb{K} . Un morphisme d'ev f entre E et F est une application qui respecte la condition de linéarité par rapport aux lois $+$ (en fait qui est un morphisme de groupe entre les groupes $(E, +)$ et $(F, +)$) et qui conserve la « linéarité par rapport à la multiplication par un scalaire » :

$$\begin{cases} \forall (x, y) \in E^2, & f(x + y) = f(x) + f(y) & \text{condition d'additivité} \\ \forall x \in E, \quad \forall \lambda \in \mathbb{K}, & f(\lambda \cdot x) = \lambda \cdot f(x) & \text{condition d'homogénéité} \end{cases} \quad (2.2)$$

Ceci est équivalent à la condition suivante (on parle de « préservation des combinaisons linéaires ») :

$$\forall (x, y) \in E^2, \quad \forall \lambda \in \mathbb{K}, \quad f(\lambda \cdot x + y) = \lambda \cdot f(x) + f(y) \quad (2.3)$$

et on utilise plutôt le terme d'application linéaire ou d'opérateur linéaire ou encore de transformation linéaire.

Dans le cas où $F = \mathbb{K}$, on ne parle pas d'application linéaire mais de forme linéaire. Une forme linéaire est donc une application linéaire définie sur E et à valeurs dans \mathbb{K} (supposé commutatif). En d'autres termes, on parle de forme au lieu d'application, mais c'est la même chose !

Si φ et ψ sont des formes linéaires et a et b des éléments de \mathbb{K} :

$$\forall x \in E, \quad (a\varphi + b\psi)(x) = a \cdot \varphi(x) + b \cdot \psi(x) \quad (2.4)$$

L'application constante de valeur $0_{\mathbb{K}}$ s'appelle la « forme linéaire nulle ».

2.2.3 Endo, iso, auto -morphisms

Un endomorphisme est un morphisme d'une structure dans elle-même.

Un isomorphisme est un morphisme f entre deux ensembles munis de la même espèce de structure, tel qu'il existe un morphisme f' dans le sens inverse, tels que $f \circ f' = \text{id}$ et $f' \circ f = \text{id}$ sont les identités des structures. Un isomorphisme est un morphisme bijectif.

Deux ensembles munis du même type de structure algébrique sont dits isomorphes s'il existe un isomorphisme entre les deux ensembles.

L'isomorphie présente un intérêt majeur, car elle permet de transposer des résultats et propriétés démontrés sur l'un des deux ensembles à l'autre.

Un automorphisme est un isomorphisme d'une structure dans elle-même, i.e. à la fois un isomorphisme et un endomorphisme. Un automorphisme est un endomorphisme bijectif.

L'identité d'un ensemble est toujours un automorphisme, quelle que soit la structure considérée.

On note :

- $\mathcal{L}_{\mathbb{K}}(E, F)$ l'espace vectoriel des applications linéaires de E dans F ;
- $\text{Isom}_{\mathbb{K}}(E, F)$ l'ensemble des isomorphismes de E dans F ;
- $\mathcal{L}_{\mathbb{K}}(E)$ l'espace vectoriel des endomorphismes de E ;
- $\text{GL}_{\mathbb{K}}(E)$ le « groupe linéaire », i.e. le groupe des automorphismes de E .

2.2.4 Espace dual d'un espace vectoriel

Définition 12 — Espace dual. On appelle espace dual d'un espace vectoriel E l'ensemble des formes linéaires sur E . Il est lui-même un \mathbb{K} -espace vectoriel, et on le note E^* ou $\text{hom}(E, \mathbb{K})$.

La structure d'un espace et celle de son dual sont très liées. Nous allons détailler quelques points en nous restreignant aux cas qui nous intéressent (cas réel, dimension finie).

Remarque. Si l'on dispose, sur l'espace vectoriel considéré E , d'un produit scalaire $\langle \cdot, \cdot \rangle$ (voir chapitre 1), alors il existe un moyen « naturel » de plonger E dans E^* , i.e. d'associer à chaque élément de E un élément du dual, et ce de manière à former un isomorphisme entre E et un sous-espace de E^* : à chaque élément $x \in E$ on associe la forme linéaire $\varphi_x : E \rightarrow \mathbb{K}; y \mapsto \langle x, y \rangle$. Alors l'application $f : E \rightarrow E^*; x \mapsto \varphi_x$ est une application linéaire injective, donc l'espace E est isomorphe au sous-espace $f(E)$ de E^* .

Si l'espace E est de dimension finie n , alors l'espace dual E^* est isomorphe à E et est donc lui aussi de dimension n . On a alors le théorème de la base duale (que je ne présente pas, car je n'ai pas parlé de base... mais peut-être pourrons-nous nous passer de ces rappels dans ce document).

Pour $x \in E$, on note $\langle \varphi, x \rangle$ pour $\varphi(x)$. Cette notation est appelée crochet de dualité.

Définition 13 — Dual topologique. Soit E un espace vectoriel topologique sur le corps \mathbb{K} . Le dual topologique E' de E est le sous-espace vectoriel de E^* (le dual algébrique de E) formé des formes linéaires continues.

Si l'espace est de dimension finie, le dual topologique coïncide avec le dual algébrique, puisque dans ce cas toute forme linéaire est continue. Mais dans le cas général, l'inclusion du dual topologique dans le dual algébrique est stricte.

Remarques topologiques. En complément, quelques mots sur les topologies faible et faible-*, afin de pouvoir faire dans de bonnes conditions une remarque sur leur usage.

Définition 14 — Topologies faible et faible-*. Soient E un espace vectoriel topologique et E' son dual topologique, i.e. l'espace vectoriel des formes linéaires continues sur E .

On appelle topologie forte sur E la topologie initiale de E ;

On appelle topologie faible sur E la topologie la plus grossière qui rende continus tous les éléments de E' ;

On appelle topologie faible-* sur E' la topologie la plus grossière qui rende continus tous les éléments de E .

Notons que si E est réflexif, i.e. si $E'' = E$ alors la topologie faible et la topologie faible-* coïncident.

Les topologies faibles ne sont en général pas métrisables, et ne peuvent donc se définir par la seule donnée des notions de convergence de suites. Cependant, c'est bien la notion de convergence faible de suites qui est utile en pratique.

On place ici des remarques faisant appel aux espaces de Lebesgue définis au chapitre 4.

Soit Ω un ouvert de \mathbb{R}^n . Pour $1 < p < \infty$, les topologies faible et faible-* sur $L^p(\Omega)$ coïncident, et la notion de convergence associée (convergence contre des fonctions test dans L^p) est :

$$\forall g \in L^{p'}(\Omega), \quad \int_{\Omega} f_n g \longrightarrow \int_{\Omega} f g$$

Pour $p = \infty$, la topologie faible-* correspond à la convergence contre des fonctions test dans L^1 ; pour $p = 1$, la topologie faible correspond à la convergence contre des fonctions test dans L^∞ . On s'interdira en revanche de considérer la convergence faible-* dans L^1 , ou la convergence faible dans L^∞ (la question de savoir si L^1 est le dual de L^∞ touche à de subtiles questions d'axiomatique, et la réponse est négative si l'on admet l'axiome du choix...)

Quel est l'intérêt d'appauvrir la topologie ? Une des motivations majeures est que moins il y a d'ouverts, plus il y a de compacts. Il est beaucoup plus facile d'être compact pour la topologie faible que pour la topologie forte.

2.2.5 Noyau et image

Définition 15 — Noyau. Le noyau du morphisme f est l'ensemble des antécédents de l'élément neutre :

$$\ker(f) = \{x \in E, f(x) = 0\} = f^{-1}(\{0\}) \quad (2.5)$$

et f est injectif si et seulement si son noyau est réduit à $\{0\}$.

Définition 16 — Image. L'image du morphisme f est l'image par f de E :

$$\text{im}(f) = \{f(x), x \in E\} = f(E) \quad (2.6)$$

et f est surjectif si et seulement si son image est égale à F .

Dans le cas d'espaces vectoriels, l'ensemble $\ker(f)$ est un sous-espace vectoriel de E et l'ensemble $\text{im}(f)$ est un sous-espace vectoriel de F .

Théorème 7 — Théorème du rang. Il est assez visible que (théorème de factorisation) f induit un isomorphisme de l'espace vectoriel quotient $E/\ker(f)$ sur l'image $\text{im}(f)$. Deux espaces isomorphes ayant même dimension, il s'en suit la relation, valable pour un espace E de dimension finie, appelée théorème du rang :

$$\dim(\ker(f)) + \dim(\text{im}(f)) = \dim(E) \quad (2.7)$$

Le nombre $\dim(\text{im}(f))$ est aussi appelé rang de f et est noté $\text{rg}(f)$.

On a également :

- l'image réciproque d'un sous-espace vectoriel de F par f est un sous-espace vectoriel de E ;
- l'image directe d'un sous-espace vectoriel de E par f est un sous-espace vectoriel de F .

2.3 Opérateur

Le terme « opérateur » a été utilisé au paragraphe précédent... regardons d'un peu plus près.

D'une manière générale, un opérateur est une application entre deux espaces vectoriels topologiques.

Un opérateur $O : E \rightarrow F$ est linéaire si et seulement si :

$$\forall (\lambda, \mu) \in \mathbb{K}^2, \quad \forall (x_1, x_2) \in E, \quad O(\lambda x_1 + \mu x_2) = \lambda O(x_1) + \mu O(x_2) \quad (2.8)$$

où \mathbb{K} est le corps des scalaires de E et F .

Lorsque $F = \mathbb{R}$, un opérateur est une fonctionnelle sur E .

Un opérateur est continu s'il est continu en tant qu'application (pour la définition de la continuité, voir chapitre suivant).

Un opérateur différentiel est un opérateur agissant sur des fonctions différentiables au sens des dérivés ordinaires ou partielles : voir définition au chapitre suivant.

On définit également les opérateurs différentiels elliptiques et hyperboliques. Mais pour cela, il faut introduire les notions de symbole et symbole principale d'un opérateur... et cela ne nous semble ni adapté à ce document, ni suffisamment « naturel » pour cet exposé. Nous nous contenterons de définir plus loin les notions d'équations aux dérivées partielles linéaires et homogènes du second-ordre dites elliptiques, hyperboliques et paraboliques.

Chapitre 3

Continuité et dérivabilité

Résumé — Dans ce chapitre, nous nous intéresserons aux notions de continuité et de dérivabilité, les espaces nécessaires ayant été définis précédemment, ainsi que les notions d'application...

Dans la mesure où les problèmes que nous souhaitons aborder, i.e. ceux issus de la physique, sont généralement décrits par des équations différentielles ou des équations aux dérivées partielles, on comprend bien que la notion de dérivation est centrale.

Mais n'oublions pas que ces notions de continuité et de différentiabilité n'ont pas toujours été définies de manière précise au cours de l'histoire et ont donné lieu à de bien terribles affrontements entre nos glorieux anciens.

3.1 Continuité et classe C^0

Histoire

Dans le manuscrit de 1673 *la Méthode inverse des tangentes ou à propos des fonctions*, Leibniz dit : « J'appelle fonctions toutes les portions des lignes droites qu'on fit en menant des droites indéfinies qui répondent au point fixe et aux points de la courbe ; comme sont abscisse, ordonnée, corde, tangente, perpendiculaire, sous-tangente, sous-perpendiculaire... et une infinité d'autres d'une construction plus composée, qu'on ne peut figurer. » Finalement, au terme d'une correspondance nourrie entre Leibniz et Jean Bernoulli, celui-ci donne en 1718 la définition suivante : « On appelle fonction d'une grandeur variable une quantité composée, de quelque manière que ce soit, de cette grandeur variable et des constantes. ». Il propose la notation ϕx .

La continuité est en quelque sorte contenue, sous-jacente à ces définitions car les fonctions considérées sont « physiques » et ne présentent au plus qu'un nombre fini de discontinuités.

Dans son *Introductio in analysin infinitorum* de 1748,

Euler définit une fonction d'une quantité variable comme « une expression analytique composée d'une manière quelconque de cette quantité variable et de nombres ou de quantités constantes ». Le mot « analytique » n'est pas davantage précisé. En fait, pour Euler, une fonction est une combinaison quelconque d'opérations prises dans le stock des opérations et des modes de calcul connus de son temps, et applicables aux nombres : opérations classiques de l'algèbre, exponentielle, logarithme, passage d'un arc à ses lignes trigonométriques..., certaines de ces opérations pouvant être itérées un nombre illimité de fois...

Dans ce même ouvrage, Euler dit qu'une fonction est continue si elle est définie par une seule expression analytique (finie ou infinie) et mixte ou discontinue si elle possède plusieurs expressions analytiques suivant ses intervalles de définition.

La définition actuelle est celle due à Bernard Bolzano dans sa théorie des fonctions en 1816 : « La fonction $f(x)$ varie suivant la loi de continuité pour la valeur x si la différence $|f(x + w) - f(x)|$ peut-être rendue plus petite que toute valeur donnée. » Il existe une notion de continuité uniforme qui est plus forte que la simple continuité et fixée par Heinrich Eduard Heine en 1872.



Euler



Bolzano



Heine

La continuité est une propriété topologique (donc indépendante de la métrique).

Définition 17 — Continuité d'une fonction en un point (version topologique). Soit f une application d'un espace topologique E dans un espace topologique F . On dit que f est continue en un point a de E si, quelque soit le voisinage W de $f(a)$ dans F , il existe un voisinage V de a dans E tel que $\forall x \in V, f(x) \in W$, c'est-à-dire que l'image réciproque de tout voisinage de $f(a)$ est un voisinage de a .

Cette définition est donnée pour la culture, car nous n'avons pas rappelé la notion de voisinage dans ce document. Cela n'a pas d'importance dans ce contexte puisque nous travaillerons sur des cas moins généraux.

Évidemment, une application de E dans F est continue si elle est continue en tout point de E . Ramenons nous à des choses plus connues et plus en lien avec ce document. Dans le cas des espaces métriques, la continuité se définit comme suit.

Définition 18 — Continuité d'une fonction en un point (version métrique). Soient (E, d) et (E', d') deux espaces métriques, $f : E \rightarrow E'$ et $a \in E$. On dit que l'application f est continue en a si :

$$\forall \varepsilon > 0, \quad \exists \eta > 0, \quad \forall x \in E, \quad \left[d(x, a) < \eta \implies d'(f(x), f(a)) < \varepsilon \right] \quad (3.1)$$

Ainsi f est continue en a si et seulement si la limite de f en a existe (elle vaut alors nécessairement $f(a)$).

Considérons la fonction f définie sur \mathbb{R}^2 par :

$$f(x, y) = \begin{cases} \frac{xy}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{si } (x, y) = (0, 0) \end{cases} \quad (3.2)$$

Elle n'a pas de limite en $(0, 0)$.

En effet, $f(x, y)$ est continue partout sur $\mathbb{R}^2 \setminus \{(0, 0)\}$. Elle est continue sur l'axe des abscisses et des ordonnées où elle est $\equiv 0$. Elle est donc séparément continue à l'origine, et par suite dans tout le plan. Mais elle n'est pas continue par rapport à l'ensemble des variables à l'origine, car sur la droite $y = mx$, elle prend la valeur $m/(1 + m^2)$ en dehors de l'origine ; or $\frac{m}{1+m^2} \neq 0$ dès que $m \neq 0$, et par conséquent elle ne tend pas vers 0 lorsque (x, y) tend vers l'origine. Une telle fonction est dite partiellement continue.

Pour une fonction réelle, on peut définir une fonction continue comme une fonction dont on peut tracer le graphe sans lever le crayon. Si l'on exclut certaines fonctions très particulières (comme les fractales), alors l'idée générale de la continuité est bien traduite par cette phrase (la fonction ne présente pas de « saut »).

Définition 19 — Continuité d'une fonction réelle en un point. Une fonction $f : I \subset \mathbb{R} \rightarrow \mathbb{R}$ sera dite continue au point $a \in I$ si :

$$\forall \varepsilon > 0, \quad \exists \eta > 0, \quad \forall x \in I, \quad \left[|x - a| < \eta \implies |f(x) - f(a)| < \varepsilon \right]. \quad (3.3)$$

La classe des fonctions continues est notée C^0 . Elle inclut par exemple les fonctions continues par morceaux ainsi que les constantes (dont la fonction nulle).

Attention à ne pas confondre la classe des fonctions continues C^0 avec C_0 l'ensemble des fonctions continues qui s'annulent à l'infini (sous-espace de l'espace des fonctions continues).

3.2 Continuité de Hölder et Lipschitz

Remarque : Dans le cas des espaces métriques, nous avons vu qu'il était possible de redéfinir la continuité à l'aide des ε plutôt que par les voisinages.

Avec Hölder et Lipschitz, la notion de « continuité uniforme » nous est proposée. La distinction entre continuité et continuité uniforme est la même que celle entre la convergence simple et la convergence uniforme dans le cas des séries (pour le lecteur qui s'en souviendrait). En effet, cette continuité uniforme ne regarde pas comment (quel ε) la fonction est continue en chaque point, mais comment elle est continue dans sa globalité, i.e. lorsque ce fameux ε n'est plus lié à la position sur la courbe, mais est fixé pour la fonction entière.



Hölder

Lipschitz

La continuité höldérienne ou condition de Hölder est une condition suffisante (mais non nécessaire) pour qu'une application définie entre deux espaces métriques soit continue.

La définition s'applique en particulier pour les fonctions d'une variable réelle.

Définition 20 — Fonction a -höldérienne. Si (E, d) et (F, d') sont deux espaces métriques, une fonction $f : E \rightarrow F$ est dite a -höldérienne s'il existe une constante $C > 0$ telle que :

$$\forall (x, y) \in E^2, \quad d'(f(x), f(y)) \leq C d(x, y)^a \quad (3.4)$$

La continuité höldérienne d'une fonction dépend donc d'un paramètre réel strictement positif $a \in]0; 1]$, et prend en compte toutes les variations de la valeur de la fonction sur son ensemble de définition. Si $0 < a \leq 1$ est fixé, l'ensemble des fonctions réelles a -höldériennes est un espace vectoriel, conventionnellement noté $\mathcal{C}^{0,a}(E, \mathbb{R})$.

Théorème 8 Toute application f qui est a -höldérienne est continue. Mieux, elle est uniformément continue, dans le sens suivant :

$$\text{Si } \varepsilon > 0, \text{ alors pour } \eta = (\varepsilon/C)^{1/a}, \quad d(x, y) < \eta \Rightarrow d(f(x), f(y)) < \varepsilon \quad (3.5)$$

Le réel η dépend de ε mais est indépendant de la variable x parcourant l'espace de définition de l'application.

Définition 21 — Fonction lipschitzienne. Lorsque $a = 1$, l'application est dit lipschitzienne. Une application lipschitzienne est plus « régulière » qu'une fonction simplement continue.

Toute fonction lipschitzienne (en tant que fonction höldérienne) est uniformément continue. Toute fonction réelle lipschitzienne est (absolument continue donc à variation bornée donc) dérivable presque partout pour la mesure de Lebesgue et sa dérivée est essentiellement bornée.

3.3 Dérivée

Le nombre dérivé en un point d'une fonction à variable et valeurs réelles est le coefficient directeur de la tangente au graphe de cette fonction en ce point, ou aussi le coefficient directeur de l'approximation affine de cette fonction en ce point : ce nombre n'est donc défini que si cette tangente, ou cette approximation, existe. La dérivée d'une fonction f est une fonction qui, à tout nombre pour lequel f admet un nombre dérivé, associe ce nombre dérivé.

Définition 22 — Dérivée d'une fonction. Soit f une application d'un ouvert Ω du corps \mathbb{K} (resp. un intervalle de \mathbb{R}) dans un espace affine normé F (resp. \mathbb{R}), alors on peut donner un sens,

pour $a \in \Omega$, à la quantité :

$$f'(a) = \lim_{\substack{h \neq 0, h \rightarrow 0 \\ a+h \in \Omega}} \frac{f(a+h) - f(a)}{h} \in \mathbb{F} \quad (3.6)$$

que l'on appelle le vecteur dérivé ou simplement la dérivée de f en a .

Histoire

Au XVII^e siècle, la compréhension mais surtout la modélisation (i.e. la mise en équations) de phénomènes physiques et techniques conduit à la création au siècle suivant de l'analyse en tant que branche des mathématiques abordant les notions de dérivation, intégration et équations différentielles.

Les échanges, les conceptions et la compréhension des infinitésimaux ont animé le monde scientifique pendant bien longtemps, la discussion était tout autant philosophique que mathématique, ce qui est somme toute assez normal compte tenu du sujet... Les fondateurs incontestés de l'analyse sont Newton et Leibniz. La portée de leurs travaux est considérable car ils vont permettre non seulement la compréhension des courbes (puis le calcul des aires), mais aussi celle du mouvement des corps. C'est véritablement une révolution où l'on passe d'une science de la statique à une science de la dynamique.

Un débat houleux sur la paternité du calcul différentiel entre mathématiciens britanniques et allemands fit rage, mais Newton et Leibniz s'en tinrent à l'écart. Toutefois, cette polémique entre les deux camps ne sera pas sans effet. Celle-ci fera que les anglais resteront à l'écart du développement général des mathématiques au XVIII^e siècle, la tradition newtonienne dominante ayant abouti à une certaine stagnation scientifique. Au début du XIX^e siècle, avec la diffusion des notations symboliques leibniziennes du calcul infinitésimal, un petit groupe de mathématicien de Cambridge se mettra à réfléchir sur le rôle et l'importance de la symbolique.



Newton Leibniz

L'opinion générale est aujourd'hui que Leibniz s'est presque certainement inspiré de certaines des idées de Newton (dont les travaux sont antérieurs, mais non publiés au moment où Leibniz publie), mais que sa contribution reste suffisamment importante pour que le mérite de l'invention du calcul différentiel soit accordé aux deux hommes.

L'approche de Newton est proche de la physique : il parle en termes de mouvement physique et ses notations ne sont employées que très rarement en dehors de la sphère des physiciens.

L'approche de Leibniz en revanche est plus géométrique et conduit à une présentation plus naturelle, encore utilisée aujourd'hui, et qui va rapidement être adoptée en Europe. C'est lui également qui introduit la notation dy/dx encore utilisée.



Archimède Oresme

Avant Newton et Leibniz, Descartes s'était intéressé au problème des tangentes à une courbe et Fermat avait introduit une notion assez proche de la dérivée lors d'une recherche d'extremum. En remontant encore plus dans l'Histoire, on peut dire que la méthode d'exhaustion, telle qu'elle a été développée et utilisée par Archimède (i.e. avec toute la rigueur nécessaire, en usant du procédé d'encadrement évitant le recourt aux ε) est réellement la première utilisation des infinitésimaux. Malgré, ou peut-être à cause de, la grande originalité de ses travaux, Archimède n'a été que peu suivi dans le monde grec et n'a pas eu de disciple direct. Les savants arabes commenceront d'ailleurs dès le IX^e siècle à s'intéresser aux procédés infinitésimaux mis en œuvre par le génial Alexandrin.

Notons également que le plus grand progrès théorique réalisé au Moyen-Âge est l'étude quantitative de la variabilité par Nicole Oresme. Éveillé par la diffusion des méthodes infinitésimales des Anciens, et d'Archimède en particulier, nourri par les spéculations scolastiques sur l'infini, le goût des mathématiciens pour les considérations infinitésimales se développe peu à peu. Ils s'appliquent d'abord à imiter le modèle archimédien, puis s'émancipent et essaient de trouver des substituts pour la méthode d'exhaustion, jugée trop lourde.

En 1748, Euler définit les deux nouvelles opérations que sont la différentiation et l'intégration dans *Introductio in analysin infinitorum* ; puis dans *Calculus differentialis* (1755) et *Institutiones calculi integralis* (1770) il essaie de mettre au point les règles d'utilisation des infiniment petits et développe des méthodes d'intégration et de résolution d'équations différentielles.

Enfin, dans la seconde moitié du XIX^e siècle, on s'intéressera aux propriétés des fonctions

dérivées, et de nombreux contre-exemples défiant l'intuition seront introduits, qui conduiront à bien cerner cette délicate notion.

La dérivabilité est elle-aussi une notion topologique et non métrique (même si on sait l'écrire en termes métrique comme ci-dessus), elle ne dépend donc pas de la norme choisie (du moment que celle-ci est compatible avec la topologie choisie...).

3.4 Fonctions de classe C^k

Il est évident que si la dérivée (telle que définie au dessus) existe partout dans Ω , alors on peut à nouveau considérer sa dérivée... et ainsi de suite.

On définit donc les classes $C^1, C^2, \dots, C^k, \dots, C^\infty$ de fonctions 1 fois, 2 fois, ..., k fois continûment dérivables ou même indéfiniment dérivables. Pour $k = 0$, on retombe sur la définition de l'ensemble des fonctions continues. Pour être très clair, une fonction est de classe C^k signifie que toutes ses dérivées jusqu'à l'ordre k sont continues dans Ω . Toutes les fonctions polynomiales sont C^∞ , car à partir d'un certain rang leur dérivée est identiquement nulle.

3.5 Différentielle

Il s'agit de généraliser la formule (3.6) à des applications de \mathbb{R}^n dans \mathbb{R}^p . La variable étant maintenant un vecteur de \mathbb{R}^n , il n'est plus question de diviser par h ... il faut donc modifier la définition pour supprimer le dénominateur. Et cela est très facile : il suffit d'approcher l'accroissement de la fonction par une application linéaire.

Définition 23 — Différentielle. Soient E et F deux espaces normés sur le corps $\mathbb{K} = \mathbb{R}$ ou \mathbb{C} , et U un ouvert de E . Une application $f : U \rightarrow F$ est dite différentiable au point $a \in U$ s'il existe une application linéaire $L \in \mathcal{L}(E, F)$ telle que :

$$f(a + h) - f(a) = L(h) + R(h) \quad (3.7)$$

où le reste $R(h)$ est un $o(\|h\|)$ lorsque h tend vers 0 dans E .

Si elle existe, l'application L est unique et on l'appelle la différentielle de f en a . On la note $L = Df(a)$, ou $f'(a)$ ou $df(a)$ ou $D_a f$ ou df_a , selon les auteurs et les circonstances.

Cela signifie qu'au voisinage de a , l'application f se comporte à peu près comme l'application affine $x \mapsto f(a) + Df(a)(x - a)$, pour laquelle on pourra utiliser les outils de l'algèbre linéaire. Géométriquement, on retrouve l'idée qu'une courbe est, au voisinage d'un point, à peu près confondues avec sa tangente¹ (droite tangente, plan tangent...). C'est pourquoi l'application linéaire $Df(a)$ est également appelée application linéaire tangente à f en a .

Évidemment, dans le cas où $E = \mathbb{R}$ et $F = \mathbb{R}$, tout ce qui a été dit coïncide avec ce que nous avons vu avec la dérivée classique.

3.6 Dérivées partielles

Nous nous intéressons maintenant au cas où une fonction dépend de plusieurs variables, i.e. au cas d'une fonction numérique d'une variable vectorielle. Avec les notation du paragraphe précédent, cela correspond au cas où $E = \mathbb{R}^n$ et $F = \mathbb{R}$.

La dérivée partielle d'une fonction de plusieurs variables est la dérivée par rapport à l'une de ses variables, les autres étant gardées constantes. La dérivée partielle de la fonction f par rapport à

1. tangente, du latin « tangerer » : toucher

la variable x est notée $\frac{\partial f}{\partial x}$ ou $\partial_x f$ ou encore f'_x . De la même manière, on peut noter des dérivées secondes,... k -ièmes par rapport à différentes variables. Par exemple $\frac{\partial^2 f}{\partial x^2} = f'_{xx} = \partial_{xx} f = \partial_x^2 f$, mais également $\frac{\partial^2 f}{\partial x \partial y} = f'_{yx} = \partial_{yx} f$, ou $\frac{\partial^2 f}{\partial y \partial x} = f'_{xy} = \partial_{xy} f$, mais également $\frac{\partial^{i+j+k} f}{\partial x^i \partial y^j \partial z^k} = f'_{kzjyix} = \partial_{kzjyix} f$.

Définition 24 — Dérivée partielle d'ordre 1 en un point. Soient Ω un ouvert de \mathbb{R}^n et f une fonction de n variables :

$$f : \begin{array}{ccc} \Omega & \rightarrow & \mathbb{R} \\ \mathbf{x} = (x_1, \dots, x_n) & \mapsto & f(\mathbf{x}) = f(x_1, \dots, x_n) \end{array} \quad (3.8)$$

On définit la dérivée partielle d'ordre 1 de f au point $\mathbf{a} = (a_1, \dots, a_n) \in \Omega$ par rapport à la i -ème variable x_i par :

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = \lim_{h \rightarrow 0} \frac{f(a_1, \dots, a_{i-1}, a_i + h, a_{i+1}, \dots, a_n) - f(a_1, \dots, a_n)}{h} \quad (3.9)$$

Remarques. Pour en revenir au cadre général qu'est celui de la différentielle, on parle ici des dérivées partielles pour désigner les dérivées dans la direction des vecteurs de base, ou plus exactement des dérivées selon les vecteurs de base.

Nous utiliserons, comme tout le monde, la notation $\frac{\partial f}{\partial x_i}$, bien qu'elle présente quelques défauts : tout d'abord elle est typographiquement lourde, ensuite elle a les apparences trompeuses d'un quotient, et enfin la mention de la variable x_i peut être source de confusions dans les calculs de dérivées de fonctions composées. On devrait donc toujours préférer la notation $\partial_i f$ ou f'_i .

La formulation générale de la différentielle donnée au paragraphe 3.5 permet de retrouver tous les cas :

- le cas « classique » d'une fonction d'une variable réelle à valeur réelle correspond au cas $E = \mathbb{R}$ et $F = \mathbb{R}$;
- le cas d'une fonction numérique d'une fonction vectorielle correspond au cas $E = \mathbb{R}^n$ et $F = \mathbb{R}$ et vient d'être présenté. Dans ce cas, la différentielle $Df(\mathbf{a})$ est la forme linéaire sur \mathbb{R}^n de composantes ${}^T \partial_1 f(\mathbf{a}), \dots, \partial_n f(\mathbf{a})$ dans la base canonique. Si \mathbb{R}^n est muni d'un produit scalaire, alors il existe un unique vecteur, appelé gradient de f en \mathbf{a} et noté $\mathbf{grad} f(\mathbf{a}) \in \mathbb{R}^n$ tel que : $Df(\mathbf{a})h = \mathbf{grad} f(\mathbf{a}) \cdot h$. Pour une définition plus opérationnelle du gradient, on se référera au paragraphe 3.8. Celui-ci a pour composantes les dérivées partielles par rapport à la base considérée ;
- le cas d'une fonction vectorielle d'une variable réelle correspond au cas $E = \mathbb{R}$ et $F = \mathbb{R}^p$. Il est à nouveau possible de diviser par h . La différentiabilité² de f équivaut alors à la dérivabilité de ses composantes f_i .
- le cas d'une fonction vectorielle d'une variable vectorielle correspond au cas le plus général $E = \mathbb{R}^n$ et $F = \mathbb{R}^p$. La différentielle est l'application linéaire définie dans les bases canoniques de \mathbb{R}^n et \mathbb{R}^p par la matrice jacobienne. La ligne i de la matrice jacobienne correspond à la différentielle de la composante f_i de f . La colonne j de la matrice jacobienne correspond à la dérivée de f dans la direction du j^{e} vecteur de la base. Nous y reviendrons de manière plus pragmatique au paragraphe 12.5.2.

Attention : Même si toutes les dérivées partielles $\frac{\partial f}{\partial x_1}(\mathbf{a}), \dots, \frac{\partial f}{\partial x_n}(\mathbf{a})$ existent en un point \mathbf{a} , la fonction peut ne pas être continue en ce point. Si l'on considère à nouveau la fonction f définie sur \mathbb{R}^2 par :

$$f(x, y) = \begin{cases} \frac{xy}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{si } (x, y) = (0, 0) \end{cases} \quad (3.10)$$

on voit qu'elle vérifie $\frac{\partial f}{\partial x}(0, 0) = \frac{\partial f}{\partial y}(0, 0) = 0$ mais elle n'a pas de limite en $(0, 0)$ comme nous l'avons vu avant.

2. On ne confondra pas « différencier » et « différentier »

Si toutes les dérivées partielles (d'ordre 1) existent et sont continues dans un voisinage de \mathbf{a} , alors f est différentiable dans ce voisinage et la différentielle est continue. Dans ce cas, on dit que f est une fonction de classe C^1 sur ce voisinage de \mathbf{a} .

Si toutes les dérivées partielles secondes de f existent et sont continues sur l'ouvert Ω , on dit que f est une fonction de classe $C^2(\Omega)$. L'ordre de dérivation peut alors être changé sans que cela modifie le résultat. C'est le théorème de Schwarz (Il s'agit de Hermann Amandus Schwarz (sans « t »), pas de Laurent Schwartz).

Théorème 9 — Théorème de Schwarz. Si $f \in C^2(\Omega)$, alors :

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \quad (3.11)$$

Définition 25 — Différentielle totale. On appelle différentielle totale de f l'expression :

$$df = \frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_n} dx_n = \sum_i \frac{\partial f}{\partial x_i} dx_i \quad (3.12)$$

3.7 Retour sur les classes C^k pour une fonction de plusieurs variables

Ce paragraphe met l'accent sur le cas multi-dimensionnel des classes C^k (ce qui est le cas avant, mais peut-être moins visiblement). Nous en profitons également pour introduire des notations et notions dont nous nous servirons plus loin.

Définition 26 — Fonctions k fois différentiables. Si Ω est un ouvert de \mathbb{R}^n , on définit l'ensemble des fonctions k fois différentiables dans Ω , à valeurs dans \mathbb{R} dont toutes les dérivées jusqu'à l'ordre k sont continues dans Ω par :

$$C^k(\Omega) = \left\{ f \in C^{k-1}(\Omega), \frac{\partial f}{\partial x_i} \in C^{k-1}(\Omega), i = 1, \dots, n \right\}, k \geq 1 \quad (3.13)$$

Définition 27 — Multi-indice. On appelle multi-indice α un n -uplet d'entiers $\alpha = (\alpha_1, \dots, \alpha_n)$, $\alpha_j \in \mathbb{N}$. Sa longueur est $|\alpha| = \alpha_1 + \dots + \alpha_n$, et on note ∂^α la quantité (l'opération) :

$$\partial^\alpha = \left(\frac{\partial}{\partial x_1} \right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n} \right)^{\alpha_n} \quad (3.14)$$

Compléments sur C^k . $C^k(\Omega)$ est l'espace vectoriel des fonctions $f : \Omega \rightarrow \mathbb{R}$ telles que $\forall \alpha, |\alpha| \leq k$, $x \mapsto \partial^\alpha f(x)$ existe et appartient à $C^0(\Omega)$.

Pour f et g dans $C^k(\Omega)$, on définit :

$$d(f, g) = \sum_{i=1}^{\infty} \frac{1}{2^i} \frac{\sum_{|\alpha| \leq k} \sup_{K_i} |\partial^\alpha f(x) - \partial^\alpha g(x)|}{1 + \sum_{|\alpha| \leq k} \sup_{K_i} |\partial^\alpha f(x) - \partial^\alpha g(x)|} \quad (3.15)$$

qui est une distance sur $C^k(\Omega)$ qui en fait un espace complet.

L'espace $C^\infty(\Omega)$ peut se définir par :

$$C^\infty(\Omega) = \bigcup_{k \in \mathbb{N}} C^k(\Omega) \quad (3.16)$$

qui est également complet pour la même distance.

Cette définition permet de clairement voir les inclusions successives des espaces C^k .

On définit également $C_b^k(\Omega)$ comme le sous-espace vectoriel des éléments de $C^k(\mathbb{R}^n)$ dont toutes les dérivées jusqu'à l'ordre k sont bornées sur Ω .

On définit alors :

$$\|f\|_{C_b^k(\Omega)} = \sum_{|\alpha| \leq k} \sup_{\Omega} |\partial^\alpha f(x)| \quad (3.17)$$

qui est une norme sur $C_b^k(\Omega)$ et en fait un espace de Banach (i.e. normé complet pour cette norme).

Définition 28 On définit, pour $k \in \mathbb{N} \cup \{\infty\}$, l'ensemble $C_c^k(\Omega)$ par : $f \in C_c^k(\Omega)$ si $f \in C^k(\Omega)$ et si f est de support compact inclus dans Ω .

Pour tout ouvert Ω de \mathbb{R}^n , et pour tout $k \in \mathbb{N} \cup \{\infty\}$:

- $C_c^k(\Omega)$ est dense dans $C^k(\Omega)$;
- $C_c^\infty(\Omega)$ est dense dans $C^k(\Omega)$.

3.8 Nabla et comparses

3.8.1 Champs de vecteurs et de scalaires

Définition 29 — Champ de vecteurs. Soit E un espace vectoriel euclidien de dimension n et Ω un ouvert de E . Un champ de vecteurs sur Ω est une application F de Ω dans E , définie par ses n fonctions composantes :

$$F : \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} F_1(x_1, \dots, x_n) \\ \vdots \\ F_n(x_1, \dots, x_n) \end{pmatrix} \quad (3.18)$$

C'est donc une fonction qui à un vecteur fait correspondre un vecteur.

Définition 30 — Champ de scalaires. Un champ de scalaires sur Ω est une application f de Ω dans \mathbb{R} ou \mathbb{C} , i.e. une fonction qui à un vecteur fait correspondre un scalaire.

La dérivée d'un champ scalaire est un champ vectoriel appelé gradient (voir plus bas).

Nabla, noté ∇ , est un pseudo-vecteur servant à noter un opérateur différentiel :

$$\nabla = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix}_{\text{cartésiennes}} = \begin{pmatrix} \frac{\partial}{\partial \rho} \\ \frac{1}{\rho} \frac{\partial}{\partial \varphi} \\ \frac{\partial}{\partial z} \end{pmatrix}_{\text{polaires}} = \begin{pmatrix} \frac{\partial}{\partial r} \\ \frac{1}{r} \frac{\partial}{\partial \theta} \\ \frac{1}{r \sin \theta} \frac{\partial}{\partial \varphi} \end{pmatrix}_{\text{sphériques}} \quad (3.19)$$

3.8.2 Gradient, divergence, rotationnel, Laplacien et D'Alembertien

Les quantités présentées ci-après apparaîtront constamment dans les problèmes physiques.

Soient \mathbf{A} un champ de vecteur et f un champ scalaire, on définit :

- Le gradient :

$$\mathbf{grad} f \equiv \nabla f \quad (3.20)$$

- La divergence :

$$\text{div } \mathbf{A} \equiv \nabla \cdot \mathbf{A} \quad (3.21)$$

— Le rotationnel :

$$\mathbf{rot} \mathbf{A} \equiv \nabla \wedge \mathbf{A} \quad (3.22)$$

— Le Laplacien (scalaire et vectoriel) :

$$\Delta f = \nabla^2 f \quad (3.23)$$

$$\Delta \mathbf{A} = \nabla^2 \mathbf{A} \quad (3.24)$$

— Le D'alembertien : il s'agit plus d'une « contraction d'écriture ». En coordonnées cartésiennes, \square s'écrit :

$$\square = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \Delta \quad (3.25)$$

où c est la vitesse de la propagation considérée (ou la vitesse de la lumière).

En coordonnées cartésiennes, il vient explicitement :

— gradient : $\mathbf{grad} f = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix}$

— divergence : $\operatorname{div} \mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}$

— rotationnel : $\mathbf{rot} \mathbf{A} = \begin{pmatrix} \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \\ \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \\ \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \end{pmatrix}$

— Laplacien scalaire : $\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$

— Laplacien vectoriel : $\Delta \mathbf{A} = \begin{pmatrix} \Delta A_x \\ \Delta A_y \\ \Delta A_z \end{pmatrix}$

De plus :

$$\operatorname{div} \mathbf{grad} f = \nabla \cdot (\nabla f) = \nabla^2 f = \Delta f \quad (3.26)$$

$$\mathbf{rot} \mathbf{grad} f = \nabla \wedge (\nabla f) = \mathbf{0} \quad (3.27)$$

$$\operatorname{div} \mathbf{rot} \mathbf{A} = \nabla \cdot (\nabla \wedge \mathbf{A}) = 0 \quad (3.28)$$

$$\mathbf{rot} \mathbf{rot} \mathbf{A} = \nabla \wedge (\nabla \wedge \mathbf{A}) = \nabla (\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} = \mathbf{grad} \operatorname{div} \mathbf{A} - \Delta \mathbf{A} \quad (3.29)$$

$$\Delta fg = f \Delta g + 2 \nabla f \cdot \nabla g + g \Delta f \quad (3.30)$$

3.8.3 Normale et dérivée normale

Définition 31 — Normale. On appelle normale au domaine Ω un champ de vecteurs $n(x)$ défini sur le bord $\Gamma = \partial\Omega$ de Ω et tel qu'en tout point $x \in \Gamma$ où le bord est régulier, $n(x)$ soit orthogonal au bord et de norme 1.

On appelle normale extérieure une normale qui pointe vers l'extérieur du domaine en tout point.

Définition 32 — Dérivée normale. On appelle dérivée normale d'une fonction régulière u sur Γ , la fonction définie sur les points réguliers de Γ par :

$$\frac{\partial u}{\partial n}(x) = \nabla u(x) \cdot n(x) \quad (3.31)$$

Il s'agit d'un produit scalaire car ∇u est un vecteur, tout comme $n(x)$.

3.8.4 Potentiel d'un champ vectoriel

Le potentiel d'un champ vectoriel est une fonction scalaire ou vectorielle qui, sous certaines conditions relatives au domaine de définition et à la régularité, permet des représentations alternatives de champs aux propriétés particulières. Ainsi :

- Un champ vectoriel irrotationnel (de rotationnel nul) peut être identifié au gradient d'un potentiel scalaire.
- Un champ vectoriel solénoïdal (de divergence nulle) peut être identifié au rotationnel d'un potentiel vecteur.

Dans les deux cas, on dit que le champ d'origine dérive d'un potentiel (allusion entre une fonction et sa primitive).

Ces potentiels permettent non seulement d'appréhender certains champs vectoriels sous un angle complémentaire (pour un traitement parfois plus aisé), mais ils légitiment des abstractions essentielles comme, par exemple en physique, l'énergie potentielle associée à un champ de forces conservatives.

Un champ de vecteurs \mathbf{A} continu et irrotationnel dérive d'un potentiel scalaire U :

$$\mathbf{A} = -\text{grad } U \quad (3.32)$$

Le même résultat se vérifie dans l'espace entier à condition que, vers l'infini, le champ décroisse « assez » rapidement.

Le potentiel scalaire n'est pas unique, il est défini à une constante près.

Un champ de vecteurs \mathbf{A} régulier et de divergence nulle dérive d'un potentiel vectoriel \mathbf{B} :

$$\mathbf{A} = -\text{rot } \mathbf{B} \quad (3.33)$$

Le même résultat se vérifie dans l'espace entier à condition que, vers l'infini, le champ décroisse « assez » rapidement.

Le potentiel vecteur n'est pas unique, il est défini à un « gradient » près (i.e. $\mathbf{B}' = \mathbf{B} + \text{grad } f$ convient également).

3.8.5 Signification « physique »

Un champ de scalaires, c'est un « truc », une application, qui à un vecteur associe un scalaire. Typiquement, on peut penser à la température. Une application $T(x, y, z)$ qui à chaque point de l'espace de coordonnées (x, y, z) associe sa température T est un champ de scalaire.

Un champ de vecteurs, c'est un « truc », une application, qui à un vecteur associe un autre vecteur. Typiquement, on peut penser à la vitesse. Une application $\mathbf{V}(x, y, z)$ qui à chaque point de l'espace de coordonnées (x, y, z) associe son vecteur vitesse \mathbf{V} est un champ de vecteurs.

Si f est une fonction de classe C^1 , alors le gradient de f au point \mathbf{a} , quand il est non nul, s'interprète comme la direction selon laquelle f varie le plus vite, i.e. la ligne de plus grande pente.

La divergence d'un champ de vecteurs mesure comment son courant déforme un volume (on devrait dire comment son flot déforme une forme volume au sens de la géométrie différentielle). L'idée est un peu la suivante : imaginons un écoulement de fluide et intéressons-nous au courant en fonction de la profondeur. Lorsque l'on est un peu en dessous de la surface et loin du sol, on peut imaginer que le courant est à peu près constant, donc une section (ou un volume) de fluide perpendiculaire à l'écoulement se retrouve, un instant plus tard, un peu plus loin, mais toujours perpendiculaire au sol, i.e. la section n'a pas varié. Si l'on est proche du fond, on peut penser que le courant est plus faible très près du sol, par exemple à cause du frottement. Si l'on considère à nouveau une section de fluide perpendiculaire au sol, alors un instant plus tard, elle n'est plus perpendiculaire au sol : près du sol, les particules ont effectué un petit déplacement, celles plus éloignées ont beaucoup plus avancé. La section n'a ni la même forme, ni la même

longueur qu'à l'instant précédent. La divergence mesure justement ce type d'écart. D'une manière plus générale, la divergence traduit la conservation (si elle est nulle) ou non d'une grandeur physique en un point : cela mesure donc, en chaque point si une grandeur (par exemple le volume comme avant) est conservative ou non.

Le rotationnel exprime la tendance qu'ont les lignes de champ d'un champ vectoriel à tourner autour d'un point : sa circulation locale sur un petit lacet entourant ce point est non nulle quand son rotationnel ne l'est pas (nous n'avons pas défini la notion de lacet, nous compterons sur l'imagination du lecteur). On rappelle que les lignes de champ sont les lignes qui, en première approche, représentent le chemin que l'on suivrait en partant d'un point. Ce sont en fait les lignes orthogonales aux équipotentielles, ou surfaces de niveau, du champ.

Concernant le Laplacien scalaire, la quantité Δf est une mesure de la différence entre la valeur de f en un point quelconque P et la valeur moyenne \bar{f} au voisinage du point P .

Le Laplacien scalaire d'une fonction peut aussi être interprété comme la courbure moyenne locale de la fonction, que l'on visualise aisément pour une fonction à une seule variable $f(x)$. La dérivée seconde (ou courbure) f'' représente la déviation locale de la moyenne par rapport à la valeur au point considéré.

Les notions de rotationnel, gradient, Laplacien... interagissent. Par exemple en mécanique des fluides, le rotationnel de la vitesse décrit une rotation de la particule fluide. Si l'écoulement est irrotationnel (son rotationnel est nul en tout point), alors le vecteur vitesse est le gradient du potentiel (on dit alors que les vitesses dérivent d'un potentiel). Si le fluide peut être considéré comme incompressible, la divergence de ce vecteur s'annule. Le laplacien du potentiel est donc nul : il s'agit d'un potentiel harmonique qui satisfait l'équation de Laplace.

3.9 Quelques théorèmes sur les intégrales

Les relations ci-après permettent de passer d'intégrales sur un domaine à des intégrales sur le bord de ce domaine.

Théorème 10 — Théorème du gradient. Ce théorème met en relation l'intégrale de volume du gradient d'un champ scalaire et l'intégrale de surface de ce même champ :

$$\int_{\Omega} \nabla f = \int_{\Gamma} f n(x) \quad (3.34)$$

où Γ est le bord du domaine Ω et f un champ scalaire (i.e. une fonction régulière).

Théorème 11 — Théorème du rotationnel. Ce théorème met en relation l'intégrale de volume du rotationnel d'un champ vectoriel et l'intégrale de surface du même champ :

$$\int_{\Omega} \text{rot } \mathbf{A} = - \int_{\Gamma} \mathbf{A} \wedge n(x) \quad (3.35)$$

où Γ est la frontière de Ω , \wedge est le produit vectoriel et $n(x)$ est la normale dirigé vers l'extérieur.

Une autre identité remarquable met en relation l'intégrale de surface du rotationnel d'un champ vectoriel et l'intégrale curviligne (ou circulation) du même champ sur la frontière. Elle découle du théorème de Green qui, pour une surface S (généralement non fermée) de frontière C , implique :

$$\iint_S \text{rot } \mathbf{A} \cdot d\mathbf{s} = \oint_C \mathbf{A} \cdot d\mathbf{l} \quad (3.36)$$

Si S est fermée, C est vide (ou réduit à un point) et le membre de droite est nul.

L'orientation de la surface et celle de la courbe frontière sont liées puisque le changement d'une orientation modifie le signe de l'intégrale correspondante. En fait, la relation est satisfaite lorsque

ces orientations sont telles que, sur un point frontière, le vecteur tangent à la surface $d\vec{s} \wedge d\vec{l}$ est orienté en direction de la surface.

Théorème 12 — Théorème de Green–Riemann. Ce théorème donne la relation entre une intégrale curviligne autour d’une courbe simple fermée C et l’intégrale double sur la région du plan S délimitée par C .

Soit C , une courbe plane simple, positivement orientée et C^1 par morceaux, S le domaine compact lisse du plan délimité par C et $Pdx + Qdy$ une 1-forme différentielle sur \mathbb{R}^2 . Si P et Q ont des dérivées partielles continues sur une région ouverte incluant S , alors :

$$\int_C P dx + Q dy = \iint_S \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy \quad (3.37)$$

Il existe une autre façon de noter ce théorème. On se place sur un domaine compact lisse du plan Ω , de bord $\partial\Omega$, en notant la forme différentielle ω . Alors la dérivée extérieure de ω s’écrit :

$$d\omega = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx \wedge dy \quad (3.38)$$

On peut alors résumer le théorème de Green par la formule :

$$\oint_{\partial\Omega} \omega = \iint_{\Omega} d\omega \quad (3.39)$$

(Le cercle sur l’intégrale précise que le bord décrit une courbe fermée).

Théorème 13 — Théorème de Green–Ostrogradsky. Soit Ω un domaine de \mathbb{R}^2 ou \mathbb{R}^3 de frontière Γ et \mathbf{A} un champ de vecteurs :

$$\int_{\Omega} \operatorname{div} \mathbf{A} = \int_{\Gamma} \mathbf{A} \cdot \mathbf{n} \quad (3.40)$$

Ce théorème prend aussi le nom de théorème du flux-divergence ou plus simplement formule de la divergence. « Ostrogradsky » peut s’écrire avec un « y » ou un « i » à la fin. Du théorème de Green–Ostrogradsky, on peut déduire la formule de Green :

Théorème 14 — Formule de Green. Soient Ω un domaine de \mathbb{R}^2 ou \mathbb{R}^3 de frontière Γ et u et v deux fonctions régulières :

$$\int_{\Omega} (\Delta u)v = - \int_{\Omega} \nabla u \cdot \nabla v + \int_{\Gamma} \frac{\partial u}{\partial n} v \quad (3.41)$$

On en déduit certaines formules utiles du calcul vectoriel. Soit Ω un domaine de \mathbb{R}^2 ou \mathbb{R}^3 de frontière Γ , \mathbf{A} et \mathbf{B} des champ de vecteurs, f et g des fonctions régulières, et $\mathbf{n}(x)$ la normale extérieure au point considéré :

$$\int_{\Omega} (\mathbf{A} \cdot \nabla f + f (\nabla \cdot \mathbf{A})) = \int_{\Gamma} f \mathbf{A} \cdot \mathbf{n}(x) \quad (3.42)$$

$$\int_{\Omega} (\mathbf{B} \cdot (\nabla \wedge \mathbf{A}) - \mathbf{A} \cdot (\nabla \wedge \mathbf{B})) = \int_{\Gamma} (\mathbf{A} \wedge \mathbf{B}) \cdot \mathbf{n}(x) \quad (3.43)$$

$$\int_{\Omega} (f \nabla^2 g + \nabla f \cdot \nabla g) = \int_{\Gamma} f \nabla g \cdot \mathbf{n}(x) \quad (3.44)$$

Ces formules sont exploitées pour obtenir des formulations faibles associées à des problèmes aux dérivées partielles.

Chapitre 4

Espaces de Lebesgue

Résumé — Les espaces L^p sont indispensables à la définition des espaces de Sobolev après lesquels nous courons depuis quelques chapitres.

La mesure de Lebesgue a été introduite, les notions d'application et de continuité ont été rappelées... nous en avons plus qu'il ne nous en faut pour définir de tels espaces.

Comme promis au paragraphe 1.3, nous fournissons maintenant quelques compléments sur l'intégration. Nous avons mentionné la mise en défaut de l'intégrale de Riemann dans le cas de l'indicatrice des rationnels. Ce contre-exemple a été historiquement fourni par Dirichlet. Regardons toutefois d'un peu plus près d'où tout cela provient.

Histoire

Dans le chapitre VI « Développement d'une fonction arbitraire en séries trigonométriques » de sa *Théorie analytique*, Fourier considère une fonction f définie dans $] -\pi/2, +\pi/2[$ dont le développement en série trigonométrique est de la forme :

$$f(x) = a_1 \sin x + a_2 \sin 2x + \dots + a_k \sin kx + \dots \quad (4.1)$$



Fourier

Le problème est de calculer les coefficients a_k dans le cas même, dit-il, où « la fonction $f(x)$ représente une suite de valeurs ou ordonnées dont chacune est arbitraire... On ne suppose point que ces coordonnées soient assujetties à une loi commune ; elles se succèdent d'une manière quelconque et chacune d'elles est donnée comme le serait une seule quantité ».

On notera au passage que la fonction considérée par Fourier n'est définie que par la donnée de ses points (si en plus ils sont en nombre fini, cela représente alors un échantillonnage). Fourier conduit son calcul selon une voie nouvelle : en multipliant l'expression précédente par $\sin kx$ et en intégrant terme à terme la série, mais sans justification, il obtient :

$$a_k = \frac{2}{\pi} \int_0^\pi f(x) \sin kx dx \quad (4.2)$$

Fourier observe que dans tous les cas envisagés, ces intégrales ont un sens et en conclut que toute fonction d'une variable peut-être représentée par une série trigonométrique.

Même manquant de rigueur, Fourier, aboutit à un résultat juste pour les fonctions qu'il considère. Or les fonctions considérées en l'occurrence « vont bien » pour l'intégrale de Riemann telle qu'elle est définie.

Définition 33 — Intégrale de Riemann. Soit f une fonction réelle définie sur un intervalle $[a; b]$. On considère une suite (x_i) , $0 \leq i \leq n$, de subdivisions de cet intervalle : $a = x_0 < x_1 < \dots < x_{n-1} < x_n = b$. Notons $\delta_i = x_i - x_{i-1}$ et $S = \delta_1 f(a + \varepsilon_1 \delta_1) + \delta_2 f(x_1 + \varepsilon_2 \delta_2) + \dots + \delta_n f(x_{n-1} + \varepsilon_n \delta_n)$, où $0 \leq \delta_i \leq 1$.

Si la somme S a la propriété, de quelque manière que les δ et les ε puissent être choisis, de s'approcher indéfiniment d'une limite fixe A , quand les δ tendent vers zéro, cette limite s'appelle la valeur de l'intégrale définie $\int_a^b f(x)dx$.

En termes modernes, on dira que pour qu'une fonction bornée soit Riemann-intégrable, il faut et il suffit que l'ensemble des points de discontinuité de f soit de mesure nulle.

Le cadre classique le plus simple pour définir une intégrale est celui des fonctions en escalier sur un intervalle $[a, b]$, ou sa complétion pour la topologie de la convergence uniforme, l'espace des fonctions réglées (admettant une limite finie à droite et à gauche). La théorie de Riemann permet d'atteindre une plus grande généralité. Cependant, Riemann lui-même a conscience que l'intégrabilité « au sens de Riemann » impose encore des conditions relativement fortes : il démontre qu'une fonction $f : [a, b] \rightarrow \mathbb{R}$ est intégrable si et seulement si, pour tout $\alpha > 0$ donné, on peut choisir une décomposition de $[a, b]$ en sous-intervalles suffisamment fins pour que la somme des longueurs des sous-intervalles sur lesquels l'oscillation de la fonction dépasse α soit arbitrairement petite.

Les conditions énoncées au paragraphe précédent peuvent sembler faibles, mais, dès la fin de son mémoire de 1829 *Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données*, Dirichlet donne l'exemple, d'une nature toute nouvelle, discontinue en tous ces points : la fonction $f(x)$ qui vaut une constante c si x est un rationnel et qui vaut une autre constante d si x est irrationnel (on appelle une telle fonction, « fonction de Dirichlet », bien que celui-ci l'ai toujours exhibée comme un monstre, existant, mais non représentatif de ce que devrait être une fonction. Lorsque $c = 1$ et $d = 0$, on parle alors plutôt de l'indicatrice de \mathbb{Q}). On voit alors directement que la Riemann-intégrabilité est mise à mal. Le travail consistera alors à définir précisément ce que l'on peut négliger, i.e. les parties de mesure nulle.



Riemann

Dirichlet

Lebesgue

Dans la théorie de Lebesgue, on va élargir la classe des fonctions intégrables (par exemple, toute fonction bornée « raisonnable », disons qui peut être décrite par un énoncé mathématique, est Lebesgue-intégrable). Pour cela, Lebesgue va commencer par définir un ensemble mesurable : c'est un ensemble dont la mesure extérieure (i.e. la borne inférieure de la mesure des ouverts le contenant) est égale à sa mesure intérieure (i.e. la borne supérieure de la mesure des fermés qu'il contient) : est-il besoin de rappeler la différence entre borne inférieure et minimum et entre borne supérieure et maximum ? La borne supérieure (ou le supremum) d'une partie d'un ensemble partiellement ordonné est le plus petit de ses majorants. Une telle borne n'existe pas toujours, mais si elle existe alors elle est unique. Elle n'appartient pas nécessairement à la partie considérée. Duallement, la borne inférieure (ou l'infimum) d'une partie est le plus grand de ses minorants...

L'intégrale de Lebesgue peut être définie de manière géométrique : pour une fonction positive f définie sur $[a; b]$, elle est égale à la mesure de dimension deux de l'ensemble $\{(x, y) \in \mathbb{R}^2 / a \leq y \leq f(x), a \leq x \leq b\}$ quand cette mesure existe. D'une manière analytique, cette définition de l'intégrale de Lebesgue devient :

Définition 34 — Intégrale de Lebesgue. soit f une fonction réelle, bornée, définie sur $[a; b]$. Supposons $m \leq f(x) \leq M$ pour $x \in [a; b]$. Pour tous ξ, η tels que $\xi \leq \eta$, on définit : $V_{\xi, \eta} = \{x \in [a; b] / \xi \leq f(x) \leq \eta\}$. Si pour tous ξ, η , $V_{\xi, \eta}$ est mesurable, alors f est mesurable. En considérant la partition $m = \xi_1 < \xi_2 < \dots < \xi_n < \xi_{n+1} = M$ et $V_i = \{x / \xi_i \leq f(x) \leq \xi_{i+1}\}$, on définit les sommes

$$\sum_{i=1}^n \xi_i m(V_i) \quad \text{et} \quad \sum_{i=1}^n \xi_{i+1} m(V_i) \quad (4.3)$$

où $m(V)$ est la mesure de l'ensemble V . L'intégrale de Lebesgue est la limite commune de ces deux sommes, dont on peut montrer qu'elle existe lorsque f est mesurable.

Voici par quelle image Lebesgue expliquait la nature de son intégrale :

« Je dois payer une certaine somme ; je fouille dans mes poches et j'en sors des pièces et des billets de différentes valeurs. Je les verses à mon créancier dans l'ordre où elles se présentent jusqu'à atteindre le total de ma dette. C'est l'intégrale de Riemann. Mais je peux opérer autrement. Ayant sorti tout mon argent, je réunis les billets de même valeur, les pièces semblables et j'effectue le paiement en donnant ensemble les signes monétaires de même valeur. C'est mon intégrale. »

Poursuivons l'interprétation géométrique à laquelle nous invite Lebesgue dans la comparaison des intégrales selon Riemann et Lebesgue. Comme dans le cas de l'intégrale de Riemann, il s'agit d'approcher l'aire sous le graphe de la fonction par une union de rectangles. Mais ces rectangles sont définis de manière différente. Dans le cas de Riemann, on s'intéresse aux variations de la fonction sur son domaine de définition : une base étant donnée (l'axe des abscisses en général), on définit le rectangle comme l'ensemble des points au-dessus de cette base, qui sont situés en-dessous de la courbe. Au contraire, dans le cas de Lebesgue, on définit le rectangle en fonction des valeurs de la fonction, sans jamais s'intéresser trop à l'espace de départ. Ce n'est donc pas la base du rectangle que l'on se donnera au départ, mais une variation dans les valeurs atteintes (côté « vertical »). Évidemment, on doit admettre que plusieurs rectangles partagent un même côté vertical. C'est ici que l'intégrale de Lebesgue va gagner toute sa complexité : alors que dans l'intégrale de Riemann, une brique élémentaire est un simple rectangle, dans celle de Lebesgue, il pourra s'agir de plusieurs ou même d'une infinité de rectangles.

L'exemple de la fonction indicatrice de \mathbb{Q} est révélateur : bien que discontinue partout, cette fonction est très facile à décrire en fonction de ses valeurs. Dans la théorie de Riemann, on tenterait vainement de découper le segment $[0, 1]$ en tout petits intervalles où cette fonction ne varie guère ; dans celle de Lebesgue, on partage $[0, 1]$ en seulement deux morceaux qui sont assez complexes (totalement discontinus) mais sur chacun desquels la fonction est effectivement constante.

Ainsi, l'intégration de Lebesgue prolonge bien celle de Riemann, et, en se « débarrassant » de la référence à l'espace de base, peut être adaptée à un cadre extrêmement général, sans qu'aucune hypothèse topologique ne soit faite sur l'espace d'intégration ; et l'expérience montre que l'on peut construire ainsi des théories maniables.

Mais attention à ne pas aller trop vite en besogne : la fonction indicatrice de l'ensemble de Cantor est intégrable au sens de Riemann bien que non réglée, et bien qu'ayant une infinité non dénombrable de points de discontinuité.

La théorie de Lebesgue éclaire bien des difficultés des discussions du XIX^e siècle (notamment sur les propriétés de différentiabilité des fonctions continues) et fournit un cadre général simplifié à de nombreux théorèmes alors que la théorie de Riemann multiplie les hypothèses et les conditions restrictives.

De plus, toute cette évolution s'est accompagnée de l'idée qu'on doit manipuler les fonctions comme des objets mathématiques en soi, des points de nouveaux espaces, les *espaces fonctionnels*. Avec l'extension du langage ensembliste, il devient naturel d'employer un langage géométrique à propos de ces espaces.

Malgré tout ce qui vient d'être dit, rendons quand même hommage à cette belle (et malgré tout puissante) intégrale de Riemann, qui a de plus le bon goût d'être souvent très bien comprise des élèves.

L'intégrale de Riemann nous apprend à nous poser la question « où se passent les choses importantes ? », question qui n'a pas de sens pour l'intégrale de Lebesgue qui ne connaît pas la question « où ? », mais seulement « combien souvent ? ».

De plus, les techniques riemanniennes consistant à localiser les problèmes et à découper l'intégrale sur plusieurs intervalles restent des outils indispensables pour aborder bien des problèmes.



Kurzweil

Henstock

Dans les années 50 a été mise au point l'intégrale de Kurzweil-Henstock (ou KH-intégrale, ou intégrale de jauge) à partir de celle de Riemann. On retravaille un peu les sous-intervalles et les points de calcul de la fonction : chaque point est appelé « marque » et l'on parle de « subdivision marquée » (pour plus de détails, il suffit faire une recherche avec « lemme de Cousin ». Ce lemme est dû à Pierre Cousin qui fut un étudiant de Henri Poincaré).

Sans faire appel à la théorie de la mesure, on dispose alors d'une intégrale aussi puissance que celle de Lebesgue (mais moins générale, car définie uniquement sur \mathbb{R}). Et l'on dispose même d'un très beau résultat : toute fonction dérivée est intégrable, et il n'est pas nécessaire d'introduire la notion d'intégrale impropre. L'intégrale de Lebesgue sur \mathbb{R} peut alors être introduite comme un cas particulier de cette intégrale.

Pour puissante qu'elle soit, la théorie de Lebesgue est impuissante à traiter la « semi-convergence » des intégrales, i.e. les situations où une fonction se trouve être intégrable du fait de compensations entre valeurs positives et négatives, alors que sa valeur absolue n'est pas intégrable. D'autres théories sont plus habiles à tirer parti des compensations : ainsi les intégrales de Denjoy ou de Perron. Pourquoi alors n'enseigne-t-on pas ces théories alternatives plutôt que celle de Lebesgue ? D'une part, parce que dans l'immense majorité des applications, le mauvais traitement des intégrales semi-convergentes s'avère sans gravité (il s'agit en fait de situations relativement exceptionnelles) ; d'autre part, parce que la théorie de Lebesgue dispose d'une grande souplesse qui lui permet de se généraliser de manière abstraite, notamment par son insensibilité à la topologie, que nous avons déjà mentionnée.

4.1 Présentation des espaces de Lebesgue L^p

Définition 35 — Ensemble des fonctions mesurables. Soit (E, \mathcal{A}, μ) un espace mesuré. Si $p \in [1; +\infty[$, on note $\mathcal{L}^p = \mathcal{L}^p(E, \mathcal{A}, \mu)$ l'ensemble de toutes les fonctions mesurables sur (E, \mathcal{A}) , à valeurs dans \mathbb{R} , telles que la fonction $|f|^p$ soit μ -intégrable.

Si $f \in \mathcal{L}^p$, on pose :

$$\|f\|_p = (|f|^p d\mu)^{1/p} \quad (4.4)$$

Théorème 15 Chaque espace \mathcal{L}^p est un espace vectoriel.

Définition 36 — Espace L^p . Soit (E, \mathcal{A}, μ) un espace mesuré. Si $p \in [1; +\infty[$, on note $L^p = L^p(E, \mathcal{A}, \mu)$ l'ensemble des classes d'équivalence des fonctions de \mathcal{L}^p pour la relation d'équivalence « égalité μ -presque partout ».

Si $f \in L^p$, on note $\|f\|_p$ la valeur commune des $\|g\|_p$ pour toutes les fonctions g appartenant à la classe de f .

De manière équivalente, on peut également définir L^p comme étant le quotient de \mathcal{L}^p par la relation d'équivalence « égalité μ -presque partout ». Il est clair que chaque espace L^p est une classe d'équivalence obtenue en identifiant les fonctions qui ne diffèrent que sur un ensemble négligeable. Ainsi, si $f \in L^p$, on appelle « représentant » de f , toute fonction mesurable $g \in \mathcal{L}^p$ qui appartient à la classe d'équivalence de f .

Comme, par construction, deux représentants d'une même classe auront la même intégrale, on appelle (abusivement, mais sans confusion) tout $f \in L^p$ une fonction de L^p et non pas une classe d'équivalence ou un représentant de la classe d'équivalence. En d'autres termes, on identifie une classe à l'un quelconque de ses représentants.

En toute généralité, on aura bien remarqué que les définitions précédentes sont liées à la mesure μ utilisée. Si l'on considère simultanément deux mesures μ et ν , les classes d'équivalences ne sont pas les mêmes respectivement à chacune de ces mesures, et l'identification d'une classe à l'un quelconque de ses représentants ne peut plus se faire.

Remarque. Sans l'identification des fonctions égales presque partout, les fonctions $\|\cdot\|_{L^p}$ ne sont que des semi-normes. Cette identification est sans conséquence pour la plupart des problèmes concrets, mais parfois désastreuse quand on veut étudier des propriétés fines de fonctions mesurables, par exemple si l'on veut étudier la dimension de Hausdorff d'un ensemble de points de discontinuité de f ... De tels problèmes ne relèvent pas de l'analyse fonctionnelle.

Théorème 16 Chaque espace L^p est un espace vectoriel.

Un espace L^p est un espace vectoriel (des classes) de fonctions dont la puissance d'exposant p est intégrable au sens de Lebesgue, où p est un nombre réel strictement positif. Le passage à la limite de l'exposant aboutit à la construction des espaces L^∞ des fonctions bornées.

Théorème 17 Chaque espace L^p est un espace de Banach lorsque son exposant p est ≥ 1 .

Théorème 18 Lorsque $0 < p < 1$, l'intégrale définit une quasi-norme qui en fait un espace complet.

Définition 37 — Exposants conjugués. Soient p et $q \in [1, +\infty]$. On dit que p et q sont des exposants conjugués si :

$$\frac{1}{p} + \frac{1}{q} = 1 \quad (4.5)$$

Théorème 19 Soient p et q des exposants conjugués, alors il existe une dualité entre les espaces d'exposants p et q .

Il suffit en effet de considérer, à $g \in L^q$ fixé (de mesure μ), l'application $\varphi_g : f \mapsto \int f g d\mu$. Cette dernière définit une forme linéaire continue sur L^p .

4.2 Construction de L^p

On considère Ω un ouvert de \mathbb{R}^n . Les fonctions f seront considérées de Ω dans \mathbb{R} ou \mathbb{C} .

On appelle $L^p = L_M^p(\Omega, \mathbb{C})$ pour $p < \infty$ l'espace des fonctions f mesurables, de $\Omega \rightarrow \mathbb{C}$, telles que $\mu(|f|^p) < \infty$ où μ est une mesure sur Ω .

Comme une fonction s'annulant presque partout est d'intégrale nulle, on peut définir l'espace $L^p(X, \mathcal{A}, \mu)$ comme le quotient de l'espace des fonctions p intégrables $L^p(X, \mathcal{A}, \mu)$ par le sous-espace vectoriel des fonctions presque nulles.

Ce quotient identifie donc les fonctions qui sont presque partout égales, autrement dit qui ne diffèrent que sur un ensemble de mesure nulle.

Définition 38 — Norme L^p . Pour $p < \infty$, on appelle norme L^p et on note $\|\cdot\|_p$, l'application définie par :

$$\|f\|_p = \left(\int_{\Omega} |f|^p \right)^{1/p} \quad (4.6)$$

— Sur \mathbb{R} , cela donne :

$$\|f\|_p = \left(\int_a^b |f(t)|^p dt \right)^{1/p} \quad (4.7)$$

— Sur un espace mesuré (X, \mathcal{A}, μ) et à valeurs réelles ou complexes :

$$\|f\|_p = \left(\int_X |f|^p d\mu \right)^{1/p} \quad (4.8)$$

Théorème 20 Si l'espace X est fini et est muni d'une mesure finie $\mu(X) < \infty$, alors tous les espaces L^p (resp. \mathcal{L}^p), pour $1 \leq p \leq \infty$ sont les mêmes.

Définition 39 — Espace des suites. Soit $X = \mathbb{N}$ muni de sa tribu $\mathcal{A} = \mathcal{P}(X)$ et de sa mesure de comptage μ . On note ℓ^p l'espace $L^p(X, \mathcal{A}, \mu) = \mathcal{L}^p(X, \mathcal{A}, \mu)$. Cet espace est l'espace des suites $(u_n)_{n \in \mathbb{N}}$ telles que :

$$\begin{cases} \text{Si } 1 \leq p < \infty : & \sum_n |u_n|^p < \infty, & \text{et } \|u_n\|_p = \left(\sum_n |u_n|^p \right)^{1/p} \\ \text{Si } p = \infty : & \sup_n |u_n|^p < \infty, & \text{et } \|u_n\|_\infty = \sup_n |u_n| \end{cases} \quad (4.9)$$

4.3 Espace L^0

L'espace L^0 est l'ensemble des fonctions mesurables. L^0 est l'espace obtenu en quotientant \mathcal{L}^0 par les fonctions nulles.

Soit φ une fonction mesurable strictement positive et μ -intégrable, alors :

$$\rho(f, g) = \int \frac{\|f - g\|}{1 + \|f - g\|} \varphi d\mu \quad (4.10)$$

définit une distance sur L^0 qui redonne la topologie de la convergence en mesure.

Théorème 21 Muni de cette distance, l'espace L^0 est complet.

Rappel : la notion de convergence (de suite) est une propriété topologique et non métrique (l'écriture avec les ε n'est que la traduction métrique de l'écriture topologique avec les boules).

4.4 Espace L^∞ et dualité avec L^1

Dans le cas où l'exposant est infini, on procède de la même manière.

L'espace $\mathcal{L}^\infty(X, \mathcal{A}, \mu)$ est défini comme l'espace vectoriel des fonctions μ -essentiellement bornées (i.e. des fonctions bornées presque partout).

L'espace $L^\infty(X, \mathcal{A}, \mu)$ est l'espace vectoriel quotient de $\mathcal{L}^\infty(X, \mathcal{A}, \mu)$ par la relation d'équivalence « $f \sim g$ » ssi « f et g sont égales presque partout ».

Théorème 22 L'espace dual de L^1 est L^∞ mais l'espace dual de L^∞ contient strictement L^1 , et d'ailleurs L^1 n'est le dual d'aucun espace de Banach.

4.5 Espace L^2

Par définition, si Ω est un ouvert donné de \mathbb{R}^n , $L^2(\Omega)$ est l'espace des fonctions (réelles ou complexes) qui sont de carré intégrable au sens de l'intégrale de Lebesgue.

Théorème 23 L'espace L^2 est un espace de Hilbert lorsqu'il est muni du produit scalaire :

$$(f, g) = \int_{\Omega} f \bar{g} \, d\mu \quad (4.11)$$

La formule des exposants conjugués conduit dans le cas $p = 2$, à $q = 2$, i.e. que L^2 s'identifie à son dual.

Si l'on reprend la remarque précédente sur la dualité, cela devient dans L^2 (et dans tout espace de Hilbert H) : en associant à tout $v \in H$ l'application $\varphi_v(u) = \langle u, v \rangle$, on peut identifier l'espace de Hilbert H à son dual H' .

Comme mentionné, l'intérêt d'avoir un espace de Hilbert est de disposer d'un produit scalaire, donc de pouvoir décomposer un vecteur. Ce que l'on retrouve dans le théorème 3, que nous réécrivons ci-dessous, et vrai dans L^2 en tant qu'espace de Hilbert :

Théorème 24 — Théorème de projection dans un espace de Hilbert. Tout vecteur u d'un espace de Hilbert H se décompose de manière unique en une somme $u = v + w$ avec $v \in K$ et $w \in K^\perp$, et les sous-espaces K et K^\perp sont supplémentaires dans H .

On rappelle que :

Définition 40 — Orthogonal d'une partie. K étant une partie de H , on note K^\perp l'ensemble des vecteurs $u \in H$ qui sont orthogonaux à tous les vecteurs de K .

Théorème 25 L'orthogonal K^\perp de toute partie de K d'un espace de Hilbert H est un sous-espace vectoriel fermé de H , donc lui-même un espace de Hilbert (théorème 1).

On a $(K^\perp)^\perp = K$.

4.6 Compléments et retour sur les fonctions continues et différentiables

Commençons par quelques exemples ultra classiques :

- La fonction $1/\sqrt{x}$ est L^1 mais pas L^2 .
- La fonction $1/|x|$ est L^2 mais pas L^1 .
- La fonction $e^{-|x|}/\sqrt{|x|}$ est L^1 mais pas L^2 .

Poursuivons par quelques inégalités bien utiles.

Théorème 26 — Inégalité de Minkowski. Soient $p \in [1, \infty]$, et f et g dans L^p , alors on a :

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p \quad (4.12)$$

Théorème 27 — Inégalité de Hölder. Soient p, q et r des nombres de $[1, \infty]$ vérifiant ;

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{r} \quad (4.13)$$

(avec la convention $1/\infty = 0$). Si $f \in L^p$ et $g \in L^q$, alors le produit $fg \in L^r$, et on a :

$$\|fg\|_r \leq \|f\|_p \|g\|_q \quad (4.14)$$

L'inégalité de Hölder conduit au théorème suivant :

Théorème 28 Si μ est une mesure finie et si $1 \leq p \leq q \leq \infty$, on a :

$$L^q \subset L^p \quad (4.15)$$

Terminons ce cours chapitre en revenant un peu aux fonctions continues et C^k , dont nous avons déjà parlé au paragraphe 3.7. Nous commençons d'ailleurs par rappeler certaines définitions déjà données.

Définition 41 Soit Ω un espace topologique. On définit $C_b(\Omega)$ comme l'espace des fonctions continues bornées de Ω dans \mathbb{R} ; $C_0(\Omega)$ comme l'espace des fonctions continues sur Ω , tendant vers 0 à l'infini ; $C_c(\Omega)$ comme l'espace des fonctions continues à support compact dans Ω . Ces trois espaces, munis de la norme sup :

$$\|f\|_\infty = \sup_{\Omega} |f| \quad (4.16)$$

sont des espaces vectoriels topologiques ; les espaces $C_b(\Omega)$ et $C_0(\Omega)$ sont des espaces de Banach, alors que $C_c(\Omega)$ ne l'est pas en général.

Si $\Omega = \mathbb{R}^n$, alors $C_0(\Omega)$ est l'adhérence de $C_c(\Omega)$.

Définition 42 Soient Ω un ouvert de \mathbb{R}^n , et $k \in \mathbb{N}$, $k \geq 1$. On définit l'espace $C^k(\Omega)$ comme l'espace des fonctions k fois dérivables dans Ω , dont toutes les dérivées sont bornées jusqu'à l'ordre k . C'est un espace de Banach quand on le munit de la norme :

$$\|f\|_{C^k} = \|f\|_\infty + \sum_{|\alpha|=k} \|\nabla^\alpha f\|_\infty \quad (4.17)$$

On note parfois cet espace $C^k(\overline{\Omega})$ pour insister sur le fait que l'on prend le supremum jusqu'au bord. Une notation plus appropriée serait sans doute $C_b^k(\Omega)$.

Définition 43 Soit Ω un ouvert de \mathbb{R}^n . L'intersection des espaces $C_c(\Omega)$ et $C^k(\Omega)$, pour tout k , est appelée espace des fonctions C^∞ à support compact ; on la note $C_c^\infty(\Omega)$ ou $\mathcal{D}(\Omega)$. Nous en reparlerons au chapitre suivant...

Puisque $C_c(\Omega)$ n'est déjà pas un espace de Banach, il semble inutile de tenter de normer $\mathcal{D}(\Omega)$.

On dispose des résultats de densité suivants :

- L'espace $C_c^0(\Omega)$ des fonctions continues à support compact est dense dans $L^p(\Omega)$ pour $1 \leq p < \infty$ (mais n'est pas dense dans $L^\infty(\Omega)$)
Dans le cas $p = 1$, cela se traduit par : $\forall f \in L^1(\Omega), \forall \varepsilon > 0, \exists g \in C_c^0(\Omega)$, tel que $\|f - g\|_{L^1(\Omega)} \leq \varepsilon$
- L'espace $C_c^\infty(\Omega)$ des fonctions infiniment dérivables à support compact est dense dans $L^p(\Omega)$ pour $1 \leq p < \infty$ (mais n'est pas dense dans $L^\infty(\Omega)$)
- L'espace $C_c^\infty(\Omega)$ est dense dans le sous espace de $L^\infty(\Omega)$ des fonctions bornées qui tendent vers 0 en l'infini.
- L'ensemble des fonctions continues à support compact de $L^p(\mathbb{R}^n)$ est dense dans $L^p(\mathbb{R}^n)$, pour $p \neq \infty$.

Les arguments de densité ont une importance pratique dans certaines démonstrations : si l'on doit démontrer qu'une propriété est vérifiée sur $L^p(\Omega)$, alors on peut commencer par montrer que cette propriété est vraie sur $C_c^\infty(\Omega)$, avant de passer à la limite en utilisant l'argument de densité : « $C_c^\infty(\Omega)$ est dense dans $L^p(\Omega)$ ».

L'espace $C_c^\infty(\Omega)$ est également noté $\mathcal{D}(\Omega)$ et appelé espace des fonctions tests. Les distributions sont définies comme les éléments de $\mathcal{D}'(\Omega)$, dual topologique de $\mathcal{D}(\Omega)$, muni d'une topologie adéquate. Nous allons les aborder dès à présent, au chapitre 5.

Chapitre 5

Espaces de Sobolev

Résumé — Les espaces de Sobolev sont des espaces fonctionnels. Plus précisément, un espace de Sobolev est un espace vectoriel de fonctions muni de la norme obtenue par la combinaison de la norme L^p de la fonction elle-même ainsi que de ses dérivées jusqu'à un certain ordre. Les dérivées sont comprises dans un sens faible, au sens des distributions afin de rendre l'espace complet.

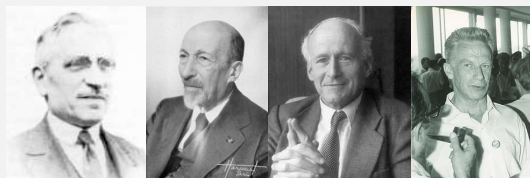
Les espaces de Sobolev sont donc des espaces de Banach. Intuitivement, un espace de Sobolev est un espace de Banach ou un espace de Hilbert de fonctions pouvant être dérivées suffisamment de fois (pour donner sens par exemple à une équation aux dérivées partielles) et muni d'une norme qui mesure à la fois la taille et la régularité de la fonction.

Les espaces de Sobolev sont un outil très important et très adapté à l'étude des équations aux dérivées partielles. En effet, les solutions d'équations aux dérivées partielles, appartiennent plus naturellement à un espace de Sobolev qu'à un espace de fonctions continues dont les dérivées sont comprises dans un sens classique (mais rien n'empêche d'avoir de la chance).

Histoire

Le xx^e siècle avait commencé par la thèse de Lebesgue *intégrale, longueur, aire*, qui constitue vraiment le début de la théorie de la mesure. La théorie de Lebesgue mène à l'étude des espaces L^p , qui permettront, sur les traces de Hilbert, Riesz et Banach, l'étude des opérateurs différentiels.

Cauchy avait publié nombre d'applications de sa théorie dans des recueils d'exercices, notamment concernant l'évaluation d'intégrales réelles, qu'il n'hésita pas à généraliser en ce qu'on appelle aujourd'hui la « valeur principale de Cauchy », un peu moins d'un siècle avant que Hadamard en ait besoin dans sa résolution des équations aux dérivées partielles dans un problème d'hydrodynamique par les « parties finies » et que Laurent Schwartz n'en vienne aux distributions.



Lebesgue Hadamard Schwartz Sobolev

L'étude des conditions de régularité des solutions des équations aux dérivées partielles permet à Sergueï Sobolev et ses continuateurs de définir ses espaces de fonctions et les théorèmes de trace en fonction des propriétés géométriques du domaine.

Quelques mots s'imposent au sujet de la théorie des distributions. Parmi tous les travaux ayant valu à leur auteur la médaille Fields, la théorie des distributions de Laurent Schwartz est l'une des très rares à être abordable par les étudiants dès leur premier cycle universitaire. C'est une partie de l'explication du réel engouement pour les distributions, l'autre partie étant sa puissance, sa commodité d'usage et surtout sa très grande beauté.

Si cette théorie fournit aux analystes un cadre général et un formalisme agréable pour l'étude des espaces fonctionnels et des équations aux dérivées partielles dans lesquels ils aiment à se placer, elle n'est pourtant pas indispensable. D'une part, les spécialistes d'équations aux dérivées partielles parviennent toujours à trouver des formulations bien adaptées à leurs problèmes, en s'inspirant des idées sous-jacentes à la théorie des distributions, mais sans y avoir recours explicitement. D'autre part, on peut étudier la plupart des espaces fonctionnels intéressants sans la notion de distribution :

par exemple, les espaces de Sobolev peuvent être définis en termes de distributions, ou bien en termes de complétion de l'espace des fonctions C^∞ à support compact pour des normes bien choisies.

En utilisant uniquement des espaces de Sobolev et leurs espaces duals, on peut obtenir la grande majorité des distributions que l'on rencontre en pratique : par exemple, la dérivée de la « fonction de Dirac » peut être vue comme un élément de l'espace dual des fonctions C^1 . Plus généralement, quand on rencontre une distribution dans un problème concret, c'est presque à coup sûr un élément du dual d'un espace de Sobolev bien choisi, au moins localement. La topologie des espaces de Sobolev étant beaucoup plus simple que celle des distributions, on comprend que l'étude des espaces de Sobolev soit plus populaire et certainement plus utile en pratique. De plus, les résultats que l'on obtient en utilisant des méthodes plus terre-à-terre sont souvent meilleurs (plus constructifs, plus quantitatifs...).

Pour autant, la théorie des distributions fournit un formalisme commode et élégant, qui apporte de l'ordre dans le paysage fonctionnel, et facilite la communication entre mathématiciens d'horizons très divers. En outre, les principes qui la sous-tendent, bien plus que les théorèmes principaux, s'avèrent d'une importance capitale en pratique. Pour toutes ces raisons, une bonne familiarité avec le langage des distributions est presque indispensable à un analyste. Schwartz lui-même avait bien conscience que le principal mérite de son approche ne résidait pas dans l'introduction d'outils nouveaux, mais dans une synthèse claire et accessible de recettes multiples qui étaient, déjà auparavant, employées dans des contextes divers.

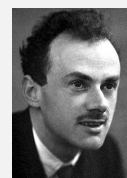
5.1 Distributions

Une distribution (également appelée fonction généralisée) est un objet qui généralise la notion de fonction et de mesure. La théorie des distributions étend la notion de dérivée à toutes les fonctions localement intégrables et au-delà.

Histoire

Généralisation de la notion de fonction :

Depuis le milieu des années 1920, les théoriciens de la physique quantique, et en particulier Dirac, utilisaient des objets étranges qu'ils manipulaient comme des fonctions. Le plus typique était la « fonction de Dirac », mystérieuse fonction qui vaut 0 partout, sauf en 0 où elle vaut $+\infty$, et dont l'intégrale est égale à 1 (en violation de toutes les règles de la théorie de l'intégration de Lebesgue). Non seulement Dirac utilisait cette fonction à des fins de calcul formel, mais encore il se permettait de la dériver à volonté, se contentant de remarquer que les dérivées successives étaient « de plus en plus singulières ».



Dirac

L'utilité et le caractère intuitif de ces objets rendait presque indispensable leur incorporation dans une théorie mathématique ; c'est ce qu'a réalisé la théorie des distributions. Leur interprétation est d'ailleurs très simple, et cause beaucoup moins de maux de tête que les interrogations que l'on peut avoir sur la nature des « fonctions » considérées par Dirac.

Réhabilitation de la dérivation :



Lebesgue



Schwartz

Au début du vingtième siècle, la théorie de l'intégration de Lebesgue a pris un essor rapide, et l'intégration apparaît désormais comme l'opération reine de l'analyse. La plupart des fonctions sont intégrables, alors que très peu sont dérivables. En outre, la théorie de Lebesgue montre que l'intégration est souvent une opération continue vis-à-vis de la convergence des fonctions, uniforme ou même simple (ponctuelle), alors que la dérivation est une opération grossièrement discontinue.

Dans la théorie de Schwartz au contraire, ces problèmes disparaissent : à toute fonction continue on pourra associer des « fonctions dérivées » à tous les ordres, selon une notion qui prolongera celle de dérivée des fonctions continûment dérivables. De manière plus générale, toute distribution sera dérivable à tout ordre, et on pourra définir une notion de convergence qui prolongera la notion de convergence uniforme, et pour laquelle l'opération de dérivée sera continue.

Extension des espaces de solutions acceptables :

Leray en 1934 et Sobolev en 1936 introduisent de nouvelles notions de solutions d'équations aux dérivées partielles, appelées solutions faibles ou solutions généralisées, qui permettent de formuler des équations aux dérivées partielles sans supposer nécessairement l'existence de dérivées au sens classique. Leur approche préfigure très bien la théorie des distributions, qui ne sera mise au point qu'une quinzaine d'années plus tard. Les contributions de Leray et Sobolev ne constituent pas les seuls travaux précurseurs de la théorie des distributions. Dans les années 30 et 40, de nombreux chercheurs vont utiliser les concepts de solutions généralisées pour étudier les solutions de diverses équations aux dérivées partielles : Courant et Hilbert, Bochner, Friedrichs, Krylov...



Leray

Sobolev

Nous arrivons maintenant à l'idée majeure de la théorie des distributions, déjà présente dans les travaux de Sobolev et Leray. Du point de vue physique, on peut la motiver comme suit. Le résultat ou l'interprétation d'une expérience physique faisant intervenir une certaine grandeur, qui varie en fonction du temps ou de l'espace, ne dépend que très rarement des valeurs ponctuelles de cette grandeur, mais plus souvent de sa valeur moyenne (ou intégrale) prise contre une fonction du temps ou de l'espace, plus ou moins localisée. En d'autres termes, plutôt qu'une fonction u , c'est plutôt une « moyenne » de la forme $\int u \varphi$ qui sera accessible en pratique. Ainsi, un signal électrique oscillant avec une période très élevée fournira, de manière apparemment paradoxale, un signal plat quand on l'utilisera pour alimenter un oscilloscope : dans le processus de moyenne inhérent à la mesure, les valeurs positives et négatives se compenseront.

Dans le cas de l'équation des ondes aussi bien que dans celui des équations de transport avec données singulières, on résout ainsi le dilemme qui nous est proposé, en reformulant l'équation par son action sur les fonctions-test.

Explicitons plus en détail la réflexion de Sobolev. Pour des raisons ayant trait à l'analyse (fonctionnelle) de certaines classes d'équations aux dérivées partielles, il cherchait à étudier les fonctions possédant une dérivée carré-intégrable, dans un cadre suffisamment général... et en particulier, sans supposer la fonction dérivable ! Il montra que l'on pouvait donner une définition utile et pratique du concept de « fonction dont la dérivée est de carré intégrable », qui ne présuppose pas l'existence d'une dérivée au sens usuel. L'espace ainsi défini est aujourd'hui appelé espace de Sobolev $W^{1,2}$ ou H^1 .

Simplification de problèmes non linéaires ardu :

L'utilisation de solutions généralisées peut venir d'une nécessité de modélisation, mais également par l'incapacité à prouver l'existence de solutions classiques. L'exemple archétypal en la matière est celui des équations de la mécanique des fluides, en particulier l'équation de Navier-Stokes incompressible, dont la première étude mathématique moderne est due à Leray. L'écriture même de l'équation de Navier-Stokes incompressible présuppose apparemment la dérivabilité de la fonction inconnue u , mais rien ne semblait permettre d'affirmer l'existence de solutions dérivables pour des données initiales assez générales. Ce problème est d'ailleurs toujours ouvert, et compte parmi les sept problèmes du millénaire de l'Institut Clay, mais nous en reparlerons au paragraphe 6.5.1. Leray eut l'idée de définir une solution de l'équation de Navier-Stokes incompressible par une formulation duale qui ne présupposait pas de régularité. De telles solutions sont aujourd'hui appelées solutions faibles.

La théorie des distributions exploite au maximum l'idée de moyenniser les solutions contre des fonctions-test qui se comportent bien : on définit au départ une classe agréable de fonctions-test, à la fois très régulières et bien localisées, à savoir les fonctions indéfiniment dérivables, à support compact. Toutes les propriétés des objets que l'on cherche à définir sont alors définies en fonction des « intégrales » de ces objets contre les fonctions-test.

Définition 44 — Espace des fonctions tests $\mathcal{D}(\Omega)$. Soit Ω un sous-espace topologique de \mathbb{R}^n (ou \mathbb{C}^n). L'espace des fonctions tests $\mathcal{D}(\Omega)$ est l'ensemble des fonctions à valeurs réelles indéfiniment dérivables de Ω à support compact inclus dans Ω .

On munit cet espace vectoriel de la topologie suivante : un ensemble $U \subset \mathcal{D}(\Omega)$ est ouvert si et seulement si $\forall K \subset \Omega$ compact et $f \in U$ dont le support est inclus dans K , il existe $\epsilon > 0$ et $k \geq 0$ tels que :

$$\{g \in \mathcal{D}(\Omega) \mid \text{supp}(g) \subset K \quad \text{et} \quad \forall x \in K, |f^{(k)}(x) - g^{(k)}(x)| \leq \epsilon\} \subset U. \quad (5.1)$$

Muni de cette topologie, $\mathcal{D}(\Omega)$ est un espace vectoriel topologique non métrisable.

Définition 45 — Distribution. Une distribution est une **forme linéaire continue** sur $\mathcal{D}(\Omega)$.

L'ensemble des distributions est donc le dual topologique de $\mathcal{D}(\Omega)$ et est par conséquent noté $\mathcal{D}'(\Omega)$.

Remarque. L'espace des distributions est extrêmement grand, et contient tous les espaces fonctionnels que nous avons mentionnés jusqu'à présent et leur dual, y compris les espaces à poids et les espaces locaux (non présentés dans ce document).

Complément topologique. L'espace \mathcal{D} n'est pas métrisable. Définir sa topologie n'est pas équivalent à définir la convergence des suites. Cependant, dans la pratique c'est presque la seule notion de convergence que l'on utilise.

Soit Ω un ouvert de \mathbb{R}^n . On muni $\mathcal{D}'(\Omega)$ de la topologie faible-* induite par $\mathcal{D}(\Omega)$. Alors :

- $\mathcal{D}'(\Omega)$ est un espace topologique complet (et un espace de Montel) ;
- Le dual topologique de $\mathcal{D}'(\Omega)$ s'identifie à $\mathcal{D}(\Omega)$, qui est donc réflexif ;
- La topologie de $\mathcal{D}'(\Omega)$ n'est pas métrisable.

Notons que la convergence dans le dual de n'importe quel espace de Sobolev implique la convergence au sens des distributions.

Toujours aussi classiquement, si T est une distribution et φ une fonction test de $\mathcal{D}(\Omega)$ alors on note $T(\varphi) = \langle T, \varphi \rangle$. (où $\langle \cdot, \cdot \rangle$ désigne comme d'habitude le crochet de dualité). Dans $\mathcal{D}'(\mathbb{R})$, l'application qui à φ associe $\varphi(0)$ est une distribution et c'est la distribution de Dirac :

$$\langle \delta, \varphi \rangle = \varphi(0) \quad (5.2)$$

Une propriété fondamentale est que toute fonction localement intégrable f représente aussi une distribution T_f définie par la forme intégrale suivante :

$$\langle T_f, \varphi \rangle = \int_{\mathbb{R}} f(x) \varphi(x) dx, \quad \forall \varphi \in \mathcal{D}(\Omega) \quad (5.3)$$

Commentaire sur le choix de l'espace \mathcal{D} . Pourquoi ce choix, qui semble assez restrictif ? Notons tout d'abord que plus on restreint la classe des fonctions-tests, plus l'espace des distributions définies sur ces fonctions sera grand. De plus, le fait que les fonctions-tests soient de classe C^∞ n'est pas très restrictif en fait, car toute fonction continue à support borné est limite uniforme de telles fonctions.

Si l'on considère un espace de fonctions tests plus grand que \mathcal{D} , alors les fonctionnelles linéaires continues sur ce nouvel espace forment un sous-espace vectoriel de \mathcal{D}' . Deux espaces de fonctions tests couramment utilisés sont :

- L'espace \mathcal{E} des fonctions indéfiniment dérivables quelconques ;
- L'espace \mathcal{S} des fonctions indéfiniment dérivables et qui décroissent, ainsi que leurs dérivées, plus vite que toute puissance de $1/x$ à l'infini, i.e. que pour tout $k > 0$ et pour tout $h > 0$, $x^k f^{(h)}(x)$ est bornée.

On obtient alors l'espace \mathcal{E}' des distributions à support compact et l'espace \mathcal{S}' des distributions dites tempérées (ou à croissance lente).

$$\begin{array}{l} \text{Espaces de fonctions tests : } \mathcal{D} \subset \mathcal{S} \subset \mathcal{E} \\ \text{Espaces de distributions : } \mathcal{D}' \supset \mathcal{S}' \supset \mathcal{E}' \end{array}$$

Pour plus de détails sur ce sujet, on consultera un cours sur la transformée de Fourier, ou un cours de traitement du signal.

5.2 Dérivées au sens des distributions

Pour définir la dérivée d'une distribution, considérons d'abord le cas d'une fonction différentiable et intégrable $f : \mathbb{R} \rightarrow \mathbb{R}$. Soit φ une fonction test, supposée régulière et à support compact. Une intégration par parties permet d'écrire :

$$\int_{\mathbb{R}} f'(x) \varphi(x) dx = - \int_{\mathbb{R}} f(x) \varphi'(x) dx \quad \text{soit} \quad I_{f'}(\varphi) = I_{-f}(\varphi') \quad (5.4)$$

En effet, puisque la fonction φ est nulle en dehors d'un ensemble borné (elle est à support compact), les termes de bords s'annulent.

Définition 46 — Dérivée d'une distribution. Si S est une distribution, cet exemple suggère que l'on puisse définir sa dérivée S' comme la forme linéaire qui à une fonction test φ fait correspondre la valeur $-S(\varphi')$. On pose donc :

$$\langle S', \varphi \rangle = -\langle S, \varphi' \rangle \quad (5.5)$$

Cette définition étend la notion classique de dérivée : chaque distribution devient indéfiniment dérivable et l'on peut montrer les propriétés usuelles des dérivées.

Cette notion peut aussi se définir comme la dérivée du produit de dualité $\langle S, \varphi \rangle$. On note $\varphi_h : x \rightarrow \varphi(x - h)$ la translatée de φ par la valeur $h \in \mathbb{R}$, alors, en utilisant la linéarité et la continuité par rapport au deuxième terme :

$$\left\langle \frac{dS}{dx}, \varphi \right\rangle = \lim_{h \rightarrow 0} \frac{\langle S, \varphi_h \rangle - \langle S, \varphi \rangle}{h} = \left\langle S, \lim_{h \rightarrow 0} \frac{\varphi_h - \varphi}{h} \right\rangle = -\left\langle S, \frac{d\varphi}{dx} \right\rangle \quad (5.6)$$

Lorsque la distribution S modélise un phénomène physique, la fonction test φ peut s'interpréter comme un instrument de mesure, $\langle S, \varphi \rangle$ en étant le résultat ; la définition ci-dessus représente alors la mesure expérimentale (au moins de pensée) de la dérivée du phénomène S à l'aide de l'instrument φ .

Nous définirons dans un autre cours (sur le traitement du signal) les notions de distribution à support compact, les distributions tempérées et la décroissance rapide utiles en analyse de Fourier ainsi que les espaces de Schwartz.

Complément au sujet des opérations sur des distributions. L'addition, la soustraction et la multiplication par un scalaire sont telles qu'on les imagine.

Le cas de la dérivation (pour lequel les distributions ont été inventées) a été traité ci-dessus à la définition 46.

Les distributions permettent de changer momentanément l'ouvert Ω sur lequel on travaille. Pour cela, on dispose des trois opérations suivantes :

- Opération de *restriction* : Soient Ω un ouvert de \mathbb{R}^n , et Ω' un ouvert de Ω . Toute distribution $T \in \mathcal{D}'(\Omega)$ définit une distribution sur Ω' , simplement par restriction de l'espace des fonctions test : $\mathcal{D}(\Omega') \subset \mathcal{D}(\Omega)$. On peut donc restreindre une distribution à un ouvert plus petit que celui sur lequel elle était définie initialement.
- Opération de *localisation* : Soient un compact K (au voisinage duquel on veut étudier la distribution), O un ouvert contenant K , et $T \in \mathcal{D}'(\Omega)$ une distribution. On peut trouver une autre distribution T' dont le support est inclus dans O , et qui coïncide avec T dans un voisinage de K . Pour cela il suffit de définir $\langle T', \varphi \rangle = \langle T, \chi_\varphi \rangle$ où χ_φ est une fonction plateau identiquement 1 au voisinage de K , à support compact inclus dans O .
- Opération de *recollement* : Soient $(\Omega_i)_{i \in J}$ une famille (éventuellement infinie) d'ouverts de \mathbb{R}^n , et Ω la réunion des Ω_i . Pour chaque i , on se donne une distribution $T_i \in \mathcal{D}'(\Omega_i)$. On peut définir dans Ω une distribution T dont la restriction à chaque Ω_i soit T_i , sous la condition nécessaire que $T_i = T_j$ dans $\Omega_i \cap \Omega_j$. On utilise le théorème de partition de l'unité pour construire cette distribution, et on montre que celle-ci est indépendante de la partition choisie.

La multiplication des distributions pose des problèmes considérables et ne se résout que partiellement.

Exemples de calculs avec les distributions. Considérons la fonction $f(x) = |x|$ sur \mathbb{R} . Alors f' est la fonction « signe », $f'' = 2\delta_0$, f''' vaut 2 fois l'application « évaluation de la dérivée en 0 ». Attention donc à ne pas confondre ces dérivées au sens des distributions avec les dérivées presque partout, qui sont nulles à partir du rang 2.

Soit $f(x) = x \log |x| - x$ dans \mathbb{R} , alors on peut écrire, au sens des distributions, $f'(x) = \log |x|$, $f''(x) = v.p.(1/x)$ (où *v.p.* est la valeur principale).

Soit (u_n) la suite de fonctions définie par $u_n(x) = (1/n) \sin(nx)$. Alors la famille (u'_n) converge au sens des distributions vers 0.

5.3 Espaces $W^{m,p}(\Omega)$

Définition 47 — Espace $W^{m,p}(\Omega)$. Soient $\Omega \subset \mathbb{R}^n$ un ouvert quelconque, p un réel tel que $1 \leq p \leq \infty$ et m un entier naturel positif. On définit l'espace de Sobolev $W^{m,p}(\Omega)$ par :

$$W^{m,p}(\Omega) = \{u \in L^p(\Omega); D^\alpha u \in L^p(\Omega)\} \quad (5.7)$$

où α est un multi-indice tel que $0 \leq |\alpha| \leq m$, $D^\alpha u$ est une dérivée partielle de u au sens faible (i.e. au sens des distributions) et L^p un espace de Lebesgue.

La norme sur $W^{m,p}(\Omega)$ est :

$$\|u\|_{W^{m,p}} = \begin{cases} \left(\sum_{0 \leq |\alpha| \leq m} \|D^\alpha u\|_{L^p}^p \right)^{1/p}, & 1 \leq p < +\infty; \\ \max_{0 \leq |\alpha| \leq m} \|D^\alpha u\|_{L^\infty}, & p = +\infty; \end{cases} \quad (5.8)$$

où $\|\cdot\|_{L^p}$ désigne la norme des espaces de Lebesgue. Muni de cette norme $W^{m,p}(\Omega)$ est un espace de Banach. Dans le cas où $p < \infty$, c'est aussi un espace séparable. La norme :

$$\|u\|_{W^{m,p}} = \begin{cases} \sum_{0 \leq |\alpha| \leq m} \|D^\alpha u\|_{L^p}, & 1 \leq p < +\infty; \\ \sum_{0 \leq |\alpha| \leq m} \|D^\alpha u\|_{L^\infty}, & p = +\infty. \end{cases} \quad (5.9)$$

est une norme équivalente à la précédente. On utilisera donc indifféremment l'une de ces normes, et on notera la norme employée $\|\cdot\|_{W^{m,p}}$ ou plutôt $\|\cdot\|_{m,p}$.

Si $p < \infty$, alors $W^{m,p}(\Omega)$ est identique à la fermeture de l'ensemble $\{u \in C^m(\Omega); \|u\|_{m,p} < \infty\}$ par rapport à la norme $\|\cdot\|_{m,p}$ où $C^m(\Omega)$ désigne l'espace de Hölder des fonctions de classe m sur Ω .

Remarque. Le théorème de Meyers-Serrin dit « $H = W$ » (qui est le titre de leur article de 1964 [63]) montre l'équivalence de deux définitions des espaces de Sobolev $W^{m,p}(\Omega) = H^{m,p}(\Omega)$, i.e. dans notre cas, donne une définition équivalente, par complétion de l'espace vectoriel normé :

$$\{u \in C^\infty(\Omega); \|u\|_{H^{m,p}} < \infty\} \quad \text{avec} \quad \|u\|_{H^{m,p}} := \left(\sum_{|\alpha| \leq m} \|D^\alpha u\|_{L^p}^p \right)^{1/p} \quad (5.10)$$

où $D^\alpha u$ est une dérivée partielle de u au sens classique ($u \in C^\infty(\Omega)$).

5.4 Espaces $H^m(\Omega)$, $H_0^m(\Omega)$ et $H^{-m}(\Omega)$

Définition 48 — Espace $H^m(\Omega)$. Dans le cas $p = 2$, on note $H^m(\Omega)$ l'espace $W^{m,2}(\Omega)$, défini par la relation (5.7).

Les espaces de Sobolev H^m ont un intérêt particulier car il s'agit alors d'espaces de Hilbert. Leur norme est induite par le produit intérieur suivant :

$$(u, v)_m = \sum_{0 \leq |\alpha| \leq m} (D^\alpha u, D^\alpha v) \quad (5.11)$$

où :

$$(u, v) = \int_{\Omega} u(x) \bar{v}(x) \, dx \quad (5.12)$$

est le produit intérieur dans $L^2(\Omega)$; produit scalaire dans le cas réel, hermitien dans le cas complexe.

Théorème 29 Si Ω est Lipschitzien, alors l'ensemble $C^\infty(\overline{\Omega})$ des fonctions régulières jusqu'au bord de Ω est dense dans $H^m(\Omega)$.

Complément. De plus, dans le cas où la transformation de Fourier peut être définie dans $L^2(\Omega)$, l'espace $H^k(\Omega)$ peut être défini de façon naturelle à partir de la transformée de Fourier (voir cours sur le traitement du signal). Par exemple si $\Omega = \mathbb{R}^n$, si \hat{u} est la transformée de Fourier de u :

$$H^m(\mathbb{R}^n) = \{u \in L^2(\mathbb{R}^n); \int_{\mathbb{R}^n} |\hat{u}(\xi)|^2 \xi^\alpha d\xi < \infty \quad (5.13)$$

pour $|\alpha| \leq m\}$, ou ce qui est équivalent :

$$H^m(\mathbb{R}^n) = \left\{ u \in L^2(\mathbb{R}^n); \int_{\mathbb{R}^n} |\hat{u}(\xi)|^2 (1 + \xi^2)^m d\xi < \infty \right\} \quad (5.14)$$

et que :

$$(u, v)_m = \int_{\mathbb{R}^n} \hat{u}(\xi) \overline{\hat{v}(\xi)} (1 + \xi^2)^m d\xi \quad (5.15)$$

est un produit hermitien équivalent à celui défini plus haut. Encore, si $\Omega =]0; 1[$, on vérifie que :

$$H^m(]0; 1[) = \left\{ u \in L^2(]0; 1[); \sum_{n=-\infty}^{+\infty} (1 + n^2 + n^4 + \dots + n^{2m}) |\hat{u}_n|^2 < \infty \right\} \quad (5.16)$$

où \hat{u}_n est la série de Fourier de u . On peut aussi utiliser la norme équivalente :

$$\|u\|^2 = \sum_{n=-\infty}^{+\infty} (1 + |n|^2)^m |\hat{u}_n|^2 \quad (5.17)$$

Définition 49 — Espace $H_0^m(\Omega)$. On définit $H_0^m(\Omega)$ comme l'adhérence dans $H^m(\Omega)$ de $\mathcal{D}(\Omega)$, ensemble des fonctions de classe C^∞ et à support compact dans Ω .

Plus exactement et de manière générale, $W_0^{m,p}$ est l'adhérence de $\mathcal{D}(\Omega)$ dans l'espace $W^{m,p}$. La définition ci-dessus correspond donc au cas $p = 2$.

Remarque. Il faut bien comprendre que cela signifie quelque chose de simple : c'est l'ensemble constitué des fonctions qui sont à la fois $\mathcal{D}(\Omega)$ et $H^m(\Omega)$. Comme on s'intéresse à la frontière (puisque l'on prend l'adhérence), et que Ω est un ouvert, la valeur que prennent les fonctions de cet ensemble sur la frontière est celle que prennent les fonctions de classe C^∞ et à support compact dans Ω , soit 0 !

Ainsi, « physiquement » les espaces $H_0^m(\Omega)$ sont les fonctions de $H^m(\Omega)$ qui sont nulles sur $\Gamma = \partial\Omega$, ainsi que leurs dérivées « normales » jusqu'à l'ordre $m - 1$. Plus mathématiquement, on dit que ces fonctions sont de trace nulle sur la frontière, comme par exemple :

$$H_0^2(\Omega) = \left\{ v/v \in H^2(\Omega), v|_\Gamma = 0, \frac{\partial v}{\partial n}|_\Gamma = 0 \right\} \quad (5.18)$$

Les espaces trace sont définis plus loin.. Les espaces $H_0^m(\Omega)$ sont des espaces de Hilbert.

Définition 50 — Espace $H^{-m}(\Omega)$. Il est possible de caractériser le dual topologique de $H_0^m(\Omega)$ de la façon suivante. Pour tout $m \geq 1$, on définit l'espace des distributions suivant :

$$H^{-m}(\Omega) = \left\{ f \in \mathcal{D}'(\Omega), f = \sum_{|\alpha| \leq m} \partial^\alpha f_\alpha, \text{ avec } f_\alpha \in L^2(\Omega) \right\} \quad (5.19)$$

muni de la norme :

$$\|f\|_{H^{-m}} = \inf \left(\sum_{|\alpha| \leq m} \|f_\alpha\|_{L^2}^2 \right)^{\frac{1}{2}} \quad (5.20)$$

l'infimum étant pris sur toutes les décompositions possibles de f sous la forme intervenant dans la définition de H^{-m} .

Ainsi défini, l'espace $H^{-m}(\Omega)$ est un espace de Hilbert isomorphe au dual topologique de $H_0^m(\Omega)$, et le crochet de dualité s'écrit :

$$\langle f, u \rangle_{H^{-m}, H_0^m} = \sum_{|\alpha| \leq m} (-1)^\alpha \int_{\Omega} f_{\alpha} \partial^{\alpha} u dx, \quad \forall f \in H^{-m}(\Omega), \quad \forall u \in H_0^m(\Omega) \quad (5.21)$$

Cette formule ne dépend pas de la décomposition de f en somme des $\partial^{\alpha} f_{\alpha}$. Remarquons que le dual de $H^m(\Omega)$ n'est pas un espace de distribution et ne possède donc pas de caractérisation aussi simple.

5.5 Prise en compte du contour du domaine

5.5.1 Trace

Dans ce paragraphe, nous essayons de présenter la notion de trace de manière simple et intuitive.

Afin de pouvoir parler de la valeur d'une fonction sur la frontière de Ω , il nous faut définir le prolongement (la trace) d'une fonction sur ce bord.

Cas $n = 1$: on considère un intervalle ouvert $I =]a; b[$ borné. On a vu que $H^1(I) \subset C^0(\bar{I})$. Donc, pour $u \in H^1(I)$, u est continue sur $[a; b]$, et $u(a)$ et $u(b)$ sont bien définies.

Cas $n > 1$: il nous faut définir la trace lorsque l'on n'a plus $H^1(I) \subset C^0(\bar{I})$. On procède ainsi :
— On définit l'espace :

$$C^1(\bar{\Omega}) = \{ \varphi : \Omega \rightarrow \mathbb{R} / \exists O \text{ ouvert contenant } \bar{\Omega}, \exists \psi \in C^1(O), \psi|_{\Omega} = \varphi \} \quad (5.22)$$

$C^1(\bar{\Omega})$ est donc l'espace des fonctions C^1 sur Ω , prolongeables par continuité sur $\partial\Omega$ et dont le gradient est lui-aussi prolongeable par continuité. Il n'y a donc pas de problème pour définir la trace de telles fonctions.

- On montre que, si Ω est un ouvert borné de frontière $\partial\Omega$ « assez régulière », alors $C^1(\bar{\Omega})$ est dense dans $H^1(\Omega)$.
- L'application linéaire continue, qui à toute fonction u de $C^1(\bar{\Omega})$ associe sa trace sur $\partial\Omega$, se prolonge alors en une application linéaire continue de $H^1(\Omega)$ dans $L^2(\partial\Omega)$, notée γ_0 , qu'on appelle application trace. On dit que $\gamma_0(u)$ est la trace de u sur $\partial\Omega$.

Pour une fonction u de $H^1(\Omega)$ qui soit en même temps continue sur $\bar{\Omega}$, on a évidemment $\gamma_0(u) = u|_{\partial\Omega}$. C'est pourquoi on note souvent par abus simplement $u|_{\partial\Omega}$ plutôt que $\gamma_0(u)$.

De manière analogue, on peut définir γ_1 , l'application trace qui permet de prolonger la définition usuelle de la dérivée normale sur $\partial\Omega$. Pour $u \in H^2(\Omega)$, on a $\partial_i u \in H^1(\Omega)$, $\forall i = 1, \dots, n$ et on peut donc définir $\gamma_0(\partial_i u)$. La frontière $\partial\Omega$ étant « assez régulière » (par exemple, idéalement, de classe C^1), on peut définir la normale $n = (n_1, \dots, n_n)^T$ en tout point de $\partial\Omega$. On pose alors :

$$\gamma_1(u) = \sum_{i=1}^n \gamma_0(\partial_i u) n_i \quad (5.23)$$

Cette application continue γ_1 de $H^2(\Omega)$ dans $L^2(\partial\Omega)$ permet donc bien de prolonger la définition usuelle de la dérivée normale. Dans le cas où u est une fonction de $H^2(\Omega)$ qui soit en même temps dans $C^1(\bar{\Omega})$, la dérivée normale au sens usuel de u existe, et $\gamma_1(u)$ lui est évidemment égale. C'est pourquoi on note souvent, par abus, $\partial_n u$ plutôt que $\gamma_1(u)$.

Exemple : traitement d'une interface - divergence faible. Soit Ω un ouvert borné constitué de deux ouverts Ω_1 et Ω_2 séparés par une interface $\Gamma = \partial\Omega_1 \cap \partial\Omega_2$. Montrons qu'une fonction vectorielle de classe C^1 sur chaque morceau Ω_1 et Ω_2 admet une divergence faible dans $L^2(\Omega)$ si et seulement si sa composante normale est continue à travers l'interface (la surface) Γ .

Soit f une fonction de Ω à valeurs vectorielles. On note f_1 et f_2 ses restrictions à Ω_1 et Ω_2 respectivement. On note également n_1 et n_2 les normales extérieures à Ω_1 et Ω_2 respectivement.

Soit $\varphi \in C_c^\infty(\Omega)$. La formule de Stokes nous permet d'écrire :

$$\begin{aligned}\int_{\Omega} f \cdot \nabla \varphi &= \int_{\Omega_1} f_1 \cdot \nabla \varphi + \int_{\Omega_2} f_2 \cdot \nabla \varphi \\ &= \int_{\Gamma} f_1 \cdot n_1 \varphi - \int_{\Omega_1} \operatorname{div}(f_1) \varphi - \int_{\Gamma} f_2 \cdot n_2 \varphi + \int_{\Omega_2} \operatorname{div}(f_2) \varphi\end{aligned}$$

Si la composante normale de f est continue à l'interface, on en déduit que :

$$\int_{\Omega} f \cdot \nabla \varphi = - \int_{\Omega} \psi \varphi$$

avec $\psi = \operatorname{div}(f_i)$ sur chaque Ω_i ($i = 1, 2$). La fonction à valeurs vectorielles f admet donc une divergence faible et $\operatorname{div}(f) = \psi$.

Réciproquement, si f possède une divergence faible, alors il existe une fonction $v \in L^2(\Omega)$ telle que :

$$\int_{\Gamma} (f_1 - f_2) \cdot n_1 \varphi = \int_{\Omega} v \varphi$$

et on en déduit que $(f_1 - f_2) \cdot n_1 = 0$ sur Γ .

5.5.2 Espace trace

Dans le cas d'exposant entier, on note souvent l'ordre avec la lettre m , dans le cas non-entier, on utilisera la lettre s , et donc les espaces seront notés : $W^{s,p}$ ou H^s .

Cas $p = 2$ et $\Omega = \mathbb{R}^n$

Dans ce cas, l'espace de Sobolev $H^s(\mathbb{R}^n)$, $s \geq 0$, peut être défini grâce à la transformée de Fourier :

$$H^s(\mathbb{R}^n) = \left\{ u \in L^2(\mathbb{R}^n) : \int_{\mathbb{R}^n} (1 + |\xi|^2)^s |\hat{u}(\xi)|^2 d\xi < +\infty \right\}. \quad (5.24)$$

$H^s(\mathbb{R}^n)$ est un espace de Hilbert muni de la norme :

$$\|u\|_{H^s}^2 = \int_{\mathbb{R}^n} (1 + |\xi|^2)^s |\hat{u}(\xi)|^2 d\xi \quad (5.25)$$

Cas $p = 2$ et $\Omega \subset \mathbb{R}^n$ quelconque

On peut alors caractériser les espaces de Sobolev d'ordre fractionnaire $H^s(\Omega)$ grâce au produit intérieur donné par :

$$(u, v)_{H^s(\Omega)} = (u, v)_{H^k(\Omega)} + \sum_{|\alpha|=k} \int_{\Omega} \int_{\Omega} \frac{(D^\alpha u(x) - D^\alpha u(y))(D^\alpha v(x) - D^\alpha v(y))}{|x - y|^{n+2s}} dx dy \quad (5.26)$$

où $s = k + T$, k est un entier tel que $0 < T < 1$ et n est la dimension du domaine $\Omega \subset \mathbb{R}^n$. La norme induite est essentiellement l'analogue pour L^2 de la continuité au sens de Hölder.

Cas $p \neq 2$ et $\Omega =]0; 1[$

On définit un opérateur D^s de dérivation d'ordre fractionnaire s par :

$$D^s u = \sum_{n=-\infty}^{\infty} (in)^s \hat{u}(n) e^{int} \quad (5.27)$$

En d'autres mots, il s'agit de prendre la transformée de Fourier, de la multiplier par $(in)^s$ et à prendre la transformée de Fourier inverse (les opérateurs définis par la séquence : transformation

de Fourier — multiplication — transformation inverse de Fourier sont appelés des multiplicateurs de Fourier). Cet opérateur permet de définir la norme de Sobolev de $H^s(]0; 1[)$ par : $\|u\|_{s,p} = \|u\|_p + \|D^s u\|_p$ et de définir l'espace de Sobolev $H^s(]0; 1[)$ comme l'espace des fonctions pour lesquelles cette norme est finie.

Cas général des espaces H^s

Soit $s > \frac{1}{2}$. Si Ω est un ouvert dont la frontière $\partial\Omega$ est « suffisamment régulière », alors on peut définir un opérateur de trace T qui à une fonction $u \in H^s(\Omega)$ lui associe sa trace, i.e sa restriction sur la frontière de Ω : $Tu = u|_{\partial\Omega}$.

Une hypothèse simple qui satisfasse la condition de régularité est que $\partial\Omega$ soit uniformément C^m pour $m \geq s$. Ainsi défini, cet opérateur de trace T a pour domaine de définition $H^s(\Omega)$ et son image est précisément $H^{s-1/2}(\partial\Omega)$.

En fait, T est d'abord défini pour les fonctions indéfiniment dérivables et cette définition est ensuite étendue par continuité à tout l'ensemble $H^s(\Omega)$. De façon intuitive, on peut dire que l'on perd en régularité « une demi-dérivée » en prenant la trace d'une fonction de $H^s(\Omega)$.

Nous nous servirons de ces espaces dans le cas des éléments finis pour un problème de continuité des contraintes à une interface entre deux milieux solides ayant des propriétés matérielles différentes. Ce problème sera abordé plusieurs fois dans ce document, et le paragraphe 12.3 fera une synthèse des stratégies possibles pour le résoudre.

Cas général des espaces $W^{s,p}$

Définir la trace d'une fonction de $W^{s,p}$ est très difficile et demande d'utiliser les techniques plus compliquées (dont les espaces de Besov).

De façon plus intuitive, on peut dire que l'on perd en régularité $1/p$ -ème de dérivée en prenant la trace d'une fonction de $W^{s,p}(\Omega)$.

5.6 Espaces $H^1(\Omega)$, $H_0^1(\Omega)$ et $H^{-1}(\Omega)$

Nous avons déjà défini les espaces $H^m(\Omega)$, $H_0^m(\Omega)$ et $H^{-m}(\Omega)$. Toutefois, dans la pratique, nous n'irons guère au delà de $m = 2$, et même le plus souvent nous nous contenterons de $m = 1$. Nous donnons ici quelques compléments dans ce dernier cas.

En application de la définition 48, nous avons :

$$H^1(\Omega) = \left\{ u \in L^2(\Omega); \forall i = 1, \dots, n, \frac{\partial u}{\partial x_i} \in L^2(\Omega) \right\} \quad (5.28)$$

Muni du produit scalaire :

$$(u, v)_1 = \int_{\Omega} \left(uv + \sum_{i=1}^N \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_i} \right) \quad (5.29)$$

$H^1(\Omega)$ est un espace de Hilbert.

En physique et en mécanique, l'espace $H^1(\Omega)$ est également appelé « espace d'énergie » au sens où il est constitué des fonctions d'énergie finie (i.e. de norme finie).

Théorème 30 — Théorème de densité. Si Ω est un borné régulier de classe C^1 , ou si $\Omega = \mathbb{R}_+^n$, ou encore si $\Omega = \mathbb{R}^n$, alors $C_c^\infty(\overline{\Omega})$ est dense dans $H^1(\Omega)$.

En pratique, il est très important de savoir si les fonctions régulières sont denses dans l'espace de Sobolev $H^1(\Omega)$. Cela justifie en partie la notion d'espace de Sobolev qui apparaît ainsi très simplement comme l'ensemble des fonctions régulières complétées par les limites des suites de fonctions régulières dans la norme de l'énergie. Cela permet de démontrer facilement de nombreuses propriétés en les établissant d'abord sur les fonctions régulières puis en utilisant un argument de densité.

Par définition de $H_0^1(\Omega)$, et en prenant en compte une remarque précédente :

$$H_0^1(\Omega) = \{v / v \in H^1(\Omega), v|_{\Gamma} = 0\} \quad (5.30)$$

On voit que sur cet espace, la condition de Dirichlet est satisfaite automatiquement sur tout le pourtour $\Gamma = \partial\Omega$. La frontière Γ est généralement partitionnée en deux sous-frontière Γ_D et Γ_N sur lesquelles on satisfait les conditions de Dirichlet et de Neumann respectivement : $\Gamma = \Gamma_D \cup \Gamma_N$

$$H_{0,D}^1(\Omega) = \{v / v \in H^1(\Omega), v|_{\Gamma_D} = 0\} \quad (5.31)$$

et $H_0^1(\Omega) \subset H_{0,D}^1(\Omega) \subset H^1(\Omega)$.

On rappelle que l'espace $H^{-1}(\Omega)$ est le dual de $H_0^1(\Omega)$. Or, grâce au théorème de représentation de Riesz-Fréchet (théorème 38 du chapitre 7 portant sur les formulations faibles), on sait que l'on peut identifier le dual d'un espace de Hilbert avec lui-même. Cependant en pratique, on n'identifie pas $H^{-1}(\Omega)$ et $H_0^1(\Omega)$. En effet, ayant défini $H_0^1(\Omega)$ comme un sous-espace strict mais dense de $L^2(\Omega)$, et ayant déjà identifié $L^2(\Omega)$ à son dual (muni du produit scalaire usuel, voir chapitre précédent), on ne peut pas en plus identifier $H^{-1}(\Omega)$ et $H_0^1(\Omega)$ (avec un autre produit scalaire). On a donc les inclusions strictes suivantes :

$$H_0^1(\Omega) \subset L^2(\Omega) \equiv (L^2(\Omega))' \subset H^{-1}(\Omega) \quad (5.32)$$

Grâce à $H^{-1}(\Omega)$, on pourrait définir une nouvelle notion de dérivation pour les fonctions de $L^2(\Omega)$, plus faible encore que la dérivée faible. Devant l'afflux de notions de dérivations, rassurons le lecteur en disant qu'elles sont toutes des avatars de la dérivation au sens des distributions (c'est l'intérêt de la théorie des distributions que d'avoir unifié ces divers types de dérivation).

L'exemple le plus simple à retenir, et le plus utile pour la suite, est que tout élément f de $H^{-1}(\Omega)$ s'écrit, au sens des distributions, sous la forme :

$$f = u + \operatorname{div} G \quad (5.33)$$

avec $u \in L^2(\Omega)$ et $G \in (L^2(\Omega))^n$.

Théorème 31 — Dérivation des fonctions composées dans les espaces de Sobolev. Soit Ω un ouvert borné de \mathbb{R}^n . Pour toute fonction $u \in H^1(\Omega)$ et toute fonction $T : \mathbb{R} \rightarrow \mathbb{R}$ de classe C^1 à dérivée bornée nous avons :

$$T(u) \in H^1(\Omega) \quad \text{et} \quad \nabla T(u) = T'(u) \nabla u \quad (5.34)$$

De plus, l'application $u \in H^1(\Omega) \mapsto T(u) \in H^1(\Omega)$ est continue.

On se contente ici de donner les résultats concernant l'espace $H^1(\Omega)$ même si des résultats similaires peuvent être démontrés pour les espaces $H^m(\Omega)$ ou les espaces $W^{1,p}(\Omega)$, avec $p \neq 2$ (voir théorème 32). Soit Ω un ouvert borné Lipschitzien de \mathbb{R}^n .

- si $n = 1$, on a une injection continue de $H^1(\Omega)$ dans l'espace de Hölder $C^{0, \frac{1}{2}}(\Omega)$;
- si $n = 2$, on a une injection continue de $H^1(\Omega)$ dans l'espace $L^p(\Omega)$ pour tout $p < \infty$ (et donc pas dans $L^\infty(\Omega)$).
- si $n \geq 3$, on a une injection continue de $H^1(\Omega)$ dans l'espace $L^{p^*}(\Omega)$ avec $p^* = \frac{2n}{n-2}$.

De plus les injection non critiques sont compactes.

Comme on va s'intéresser par la suite à la discrétisation de problème aux dérivées partielles, le cadre de domaines bornés nous suffira.

5.7 Espaces $H(\text{div})$ et $H(\text{rot})$

Définition 51 — Espace $H(\text{div}, \Omega)$. L'espace $H(\text{div}, \Omega)$ est un espace intermédiaire entre $L^2(\Omega)$ et $H^1(\Omega)$, qui est défini par :

$$H(\text{div}, \Omega) = \{f \in L^2(\Omega)^n; \text{div } f \in L^2(\Omega)\} \quad (5.35)$$

C'est un espace de Hilbert, muni du produit scalaire :

$$\langle f, g \rangle = \int_{\Omega} (f(x) \cdot g(x) + \text{div } f(x) \text{div } g(x)) \, dx \quad (5.36)$$

et de la norme $\|f\|_{H(\text{div}, \Omega)} = \sqrt{\langle f, f \rangle}$. Un des intérêts de l'espace $H(\text{div}, \Omega)$ est qu'il permet de démontrer un théorème de trace et une formule de Green avec encore moins de régularité que dans $H^1(\Omega)$. En effet, si $f \in H(\text{div}, \Omega)$, on ne « contrôle » qu'une seule combinaison de ses dérivées partielles, et non pas toutes comme dans $H^1(\Omega)$, mais on peut quand même donner un sens à la trace normale $f \cdot n$ sur $\partial\Omega$.

Définition 52 — Espace $H(\text{rot}, \Omega)$. L'espace $H(\text{rot}, \Omega)$ est défini par :

$$H(\text{rot}, \Omega) = \{f \in (L^2(\Omega))^2; \text{rot}(f) \in L^2(\Omega)\} \quad (5.37)$$

$H(\text{rot}, \Omega)$ est un espace de Hilbert muni de la norme :

$$\|f\|_{H(\text{rot}, \Omega)} = \left(\|f\|_{L^2(\Omega)}^2 + \|\text{rot}(f)\|_{L^2(\Omega)}^2 \right)^{1/2} \quad (5.38)$$

L'application trace tangentielle $f \mapsto f \wedge n$ est continue de $H(\text{rot}, \Omega)$ sur un espace de Hilbert de fonctions définies sur $\partial\Omega$ qui n'est pas précisé. On peut introduire l'espace $H_0(\text{rot}, \Omega)$ des fonctions de $H(\text{rot}, \Omega)$ de trace tangentielle nulle. Cet espace peut être utilisé pour la modélisation de phénomènes électromagnétiques par les équations de Maxwell.

En fait, il est possible de définir de nombreux espaces de Sobolev en fonction du problème considéré et c'est tout l'intérêt.

5.8 Inégalités utiles

Dans ce paragraphe, nous présentons quelques inégalités utilisées pour borner une fonction à partir d'une estimation sur ses dérivées et de la géométrie de son domaine de définition. Nous en aurons besoin pour les formulations faibles ainsi que pour les problèmes d'homogénéisation.

Définition 53 — Exposant conjugué de Sobolev. On s'intéresse à Ω un ouvert de \mathbb{R}^n (qui peut être \mathbb{R}^n tout entier), ainsi qu'aux espaces de type $W^{m,p}(\Omega)$.

On appelle conjugué de Sobolev du nombre p , le nombre p^* , défini par la relation :

$$\frac{1}{p^*} = \frac{1}{p} - \frac{m}{n} \quad (5.39)$$

Théorème 32 — Injections continues de Sobolev. Soit Ω un ouvert de \mathbb{R}^n . Si Ω est borné et a une frontière Lipschitz-continue, alors, pour tout entier $m \geq 0$ et pour tout $p \in [1; +\infty[$, on

dispose des inclusions avec injections continues suivantes :

$$\begin{cases} W^{m,p}(\Omega) \hookrightarrow L^{p^*} & \text{si } m < \frac{n}{p} \\ W^{m,p}(\Omega) \hookrightarrow L^q & \forall q \in [1; +\infty[, \text{ si } m = \frac{n}{p} \\ W^{m,p}(\Omega) \hookrightarrow C(\overline{\Omega}) & \text{si } \frac{n}{p} < m \end{cases} \quad (5.40)$$

Théorème 33 — Inégalité de Poincaré. Soit p , tel que $1 \leq p < \infty$, et Ω un ouvert de largeur finie (i.e. borné dans une direction). Alors il existe une constante C , dépendant uniquement de Ω et p , telle que, pour toute fonction $u \in W_0^{1,p}(\Omega)$, on ait :

$$\|u\|_{L^p(\Omega)} \leq C \|\nabla u\|_{L^p(\Omega)} \quad (5.41)$$

Théorème 34 — Inégalité de Poincaré-Friedrichs. Si Ω est borné de diamètre d , alors la constante de Poincaré précédente s'exprime en fonction de ce diamètre. On a, pour toute fonction $u \in W_0^{k,p}(\Omega)$:

$$\|u\|_{L^p(\Omega)} \leq d^p \left(\sum_{|\alpha| \leq k} \|D^\alpha u\|_{L^p(\Omega)}^p \right)^{1/p} \quad (5.42)$$

En particulier pour $p = 2$:

$$\forall u \in H_0^1(\Omega), \quad \|u\|_{L^2(\Omega)}^2 \leq d^2 \|\nabla u\|_{L^2(\Omega)}^2 \quad (5.43)$$

Théorème 35 — Inégalité de Poincaré-Wirtinger. Soit p , tel que $1 \leq p < \infty$ et Ω un domaine (i.e. un ouvert connexe) lipschitzien (i.e. borné et « à frontière lipschitzienne ») de l'espace euclidien \mathbb{R}^n . Alors il existe une constante C , dépendant uniquement de Ω et p , telle que, pour toute fonction $u \in W^{1,p}(\Omega)$, on ait :

$$\|u - u_\Omega\|_{L^p(\Omega)} \leq C \|\nabla u\|_{L^p(\Omega)} \quad (5.44)$$

où $u_\Omega = \frac{1}{|\Omega|} \int_\Omega u(y) dy$ est la valeur moyenne de u sur Ω , $|\Omega|$ désignant la mesure de Lebesgue du domaine Ω .

Histoire

Arthur Korn a été un étudiant de Henri Poincaré.

Il est plus connu comme pionnier des télécommunications, plus précisément comme l'inventeur de la téléphotographie. Il met au point un téléautographe, un système de transmission des images fixes à distance, par le biais du fil télégraphique, en recourant aux propriétés photoélectriques du sélénium.



Poincaré

Friedrichs

Wirtinger

Korn

Théorème 36 — Inégalité de Korn. Soit Ω un ouvert borné de \mathbb{R}^n de la frontière est « suffisamment » régulière (par exemple C^1 par morceaux). Notons $\varepsilon(v) = \frac{1}{2} (\nabla v + (\nabla v)^T)$ le linéarisé du tenseur des déformations défini pour tout $v \in V = (H^1(\Omega))^n$ ou $(H_0^1(\Omega))^n$. Alors il existe

une constante (dite constante de Korn) $C_K > 0$ telle que :

$$\forall v \in V, \quad \|v\|_V \leq C_K \left(\|v\|_{L^2(\Omega)}^2 + \|\varepsilon(v)\|_{L^2(\Omega)}^2 \right) \quad (5.45)$$

Cette inégalité est assez puissante puisqu'elle contient, dans son membre de gauche, toutes les dérivées partielles de v , alors que son membre de droite ne fait intervenir que certaines combinaisons linéaires des dérivées partielles.

L'inégalité inverse de l'inégalité de Korn étant évidente, on en déduit que les deux membres définissent des normes équivalentes.

En mécanique, l'inégalité de Korn dit que l'énergie élastique (qui est proportionnelle à la norme du tenseur des déformations dans $L^2(\Omega)$) contrôle (i.e. est supérieure à) la norme du déplacement dans $H^1(\Omega)$ à l'addition près de la norme du déplacement dans $L^2(\Omega)$. Ce dernier point permet de prendre en compte les « mouvements de corps rigides », i.e. les déplacements u non nuls mais d'énergie élastique nulle.

Résumé des outils d'analyse fonctionnelle

Norme, produit scalaire, espaces

Soit E un espace vectoriel.

$\|\cdot\| : E \rightarrow \mathbb{R}_+$ est une norme sur E ssi elle vérifie :

1. $(\|x\| = 0) \implies (x = 0)$;
2. $\forall \lambda \in \mathbb{R}, \forall x \in E, \|\lambda x\| = |\lambda| \|x\|, \lambda \in \mathbb{K}$;
3. $\forall x, y \in E, \|x + y\| \leq \|x\| + \|y\|$ (inégalité triangulaire).

La distance issue de la norme est la distance définie par $d(x, y) = \|x - y\|$.

Toute forme bilinéaire symétrique définie positive $\langle \cdot, \cdot \rangle : E \times E \rightarrow \mathbb{R}$ est un produit scalaire sur E . Elle vérifie les propriétés :

1. bilinéarité : $\forall x, y, z \in E, \forall \lambda, \mu \in \mathbb{R}, \langle x, \lambda y + \mu z \rangle = \lambda \langle x, y \rangle + \mu \langle x, z \rangle$;
2. symétrie : $\forall x, y \in E \quad \langle x, y \rangle = \langle y, x \rangle$;
3. positivité : $\forall x \in E \quad \langle x, x \rangle \geq 0$;
4. définie : $\forall x \in E \quad (\langle x, x \rangle = 0 \implies x = 0)$.

La norme induite par le produit scalaire est $\|x\| = \sqrt{\langle x, x \rangle}$.

L'inégalité triangulaire donne alors l'inégalité de Cauchy-Schwarz : $|\langle x, y \rangle| \leq \|x\| \|y\|$.

Un espace vectoriel muni d'une norme est appelé espace normé.

Un espace vectoriel muni d'un produit scalaire est appelé espace préhilbertien, qui est donc également un espace normé pour la norme induite.

Exemple : Pour $E = \mathbb{R}^n$ et $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, on a les normes :

$$\|x\|_1 = \sum_{i=1}^n |x_i| \quad \|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2} \quad \|x\|_\infty = \sup_i |x_i|$$

Le produit scalaire défini par $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ a pour norme induite la norme $\|\cdot\|_2$.

Soit E un espace vectoriel et $(x_n)_n$ une suite de E . $(x_n)_n$ est une suite de Cauchy ssi $\forall \varepsilon > 0, \exists N, \forall p > N, \forall q > N : \|x_p - x_q\| < \varepsilon$.

Toute suite convergente est de Cauchy. La réciproque est fausse.

Un espace vectoriel est complet si et seulement si toute suite de Cauchy y est convergente.

Un espace de Banach est un espace normé complet.

Un espace de Hilbert est un espace préhilbertien complet.

Un espace euclidien est un espace de Hilbert de dimension finie.

Espaces fonctionnels

Un espace fonctionnel est un espace vectoriel dont les éléments sont des fonctions.

Dans la suite, nous considérons les fonctions définies sur un ouvert $\Omega \subset \mathbb{R}^n$ et à valeurs dans \mathbb{R} ou \mathbb{R}^p .

Une fonction u est mesurable ssi $\{x/|u(x)| < r\}$ est mesurable $\forall r > 0$.

On définit les espaces $\mathcal{L}^p(\Omega)$, pour $1 \leq p < \infty$ par :

$$\mathcal{L}^p(\Omega) = \left\{ u : \Omega \rightarrow \mathbb{R}, \text{ mesurable, et telle que } \int_{\Omega} |u|^p < \infty \right\}$$

$\mathcal{L}^p(\Omega)$ est la classe d'équivalence des fonctions de $\mathcal{L}^p(\Omega)$ pour la relation d'équivalence « égalité presque partout » : on confondra deux fonctions dès qu'elles sont égales presque partout, i.e. lorsqu'elles ne diffèrent que sur un ensemble de mesure nulle.

Les formes :

$$\|u\|_{L^p} = \left(\int_{\Omega} |u|^p \right)^{1/p}, 1 \leq p < \infty \quad \text{et} \quad \|u\|_{L^\infty} = \sup_{\Omega} |u|$$

ne sont pas des normes sur $\mathcal{L}^p(\Omega)$ (en effet, $\|u\|_{L^p} = 0$ implique que u est nulle presque partout dans $\mathcal{L}^p(\Omega)$ et non pas $u = 0$, d'où la définition des espaces $L^p(\Omega)$).

Ces formes sont des normes sur $L^p(\Omega)$ et en font des espaces de Banach (i.e. complets).

Dans le cas particulier $p = 2$, on obtient l'espace $L^2(\Omega)$ des fonctions de carré sommable pp. sur Ω . On remarque que la norme $\|u\|_{L^2} = \sqrt{\int_{\Omega} u^2}$ est induite par le produit scalaire $(u, v)_{L^2} = \int_{\Omega} uv$, ce qui fait de l'espace $L^2(\Omega)$ un espace de Hilbert.

Dérivée généralisée

Les éléments des espaces L^p ne sont pas nécessairement des fonctions très régulières. Leurs dérivées partielles ne sont donc pas forcément définies partout. C'est pourquoi on va étendre la notion de dérivation. Le véritable outil à introduire pour cela est la notion de distribution. Une idée simplifiée en est la notion de dérivée généralisée (certes plus limitée que les distributions, mais permettant de sentir les aspects nécessaires pour aboutir aux formulations variationnelles).

Fonctions tests

On note $\mathcal{D}(\Omega)$ l'espace des fonctions de Ω vers \mathbb{R} , de classe C^∞ , et à support compact inclus dans Ω . $\mathcal{D}(\Omega)$ est parfois appelé espace des fonctions-tests.

Théorème : $\overline{\mathcal{D}(\Omega)} = L^2(\Omega)$

Dérivée généralisée

Soit $u \in C^1(\Omega)$ et $v \in \mathcal{D}(\Omega)$. Par intégration par parties (ou Green) on a l'égalité :

$$\int_{\Omega} \partial_i u \varphi = - \int_{\Omega} u \partial_i \varphi + \int_{\partial\Omega} u \varphi n = - \int_{\Omega} u \partial_i \varphi$$

(car φ est à support compact donc nulle sur $\partial\Omega$).

Le terme $\int_{\Omega} u \partial_i \varphi$ a un sens dès que $u \in L^2(\Omega)$, donc $\int_{\Omega} \partial_i u \varphi$ aussi, sans que u ait besoin d'être C^1 . Il est donc possible de définir $\partial_i u$ même dans ce cas.

Cas $n = 1$: Soit I un intervalle de \mathbb{R} , pas forcément borné. On dit que $u \in L^2(I)$ admet une dérivée généralisée dans $L^2(I)$ ssi $\exists u_1 \in L^2(I)$ telle que $\forall \varphi \in \mathcal{D}(I)$, $\int_I u_1 \varphi = \int_I u \varphi'$.

Par itération, on dit que u admet une dérivée généralisée d'ordre k dans $L^2(I)$, notée u_k , ssi : $\forall \varphi \in \mathcal{D}(I)$, $\int_I u_k \varphi = (-1)^k \int_I u \varphi^{(k)}$.

Ces définitions s'étendent au cas où $n > 1$.

Théorème : Quand elle existe, la dérivée généralisée est unique.

Théorème : Quand u est de classe $C^1(\overline{\Omega})$ la dérivée généralisée est égale à la dérivée classique.

Espaces de Sobolev

Espaces H^m

L'espace de Sobolev d'ordre m est défini par :

$$H^m(\Omega) = \{u \in L^2(\Omega) / \partial^\alpha u \in L^2(\Omega), \forall \alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n \text{ tel que } |\alpha| = \alpha_1 + \dots + \alpha_n \leq m\}$$

On voit que $H^0(\Omega) = L^2(\Omega)$.

$H^m(\Omega)$ est un espace de Hilbert avec le produit scalaire et la norme induite :

$$(u, v)_m = \sum_{|\alpha| \leq m} (\partial^\alpha u, \partial^\alpha v)_0 \quad \text{et} \quad \|u\|_m = \sqrt{(u, u)_m}$$

Théorème : Si Ω est un ouvert de \mathbb{R}^n de frontière « suffisamment régulière », alors on a l'inclusion : $H^m(\Omega) \subset C^k(\overline{\Omega})$ pour $k < m - \frac{n}{2}$.

Trace

Dans les problèmes physiques que nous rencontrerons (voir partie 2), nous devons pouvoir imposer des conditions aux limites. Ceci est vrai, que l'on s'intéresse aux formulations forte ou faible.

Pour cela, il faut que la valeur d'une fonction sur la frontière soit définie. C'est justement ce que l'on appelle sa trace. La trace est le prolongement d'une fonction sur le bord de l'ouvert Ω .

De manière analogue, il est possible de prolonger la définition de la dérivée normale sur le contour de Ω , ce qui permet de prendre en compte des conditions aux limites de type Neumann par exemple.

Espace $H_0^1(\Omega)$

Soit Ω ouvert de \mathbb{R}^n . L'espace $H_0^1(\Omega)$ est défini comme l'adhérence de $\mathcal{D}(\Omega)$ pour la norme $\|\cdot\|_1$ de $H^1(\Omega)$.

Théorème : Par construction $H_0^1(\Omega)$ est un espace complet. C'est un espace de Hilbert pour la norme $\|\cdot\|_1$.

$H_0^1(\Omega)$ est le sous-espace des fonctions de $H^1(\Omega)$ de trace nulle sur la frontière $\partial\Omega$ (on a : $H_0^1(\Omega) = \ker \gamma_0$ où γ_0 est l'application trace).

Pour toute fonction u de $H^1(\Omega)$, on peut définir :

$$|u|_1 = \sqrt{\sum_{i=1}^n \|\partial_i u\|_0^2} = \sqrt{\int_{\Omega} \sum_{i=1}^n (\partial_i u)^2}$$

Inégalité de Poincaré : Si Ω est borné dans au moins une direction, alors il existe une constante $C(\Omega)$ telle que $\forall u \in H_0^1(\Omega) ; \|u\|_0 \leq C(\Omega)|u|_1$. On en déduit que $|\cdot|_1$ est une norme sur $H_0^1(\Omega)$, équivalente à la norme $\|\cdot\|_1$.

Corollaire : Le résultat précédent s'étend au cas où l'on a une condition de Dirichlet nulle seulement sur une partie de $\partial\Omega$, si Ω est connexe.

On suppose que Ω est un ouvert borné connexe, de frontière C^1 par morceaux. Soit $V = \{v \in H^1(\Omega) ; v = 0 \text{ sur } \Gamma_0\}$ où Γ_0 est une partie de $\partial\Omega$ de mesure non nulle. Alors il existe une constante $C(\Omega)$ telle que $\forall u \in V ; \|u\|_{0,V} \leq C(\Omega)|u|_{1,V}$, où $\|\cdot\|_{0,V}$ et $|\cdot|_{1,V}$ sont les norme et semi-norme induites sur V . On en déduit que $|\cdot|_{1,V}$ est une norme sur V équivalente à la norme $\|\cdot\|_{1,V}$.



PROBLÈME CONTINU

Chapitre 6

Problèmes physiques : équations différentielles et aux dérivées partielles

Résumé — La première partie était uniquement des « rappels » de notions mathématiques nécessaires pour disposer d'un certain nombre d'outils. Maintenant que nous avons ces outils, nous allons nous intéresser à des problèmes concrets de la physique (au sens général).

Cette deuxième partie va donc présenter des problèmes physiques, sous leur forme forte (dans ce chapitre), puis sous forme faible et variationnelle ; la théorie de la formulation faible étant exposée entre temps.

6.1 Introduction

Une équation différentielle est une relation entre une ou plusieurs fonctions inconnues et leurs dérivées. L'ordre d'une équation différentielle correspond au degré maximal de dérivation auquel l'une des fonctions inconnues a été soumise.

Pour les méthodes explicites de résolution des équations différentielles, on ira voir au chapitre C.

Histoire

Les problèmes posés ou menant à des équations différentielles sont aussi vieux que l'analyse elle-même (XVII^e-XVIII^e, voir notes précédentes sur les fonctions et sur la continuité et la dérivabilité). Avant même qu'on ait complètement élucidé la question des infiniment petits l'on se préoccupe déjà de résoudre des problèmes de tangente, qui mènent invariablement à une équation différentielle. Dès les débuts de la mécanique classique, dite newtonienne, on est confronté à l'intégration de systèmes d'équations différentielles du second ordre (voir exemples ci-dessous).

On s'habitue progressivement à ce que l'inconnue d'une équation puisse être une fonction, même si la fonction a encore à l'époque un statut flou.

On résout les équations différentielles par la méthode des séries entières sans s'encombrer des notions de convergence, bien que l'on entrevoit parfois certaines difficultés dues justement à ces problèmes de convergence... et on en arrive presque aux séries asymptotiques qui ne seront conceptualisées qu'au XIX^e siècle.

Un autre problème reste d'écrire, à l'aide de fonctions simples, les solutions des équations différentielles. La liste des équations différentielles que l'on sait résoudre de cette manière est plutôt maigre : Leibniz et sa méthode de décomposition en éléments simples, les Bernoulli et les équations différentielles linéaires du premier ordre, puis Riccati...

La découverte par Clairaut en 1734 de l'existence d'une solution singulière à l'équation $y - xy' + f(y') = 0$ (à la famille de droites $y = Cx + f(C)$ qui est l'expression générale des courbes intégrales, il faut adjoindre l'enveloppe de cette famille pour avoir toutes les solutions analytiques de l'équation) relance la dynamique. D'Alembert, en 1748, trouve un second cas d'intégrale singulière.



Riccati



Clairaut



D'Alembert



Euler

Mais c'est Euler et Lagrange qui élucident ce qui se passe en général en exhibant la courbe qui est le lieu des points singuliers.

Si l'on sait, dès la fin du XVII^e siècle intégrer les équations différentielles linéaires du premier et du second ordre à coefficients constants par des sommes d'exponentielles, il faut attendre 1760 pour que la théorie vienne à bout des équations différentielles linéaires à coefficients constants d'ordre quelconque. En 1739, Euler rencontre une équation différentielle linéaire à coefficients constants du quatrième ordre sur un problème de vibration des tiges qu'il ne sait pas intégrer. C'est en 1743 qu'il forme ce qu'on appelle aujourd'hui l'équation caractéristique, qu'il complète un peu plus tard lorsqu'une racine de cette équation polynomiale est multiple.

D'Alembert remarque que pour les équations différentielles non homogènes, l'adjonction d'une solution particulière à la solution générale de l'équation homogène donne la solution générale de l'équation homogène. Lagrange introduit la méthode de variation des constantes pour résoudre par quadratures l'équation linéaire non homogène lorsque l'on connaît la solution générale de l'équation homogène.



Clairaut, D'Alembert, Euler, Lagrange et Laplace consacrent de nombreux mémoires au problème des n corps. Le premier exemple du problème de Sturm-Liouville est donné par D'Alembert à propos de la vibration d'une corde non homogène.

Les mathématiciens du XVIII^e siècle admettent sans discussion l'existence de solutions des équations différentielles et des équations aux dérivées partielles sans chercher le domaine d'existence de ces solutions. Il faut attendre Cauchy, vers 1820, pour que soit abordée l'existence d'une solution à l'équation différentielle $y' = f(x, y)$, où f est supposée continument différentiable en chaque variable. Sans connaître les travaux de Cauchy, Lipschitz, en 1868, retrouve le résultat de Cauchy mais s'affranchit de l'hypothèse de différentiabilité de f au profit de la condition dite aujourd'hui de Lipschitz. En 1837, Liouville utilise, pour le cas particulier d'une équation linéaire du second ordre une méthode d'approximation successive : on construit une suite de fonctions convergeant vers la solution.

Si l'on considère une fonction f , sa dérivée f' exprime sa variation : positive f est croissante (et plus sa valeur est grande, plus la croissance est rapide), négative f est décroissante...

À partir de là on peut considérer une population de personnes. Le nombre de total de personnes à un instant t est donné par $f(t)$. Plus la population est nombreuse, plus elle se reproduit : en d'autres termes, la vitesse de croissance de la population est proportionnelle à la taille de la population. On peut donc écrire :

$$f'(t) = k f(t) \quad (6.1)$$

C'est une équation différentielle très simple, mais qui modélise le problème dit de dynamique de population.

Un autre problème simple, toujours en dynamique des populations, est celui de deux populations interdépendantes : les proies $g(t)$ et le prédateurs $m(t)$. Les proies se reproduisent et sont mangées par les prédateurs, alors que les prédateurs meurent sauf s'ils peuvent manger des proies.

Cela conduit au système de Lotka-Volterra :

$$\begin{cases} g'(t) = A g(t) - B g(t) m(t) \\ m'(t) = -C m(t) + D g(t) m(t) \end{cases} \quad (6.2)$$

où il est nécessaire de résoudre les équations simultanément (on dit que les équations sont couplées).

Pour revenir à des exemples plus connus des lecteurs, on n'oubliera pas que la relation fondamentale de la dynamique de Newton :

$$m\ddot{x} = f(x) \quad (6.3)$$

est une d'équation différentielle du second ordre. On pourrait multiplier les exemples comme l'équation de l'oscillation d'une masse suspendue à un ressort :

$$\ddot{x} = -c\dot{x} - \omega^2 x \quad (6.4)$$

avec ($c > 0$) ou sans ($c = 0$) frottement.

Comme on le voit sur les exemples ci-dessus, et comme on s'en souvient sans doute, il est nécessaire, pour déterminer complètement la solution d'une équation différentielle (ou d'une équation aux dérivées partielles), de disposer de valeurs de la solution en certains points. On appelle ces « contraintes », des conditions aux limites.

Histoire

Concernant les équations aux dérivées partielles, elles sont initialement résolues par des méthodes ad-hoc, le plus souvent géométriques. On n'écrit pas l'équation aux dérivées partielles linéaires. On ne s'étonnera donc pas que la première équation aux dérivées partielles n'apparaisse qu'assez tard.

La théorie de l'intégration des équations aux dérivées partielles du premier ordre ne date que de 1734. L'idée d'Euler est de ramener ladite intégration à celle des équations différentielles ordinaires. Euler montre qu'une famille de fonctions dépendant de deux paramètres vérifie une équation aux dérivées partielles du premier ordre en éliminant ces paramètres entre les dérivées partielles. Inversement, une telle équation admet une solution dépendant d'une fonction arbitraire. Il parvient ainsi à intégrer plusieurs équations de ce type.



Cauchy

Lipschitz

Dirichlet

Neumann

Lagrange, dans un mémoire de 1785, résume les connaissances de l'époque sur ces questions. Il ne sait intégrer que onze types d'équations aux dérivées partielles du premier ordre.

La solution de l'intégration de ces équations allait venir d'un mathématicien peu connu, mort en 1784, Paul Charpit. Son mémoire, présenté en 1784 à l'académie des sciences n'a jamais été publié et est resté longtemps une énigme. Une copie de ce mémoire a été trouvée en 1928.

6.2 Conditions aux limites

Comme nous venons de le dire, les conditions aux limites sont des contraintes (des valeurs) que l'on impose à la fonction solution (ou à certaines de ces dérivées) sur tout ou partie du domaine Ω , et/ou sur toute ou partie de sa frontière $\Gamma = \partial\Omega$. D'un point de vue « vocabulaire », certains types de conditions aux limites standard existent qui sont listées ci-après.

6.2.1 Dirichlet – valeurs aux bords

On parle de condition de Dirichlet imposée à une équation différentielle ou une équation aux dérivées partielles, lorsque l'on spécifie les valeurs que la solution doit vérifier sur toute ou partie de la frontière du domaine. Il peut s'agir par exemple d'un déplacement imposé (par exemple nul) en des points d'une structure (points d'appuis rigides).

6.2.2 Neumann – gradients aux bords

On parle de condition de Neumann imposée à une équation différentielle ou une équation aux dérivées partielles, lorsque l'on spécifie les valeurs des dérivées que la solution doit vérifier sur la frontière du domaine (flux, contraintes...)

6.2.3 Robin – relation gradient/valeurs sur le bord

On parle de condition de Robin ou condition de Fourier ou condition d'impédance imposée à une équation différentielle ou une équation aux dérivées partielles, lorsque l'on spécifie une relation linéaire entre les valeurs de la fonction et les valeurs de la dérivée de la fonction qui est solution du problème sur toute ou partie de la frontière du domaine. C'est donc une pondération de conditions de Dirichlet et de Neumann.

6.2.4 Condition aux limites dynamique

On parle de condition aux limites dynamique imposée à une équation différentielle ou une équation aux dérivées partielles, lorsque l'on spécifie une combinaison linéaire entre la dérivée temporelle et la dérivée normale que la solution doit vérifier sur toute ou partie la frontière du domaine.

6.2.5 Condition aux limites mêlée

On parle de condition aux limites mêlée lorsque l'on juxtapose plusieurs conditions aux limites différentes sur toute ou partie de la frontière du domaine.

6.3 Types d'équation aux dérivées partielles

La forme générale d'une équation aux dérivées partielles linéaire, scalaire, d'ordre 2 est :

$$au + c\nabla u + \operatorname{div}(A\nabla u) = f \quad (6.5)$$

où $a : \Omega \rightarrow \mathbb{R}$, $c : \Omega \rightarrow \mathbb{R}^n$, $A : \Omega \rightarrow \mathbb{R}^{n \times n}$ sont les coefficients de l'équation aux dérivées partielles linéaire. Dans le cas où u est scalaire ($n = 1$) et les coefficients sont constants, on obtient :

$$\alpha \frac{\partial^2 u}{\partial x^2} + \beta \frac{\partial^2 u}{\partial x \partial y} + \gamma \frac{\partial^2 u}{\partial y^2} + \delta \frac{\partial u}{\partial x} + \epsilon \frac{\partial u}{\partial y} + \gamma u = f \quad (6.6)$$

où $\alpha, \beta, \gamma, \delta, \epsilon, \gamma$ sont des scalaires. Cette équation est dite :

- elliptique si $\beta^2 - 4\alpha\gamma > 0$;
- parabolique si $\beta^2 - 4\alpha\gamma = 0$;
- hyperbolique si $\beta^2 - 4\alpha\gamma < 0$.

On peut dire que :

- les problèmes elliptiques vont concerner les problèmes stationnaires tels que ceux de la mécanique, la thermique, l'électrostatique, les membranes élastiques, l'écoulement potentiel ;
- les problèmes paraboliques vont concerner les modèles instationnaires tels que la diffusion thermique (équation de la chaleur), chimique, neutronique, les fluides visqueux incompressibles (on parlera d'irréversibilité, de décroissance, de principe du maximum, de propagation à vitesse infinie) ;
- les problèmes hyperboliques vont concerner les modèles instationnaires tels que la propagation des ondes, l'électromagnétisme, l'élastodynamique (on parlera de réversibilité, de conservation de l'énergie, de propagation à vitesse finie).

6.4 Phénomènes de propagation et de diffusion

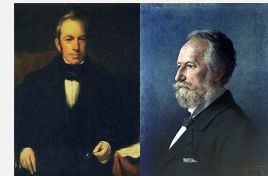
Dans ce paragraphe, nous avons regroupé les notions de propagation et de diffusion. La propagation et la diffusion sont liées. La diffusion, c'est en quelque sorte l'homogénéisation d'une grandeur dans un milieu et dans le temps. Or pour qu'il y ait diffusion, il faut bien que ledit phénomène se propage.

Nous nous placerons dans un cadre beaucoup plus macroscopique, où les interactions entre particules sont quantifiées au travers de variables macroscopiques continues, comme par exemple la température, la pression, mais également le frottement fluide, la conductivité thermique, la masse volumique... L'analogie de formulation mathématique nous fera également considérer les cas de propagation des ondes dans des milieux donnés. C'est d'ailleurs essentiellement via la propagation des ondes que nous aborderons les formulations... l'équation de la chaleur constituant le liant idéal pour illustrer les liens entre propagation et diffusion.

Histoire

Le déplacement des atomes, ions ou molécules dans un milieu, que celui-ci soit solide (cristallin ou amorphe), liquide ou gazeux, est appelé de manière générale « migration ».

Au sens large la diffusion désigne des transferts obéissant aux lois de Fick, i.e. dont la résultante macroscopique vérifie l'équation de diffusion. La turbulence entraîne ainsi une forte diffusion dans les fluides. La diffusion moléculaire est la migration sous l'effet de l'agitation thermique, à l'exception des autres phénomènes. Elle intervient par exemple dans des procédés d'amélioration des caractéristiques mécaniques (traitements de surface comme la nitruration ou cémentation), la résistance à la corrosion et les procédés d'assemblage par brasage.



Brown

Fick

En 1827, le botaniste Robert Brown observe le mouvement erratique de petites particules de pollen immergées dans de l'eau. Il ne s'agit pas d'un phénomène de diffusion, puisque ce qui bouge est une particule macroscopique, mais cette « marche aléatoire », que l'on appellera « mouvement brownien », servira de modèle pour la diffusion : Le mouvement brownien, ou processus de Wiener, est le mouvement aléatoire d'une « grosse » particule immergée dans un fluide et qui n'est soumise à aucune autre interaction que des chocs avec les « petites » molécules du fluide environnant. Il en résulte un mouvement très irrégulier de la grosse particule. Ce mouvement, qui permet de décrire avec succès le comportement thermodynamique des gaz (théorie cinétique des gaz), est aussi très utilisé dans des modèles de mathématiques financières.

En 1855, Adolph Fick propose des lois phénoménologiques, empiriques, inspirées de la loi de Fourier pour la chaleur (établie en 1822). C'est Albert Einstein qui démontrera les lois de Fick en 1905 avec ses travaux sur la loi stochastique. La première loi de Fick énonce que « le flux de diffusion est proportionnel au gradient de concentration ».

6.4.1 Équations de Laplace et Poisson

L'équation de Laplace, ou Laplacien est l'équation :

$$\Delta\varphi = 0 \quad (6.7)$$

Elle est elliptique. Elle apparaît dans de nombreux problèmes physiques : astronomie, électrostatique, mécanique des fluides, propagation de la chaleur, diffusion, mouvement brownien, mécanique quantique. Les fonctions solutions de l'équation de Laplace sont appelées les fonctions harmoniques. Toute fonction holomorphe est harmonique.

L'équation de Poisson est l'équation :

$$\Delta\varphi = f \quad (6.8)$$

où f est une fonction donnée. Elle est elliptique. Par exemple, en électrostatique, on exprime le potentiel électrique V associé à une distribution connue de charges ρ (dans le vide) par la relation :

$$\Delta V = -\frac{\rho}{\epsilon_0} \quad (6.9)$$

En gravitation universelle, le potentiel gravitationnel Φ est relié à la masse volumique μ par la relation :

$$\Delta\Phi = 4\pi G \mu \quad (6.10)$$

6.4.2 Équation d'onde, phénomènes vibratoires

Il s'agit encore d'une extension de l'équation de Laplace. L'équation d'onde est l'équation générale qui décrit la propagation d'une onde (sonore ou électromagnétique dont la lumière), qui peut être représentée par une grandeur scalaire ou vectorielle. Dans le cas vectoriel, en espace libre, dans un milieu homogène, linéaire et isotrope, l'équation d'onde s'écrit :

$$\Delta u = -\frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} \quad \text{ou} \quad \square u = 0 \quad (6.11)$$

où la fonction d'onde inconnue est notée $u(x, y, z, t)$, t représente le temps, et le nombre c représente la célérité ou vitesse de propagation de l'onde u (rappel : \square est le d'Alembertien). Elle est hyperbolique.

Dans le cas de l'acoustique, alors cette équation est obtenue à partir des équations de la mécanique ainsi que de l'hypothèse de fluide parfait :

- conservation de la masse : $\dot{\rho} + \text{div}(\rho u) = 0$ avec ρ la masse volumique du milieu et u sa vitesse ;
- conservation de la quantité de mouvement : $\text{div}(\sigma) = \rho \ddot{u}$, avec σ le tenseur des contraintes
- le fluide est parfait, et sa relation de comportement est donc : $\sigma = -p'I$ avec p' la pression dans le fluide.

À partir de ces trois équations, on retrouve l'équation des ondes sous la forme :

$$\Delta p' - \frac{1}{c^2} \ddot{p}' = 0 \quad (6.12)$$

avec c la célérité de l'onde acoustique dans le milieu (vitesse du son).

L'équation des ondes permet également de modéliser le problème de l'évolution de la déformation d'une membrane tendue (peau de tambour). Dans ce cas là, on note plutôt :

$$\sigma \Delta u = \rho \ddot{u} \text{ dans } \Omega \quad (6.13)$$

où Ω est notre domaine qui est une surface, donc $\Omega \subset \mathbb{R}^2$.

La même équation en dimension deux ou trois modélise la plupart des phénomènes vibratoires. En dimension 1, cette équation s'appelle l'équation des cordes vibrantes.

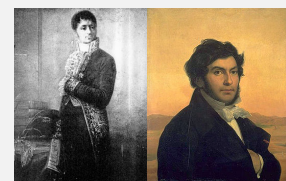
6.4.3 Équation de la chaleur

L'équation de la chaleur est une équation aux dérivées partielles parabolique, introduite au début du XIX^e siècle par Fourier pour décrire le phénomène physique de conduction thermique.

Histoire

En dépit des railleries de Hugo et Stendhal Joseph Fourier laissera son nom à la postérité à plus d'un titre :

en tant qu'égyptologue, une spécialité qu'il bouleverse à la suite de sa participation à l'expédition napoléonienne de 1798 (en 1810, il créera l'Université Royale de Grenoble, dont il deviendra le recteur, et y remarquera Jean-François Champollion. Ils sont enterrés au cimetière du Père-Lachaise à côté l'un de l'autre) ; en tant qu'homme politique puisqu'il est préfet d'Isère sous Napoléon Bonaparte et sous la Restauration ; en tant qu'administrateur, comme réorganisateur des statistiques françaises ; et enfin en tant que scientifique, la consécration de sa carrière étant son élection au poste de secrétaire perpétuel de l'Académie des Sciences. Pour tous les mathématiciens, il reste le fondateur de l'analyse de Fourier, qu'il crée au début du XIX^e siècle pour étudier mathématiquement la répartition de chaleur dans un corps conducteur.



Fourier Champollion

Quelques repères chronologiques : En 1807, Fourier écrit l'équation de la chaleur. En 1808, son mémoire est accepté par l'Académie des Sciences en dépit de sérieuses critiques concernant la rigueur des démonstrations (Fourier a en particulier une controverse avec Poisson). En 1822 a lieu la publication du *Traité analytique de la chaleur*, version remaniée et augmentée de son mémoire, qui s'imposera comme l'un des ouvrages scientifiques majeurs du dix-neuvième siècle.

La lecture du *Traité analytique* est toujours profitable. La première phrase sonne comme une profession de foi : « Les causes primordiales ne nous sont point connues ; mais elles sont assujetties à des lois simples et constantes, que l'on peut découvrir par l'observation, et dont l'étude est l'objet de la philosophie naturelle ». Fourier confesse donc son ambition de décrire des phénomènes physiques, jusque là jugés inaccessibles, par le moyen des mathématiques ; il le confirme dans le choix d'une devise en latin (attribuée par Fourier à Aristote, mais probablement inventée par lui-même) : *Et ignem regunt numeri* (Même le feu est régi par les nombres).

L'idée maîtresse de Fourier, celle que l'on lui attribue le plus naturellement, consiste à décomposer une fonction périodique arbitraire sous forme de série trigonométrique déjà mentionnée à l'équation (4.1). Des décompositions de ce style ne sont pas à proprement parler nouvelles : elles faisaient partie du répertoire des analystes de la fin du XVIII^e siècle, à commencer par le plus impressionnant d'entre eux, le Suisse Euler. Mais Fourier est le premier à avoir l'intuition du caractère universel et particulièrement commode de cette décomposition trigonométrique.

Fourier donne quelques exemples frappants pour montrer que « n'importe quelle » fonction est décomposable en série trigonométrique : une fonction « en dents de scie » et une autre « en créneaux ». Ainsi, même des fonctions discontinues peuvent être exprimées en séries de Fourier, pourvu que l'on définisse la fonction aux points de discontinuité en utilisant la « condition de demi-somme » :

$$f(x) = \frac{f(x^+) + f(x^-)}{2} \quad (6.14)$$

D'ailleurs, un théorème de Dirichlet dit que pour une fonction f définie sur $[-\pi, \pi]$ n'ayant qu'un nombre fini de discontinuités finies et ayant ailleurs une dérivée continue, son développement converge pour tout $x \in [-\pi, \pi]$, que somme vérifie la condition de demi-somme en tout point de $] - \pi, \pi[$, et qu'aux extrémités on a $f(\pm\pi) = (f(-\pi^+) + f(\pi^-))/2$.



Dirichlet

Cela pouvait paraître surprenant à l'époque de Fourier, mais bien sûr ce n'est pas contradictoire avec les théorèmes connus : une série de fonctions continues n'est pas en général continue. L'acte de foi suivant consiste à admettre que des décompositions similaires s'appliquent à des fonctions non périodiques (définies sur \mathbb{R} tout entier). Dans tous les cas, la formule donnant les coefficients de Fourier (4.2) conduit immédiatement à quelques questions, et notamment : comment trouver les coefficients ? (problème de l'analyse) En quel sens faut-il comprendre la convergence de ces séries ?

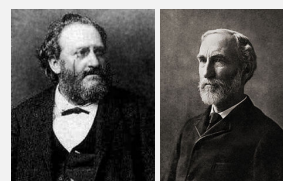
La convergence des séries de Fourier est reconnue dès le début comme un problème important et délicat, sur lequel Fourier lui-même avait été attaqué. Le résultat qui fait référence en la matière est le théorème de Dirichlet (1829) :

Théorème 37 — Théorème de Dirichlet. Soit $f : \mathbb{T} \rightarrow \mathbb{C}$, continue par morceaux, satisfaisant la condition de demi-somme (6.14) et C^1 de part et d'autre de chaque point de discontinuité, alors :

$$\forall x \in \mathbb{T}, \quad \sum_{k \in \mathbb{Z}} c_k(f) e^{2i\pi kx} = f(x) \quad (6.15)$$

Jusqu'en 1873 les spécialistes pensaient que l'hypothèse de continuité C^1 de part et d'autre était superflue ; mais dans son article *Eine neue Theorie der Convergenz und Divergenz von Reihen mit positiven Gliedern*, du Bois-Reymond donna un contre-exemple sous la forme d'une fonction continue périodique dont la série de Fourier diverge en un point. Si l'on veut avoir la propriété de reconstruction ponctuelle, on doit supposer une certaine régularité ; tout au mieux on peut relaxer l'hypothèse C^1 .

Même quand elle est vraie, la convergence ponctuelle des séries de Fourier peut être médiocre. Un exemple bien connu est le phénomène de Gibbs, qui pose des problèmes pratiques : la série de Fourier d'une fonction créneau présente des oscillations près des points de discontinuité, dont l'ampleur ne s'atténue pas quand le nombre de fréquences tend vers l'infini.



du Bois-Reymond Gibbs

Soit Ω un domaine de \mathbb{R}^3 de frontière $\Gamma = \partial\Omega$ et $T(x, t)$ un champ de température sur ce domaine (champ de scalaires). En présence d'une source thermique dans le domaine, et en l'absence

de convection, i.e. de transport de chaleur (i.e. on s'intéresse à la propagation de la température au sein d'un milieu « stable », i.e. qui ne bouge pas ; on ne s'intéresse pas à la propagation de chaleur due par exemple à l'existence d'un courant d'air, d'un courant de convection. La convection est plutôt un problème de mécanique des fluides.), l'équation de la chaleur s'écrit :

$$\forall x \in \Omega, \quad \frac{\partial T(x, t)}{\partial t} = D \Delta T(x, t) + \frac{f}{\rho c} \quad (6.16)$$

où :

- Δ est l'opérateur Laplacien ;
- D est le coefficient de diffusivité thermique (en m^2/s) ;
- f une éventuelle production volumique de chaleur (en W/m^3) ;
- ρ est la masse volumique du matériau (en kg/m^3) ;
- c la chaleur spécifique massique du matériau (en $\text{J}/\text{kg}\cdot\text{K}$).

L'équation de la chaleur est donc une équation de la forme :

$$\dot{u} - \Delta u = f \quad (6.17)$$

Elle est parabolique. Pour que le problème soit bien posé, il faut spécifier :

- une condition initiale :

$$\forall x \in \Omega, \quad T(x, 0) = T_0(x) \quad (6.18)$$

- une condition aux limites sur le bord du domaine, par exemple :
 - de Dirichlet :

$$\forall x \in \partial\Omega, \quad T(x, t) = 0 \quad (6.19)$$

- de Neumann :

$$\forall x \in \partial\Omega, \quad \frac{\partial T(x, t)}{\partial n} = n(x) \cdot \nabla T(x, t) = 0 \quad (6.20)$$

où $n(x)$ est le vecteur normal unitaire au point x .

L'équation de la chaleur, introduite initialement pour décrire la conduction thermique, permet de décrire le phénomène de diffusion. La diffusion est un phénomène de transport irréversible qui se traduit à terme par une homogénéisation de la grandeur considérée (par exemple la température dans un domaine, la concentration en produits chimiques dans une solution...). D'un point de vue phénoménologique, et au premier ordre, ce phénomène est régi par une loi de Fick (par exemple sorption d'eau dans les matériaux composites, diffusion d'actifs au travers de la peau). Dans l'équation précédente, T représente alors la répartition de la grandeur considérée (eau dans un composite, concentration d'un constituant chimique...) et le terme source dans Ω est souvent nul (i.e. $f = 0$).

Histoire

À propos de l'effet régularisant de la chaleur :

L'équation de la chaleur a un « effet régularisant » : quelque soit la distribution initiale de la température pour un problème donné, même discontinue, on aboutit rapidement à une distribution continue et même lisse de la température.

En 1956, Louis Nirenberg propose un nouveau problème à John Nash (qui est, selon Warren Ambrose, « aussi pressé que X, aussi exaspérant que Y, quels que soient X et Y ») : la continuité des solutions de l'équation de la chaleur dans les milieux discontinus, i.e. lorsque le coefficient de conductivité est quelconque, et peut même varier brutalement d'un point à l'autre et que l'on ne dispose seulement que des bornes inférieure et supérieure sur la conductivité.

John Nash résout le problème, notamment à l'aide du concept d'entropie de Boltzmann, mais n'obtiendra pas le triomphe escompté, le problème ayant été résolu, en même temps, mais par une autre méthode, par l'italien De Giorgi, qui lui deviendra célèbre.



Nirenberg Nash De Giorgi

Mais John Nash est célèbre pour bien d'autres travaux... dont l'un d'eux lui a valu le prix Nobel d'économie en 1994. On pourra regarder le film « un homme d'exception » (*A Beautiful Mind*) de Ron Howard en 2001 sur John Nash.

L'effet régularisant. Considérons l'équation de la chaleur (6.16) avec condition initiale $T_0(x)$ et $f = 0$, que nous réécrivons :

$$\begin{cases} \frac{\partial T}{\partial t} - \Delta T = 0 & \text{dans }]0; \infty[\times \mathbb{R}^N \\ T(x, 0) = T_0(x) & \text{dans } \mathbb{R}^N \end{cases}$$

Nous nous proposons de montrer que la solution T appartient en fait à $C^\infty(\mathbb{R}^N \times \mathbb{R}_*^+)$.

Soit $\hat{T}(\xi, t) = \hat{T}_0(\xi)e^{-|\xi|^2 t}$ la transformée de Fourier de T . Pour tout multi-indice α et pour tout $t > 0$, la fonction $(\xi, t) \mapsto |\xi|^\alpha \hat{T}(\xi, t)$ est continue en temps et à valeurs dans $L^2(\mathbb{R}^N)$. Par transformation de Fourier inverse, $\partial^\alpha T$ est donc un élément de $C(\mathbb{R}_*^+, L^2(\mathbb{R}^N))$, i.e. pour tout entier m , T est dans $C(\mathbb{R}_*^+, H^m(\mathbb{R}^N))$. Les injections de Sobolev nous permettent donc de dire que $T(t)$ appartient à $C(\mathbb{R}_*^+, C^\infty(\mathbb{R}^N))$. La même analyse tient pour $\partial^n T / \partial t^n$, et donc, on en déduit que $T \in C^\infty(\mathbb{R}_*^+, C^\infty(\mathbb{R}^N)) = C^\infty(\mathbb{R}^N \times \mathbb{R}_*^+)$.

6.5 Mécanique des fluides

La mécanique des fluides est l'une des deux composantes de la mécanique des milieux continus ; l'autre composante étant la mécanique des solides qui sera traitée un peu plus loin.

6.5.1 Équation de Navier-Stokes

Les équations de Navier-Stokes sont des équations aux dérivées partielles non linéaires qui décrivent le mouvement des fluides newtoniens (visqueux) dans l'approximation des milieux continus. Elles modélisent par exemple les mouvements de l'air de l'atmosphère, les courants océaniques, l'écoulement de l'eau dans un tuyau, et de nombreux autres phénomènes d'écoulement de fluides.

Nous considérons l'équation de Navier-Stokes sous la forme :

$$\rho \dot{u} + \rho(u \cdot \nabla)u - \eta \Delta u - \left(\zeta + \frac{\eta}{3}\right) \text{grad}(\text{div } u) = \rho f - \nabla p \quad (6.21)$$

où :

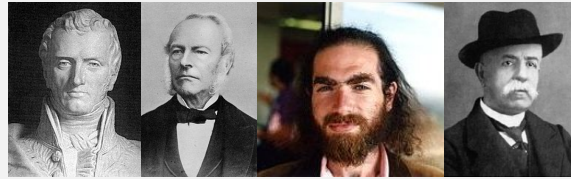
- u est la vitesse du fluide ;
- p est la pression dans le fluide ;
- ρ est la masse volumique du fluide ;
- η est la viscosité dynamique en cisaillement du fluide ;
- ζ est la viscosité dynamique en compression du fluide ;
- f est une force massique s'exerçant dans le fluide (par exemple : pesanteur).

Si de plus le fluide est incompressible (bonne approximation pour les liquides), alors $\text{div } u = 0$ et l'équation se simplifie :

$$\rho \dot{u} + \rho(u \cdot \nabla)u - \eta \Delta u = \rho f - \nabla p \quad (6.22)$$

Des 23 problèmes de Hilbert énoncés en 1900, la moitié est aujourd'hui complètement résolue, l'un a été démontré indécidable, cinq ne le sont pas du tout, les autres étant partiellement traités.

Des sept problèmes de l'institut Clay (ou problèmes du prix du millénaire, dotés chacun d'un million de dollars américains) énoncés en 2000, seule la conjecture de Poincaré a été démontrée en 2003 par Grigori Perelman (qui a refusé le prix, ainsi que la médaille Field. D'ailleurs il n'a pas publié son travail dans un journal, mais l'a rendu publique et gratuit sur internet...). Notons au passage que Grigori Perelman est analyste, i.e. un spécialiste de équations aux dérivées partielles, un sujet que beaucoup de topologues avaient l'habitude de regarder de loin. Ils ont changé d'avis depuis !

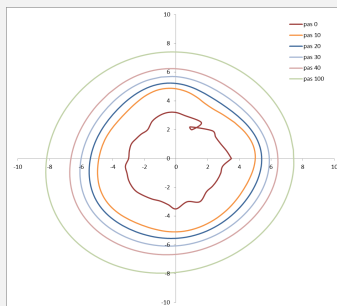


Navier

Stokes

Perelman

Ricci



Pour information, la preuve de Perelman suit le programme de Hamilton, qui le premier a pensé à utiliser le flot de Ricci pour s'attaquer à ce problème. Deux mots en aparté sur le flot de Ricci dont on entend beaucoup parler en ce moment...

Regardons se qui se passe dans le plan. On part d'une courbe (au centre de l'image ci-contre) qui est « biscornue » et peut même comporter des boucles. Si on la modifie de sorte que chacun de ses points est bougé proportionnellement à sa courbure, alors on obtient une seconde courbe, un peu plus lisse... et en itérant le procédé, on finit par retomber sur un cercle (i.e. sur une sphère en dimension 2), et donc on « voit » que la conjecture de Poincaré est vraie. En d'autres termes, le flot de Ricci permet d'homogénéiser la courbure de la courbe (de la variété dans le cas général).

Le problème des équations de Navier-Stokes figure parmi ces sept problèmes, c'est dire si le problème n'est pas simple. Et c'est, des sept problèmes du millénaire, le seul qui soit en lien direct avec la physique (en lien très direct...). De plus, l'Institut Clay n'a pas fixé comme objectif la démonstration de l'existence de solutions pour les équations de Navier-Stokes et la démonstration que les solutions, si elles existent, sont uniques. Il a « simplement » proposé de récompenser des « progrès essentiels » sur le sujet. Pourtant des calculs en mécanique des fluides sont effectués chaque jour, par exemple au sein de bureaux d'études pour dimensionner des voitures en aérodynamique.

On n'a pas pu prouver, pour ces équations, qu'il existe en toutes circonstances une solution parfaitement déterminée par les conditions initiales. Pour expliquer sommairement où réside la difficulté avec ces équations, on pourrait dire qu'avec les équations de Navier-Stokes, on passe du discret au continu.

Le problème c'est que l'on n'est pas sûr que ce qui est calculé est réellement une solution approchée et si le résultat est fidèle à la réalité : nous nous trouvons dans un cas où l'existence et l'unicité des solutions n'est pas facile à prouver. S'il existe des cas où des solutions régulières sont parfaitement déterminées à partir des conditions aux limites, cela n'est pas vrai en toutes circonstances et pour tous temps (on ne sait pas si une solution régulière ne peut pas devenir turbulente près un temps suffisamment long par exemple).

La non linéarité de ces équations joue des tours conduisant à des solutions qui peuvent par exemple ne pas se superposer sans s'influencer (alors que ce n'est pas le cas avec les équations linéaires). La viscosité n'est pas exempte de problème non plus, puisqu'elle conduit aux turbulences (que l'on pourrait qualifier de problème de passage d'échelle). D'ailleurs une description à un niveau intermédiaire entre une description atomiste (microscopique) et une description continue (macroscopique) a été faite par Boltzmann (niveau mésoscopique, description statistique), mais nous ne la présenterons pas ici. Les travaux de Cédric Villani (médaillé Fields 2010) ont répondu en grande partie au problème de la théorie cinétique (qui décrit un système constitué d'un très grand nombre de particules en interaction). Il a également prouvé des formes quantifiées du second principe de la thermodynamique : Boltzmann avait prouvé la croissance de l'entropie, et Cédric Villani en collaboration avec Laurent Desvillettes et Giuseppe Toscani a permis de quantifier la production d'entropie et le retour à l'équilibre. Ses travaux sur l'effet Landau (phénomène de relaxation non collisionnelle) ont des répercussions théoriques et pratiques, par exemple dans les modèles de mécanique classique en astrophysique.



Boltzmann Villani Desvillettes Toscani

La description faible du problème, qui sera présentée dans un autre chapitre, comporte elle-aussi ses problèmes car on n'est pas sûr de la convergence vers la solution forte.

Bien que des résultats d'existences et d'unicité existent, aucun ne permet soit de conclure à l'existence de solutions régulières pour tout temps, soit à l'apparition de singularité... On retrouve ici la notion de chaos, déterministe ou non, que nous n'aborderons pas dans ce document, mais pour lequel il est aisé de trouver de la documentation. Signalons quand même qu'il est identique de parler de dynamique non linéaire ou de théorie du chaos. On pourra donc se servir de ces deux points d'entrée pour toute recherche sur le sujet.

J'ajoute une dernière remarque prise dans le chapitre VIII «La géométrie des turbulences» de l'ouvrage *Les objets fractals* de Benoît Mandelbrot : «P.-S. 1989. J'ai précisé ces idées dans Mandelbrot 1976c [i.e. Géométrie fractale de la turbulence. Dimension de Hausdorff, dispersion et nature des singularités du mouvement des fluides. Comptes Rendus (Paris) : 282A, 119-120], en émettant les conjectures que les singularités des équations de Navier-Stokes et d'Euler sont des fractales. Ces conjectures paraissent être en bonne voie de se confirmer, au delà même de ce qui est dit au chapitre 11 de *Fractal Geometry of Nature*.»

6.5.2 Équation de Stokes

Il s'agit d'un cas particulier de l'équation de Navier-Stokes (termes inertiels absents ou négligés). Lorsqu'un fluide visqueux s'écoule lentement en un lieu étroit ou autour d'un petit objet, les effets visqueux dominent sur les effets inertiels. Son écoulement est alors appelé écoulement de Stokes. On parle aussi d'écoulement à faible nombre de Reynolds (le nombre de Reynolds mesure le poids relatif des termes visqueux et inertiel dans l'équation de Navier-Stokes). Ce nombre de Reynolds est alors beaucoup plus petit que 1.

L'équation de Stokes, qui décrit l'écoulement d'un fluide newtonien incompressible en régime permanent et à faible nombre de Reynolds, s'écrit dans Ω :

$$-\eta \Delta u + \nabla p = \rho f \quad (6.23)$$

On rappelle que si de plus le fluide est incompressible, alors on a également $\operatorname{div} u = 0$ dans Ω .

Contrairement à l'équation de Navier-Stokes, l'équation de Stokes est linéaire. Les écoulements solutions de cette équation possèdent par conséquent des propriétés bien particulières :

unicité : pour des conditions aux limites données (valeur de la vitesse au niveau des parois et/ou à l'infini), il existe un et un seul écoulement vérifiant l'équation de Stokes ;

additivité : les solutions de l'équation de Stokes vérifient le principe de superposition : si u_1 et u_2 sont solutions, alors toute combinaison linéaire $\lambda_1 u_1 + \lambda_2 u_2$ le sera aussi (ceci n'est pas incompatible avec la propriété d'unicité : seul l'écoulement vérifiant les bonnes conditions aux limites sera observé) ;

réversibilité : si un champ de vitesse u est solution de l'équation, alors $-u$ l'est aussi, à condition de changer le signe des gradients de pression, ainsi que des vitesses aux parois et à l'infini ; cette propriété est contraire à l'intuition, fondée sur notre expérience des écoulements macroscopiques : la réversibilité des écoulements à bas nombre de Reynolds a ainsi poussé les êtres vivants de très petite taille à développer des moyens de propulsion originaux.

paradoxe de Stokes : il faut prendre garde au fait que les solutions mathématiques de l'équation de Stokes, dans un cas donné ou dans certaines régions du domaine de solution, peuvent être physiquement fausses. Ceci est dû au « paradoxe de Stokes » à savoir que les conditions physiques permettant de ramener l'équation de Navier-Stokes à l'équation de Stokes ne sont pas nécessairement réalisées dans tout le domaine de solution, à priori. On aboutit alors à des solutions présentant des comportements potentiellement aberrants dans certaines limites. C'est le cas par exemple « à l'infini » où souvent le terme inertiel finit par l'emporter sur le terme visqueux, sans qu'on puisse le préjuger à priori.

6.5.3 Équation d'Euler

L'équation d'Euler établie en 1755 en appliquant le principe fondamental de la dynamique à une particule fluide (i.e. sous forme locale), s'applique dans le cas d'un fluide parfait, i.e. un fluide non visqueux et sans conductivité thermique. Le fluide peut être incompressible ou compressible (à condition, dans ce dernier cas, de se placer dans l'hypothèse de vitesses faibles). Il s'agit d'un cas particulier de l'équation de Navier-Stokes.

Toujours avec les mêmes notations, elle s'écrit dans $\Omega \times \mathbb{R}^+$:

$$\rho \dot{u} + \nabla p = \rho f \quad (6.24)$$

Et, si de plus le fluide est incompressible, alors on a également $\operatorname{div} u = 0$ dans Ω . Dans le cas où la description du problème est une description eulérienne (i.e. le champ de vitesse est associé à chaque point et est considéré de manière instantanée), on peut alors écrire :

$$\rho f - \nabla p = \rho \dot{u} = \rho (\dot{u} + (u \cdot \nabla)u) = \rho \left(\dot{u} + \frac{1}{2} \nabla u^2 + (\nabla \wedge u) \wedge u \right) \quad (6.25)$$

Bien que ces équations correspondent à une simplification de l'équation de Navier-Stokes, elles n'en sont pas plus stables... bien au contraire.

6.6 Équations de la mécanique des milieux continus des solides

La mécanique des milieux continus traite aussi bien de la déformation des solides que de l'écoulement des fluides. Ce dernier point a déjà été abordé auparavant aux paragraphes sur les équations de Navier-Stokes et de Stokes. Intéressons-nous maintenant au cas de la déformation des solides.

6.6.1 Notions générales conduisant aux équations de la mécanique

Comme ce document est initialement prévu pour un public d'ingénieurs mécaniciens, c'est dans cette section que se trouve du « matériel » qui aurait pu être présenté avant. Il va nous permettre de montrer que les équations de la mécanique comme celles de la chaleur, de Faraday, d'Ohm... viennent toutes d'une équation d'équilibre (ou de conservation) et d'une équation de flux (ou loi de comportement). Il est donné à titre de culture général.

Équation aux dérivées partielles elliptique linéaire

Si l'on considère le cas scalaire, i.e. celui d'une fonction inconnue u définie sur un domaine ouvert $\Omega \subset \mathbb{R}^2$, une équation aux dérivées partielles elliptique linéaire se met sous la forme :

$$a(x, y)u_{xx}(x, y) + b(x, y)u_{yy}(x, y) + c(x, y)u_x(x, y) + d(x, y)u_y(x, y) + eu(x, y) = f(x, y)$$

(6.26)

avec $a(x, y)b(x, y) > 0$.

Dans le cas où la fonction u est définie sur $\Omega \subset \mathbb{R}^n$, une équation aux dérivées partielles elliptique linéaire est définie par :

$$\mathcal{L}u(x) = f(x), \quad \forall x \in \Omega \quad (6.27)$$

où l'opérateur différentiel \mathcal{L} est défini par :

$$\mathcal{L}u = \sum_{i,j=1}^n a_{ij}(x)u_{x_i x_j}(x) + \sum_{i=1}^n b_i(x)u_{x_i}(x) + c(x)u(x) \quad (6.28)$$

avec les valeurs propres de la matrice $A = a_{ij}$ qui sont toutes non nulles et de même signe. On rappelle que la matrice A peut toujours être choisie symétrique du fait de la symétrie $u_{x_i x_j} = u_{x_j x_i}$ et donc que les valeurs sont réelles.

D'ailleurs, en dimension finie, le fait que toutes les valeurs propres sont strictement positives est équivalent au fait que la matrice A est définie positive.

Flux conservatif

Les équations aux dérivées partielles elliptiques conduisent à la notion physique de flux conservatif donné par un gradient. Considérons le cas scalaire. On associe à la fonction $u(x)$ un flux $q(x)$ qui est nul vers l'extérieur :

$$\int_{\partial\omega} q(x) \cdot n_x ds = 0, \quad \forall \text{ volume } \omega \subset \Omega \quad (6.29)$$

En utilisant la formule de divergence-flux, il vient : $\int_{\omega} \operatorname{div}_x q(x) dx = 0$, et en supposant la fonction $q(x)$ suffisamment régulière on en déduit $\operatorname{div}_x q(x) = 0, \forall x \in \Omega$.

En supposant que ce flux est une fonction linéaire du gradient ∇u , et qu'il est orienté dans la direction opposée (physiquement, les flux se font souvent de façon opposée au gradient d'une grandeur), on écrit $q(x) = -a(x)\nabla u$ et on obtient une loi de conservation à l'équilibre du type :

$$\operatorname{div}(a(x)\nabla u(x)) = 0 \quad (6.30)$$

Pour $a(x) \equiv 1$, on retrouve la plus simple des équations elliptiques, l'équation de Laplace :

$$\operatorname{div} \nabla u(x) = \Delta u(x) = 0 \quad (6.31)$$

Équilibre des lois de conservation en comportement linéaire

Le raisonnement précédent peut être mené en considérant les lois de conservations dans les milieux continus. Les ingrédients principaux qui vont nous conduire à une grande famille d'équations elliptiques sont les suivants :

- Conservation d'une grandeur ;
- Milieu immobile et régime stationnaire ;
- Loi de comportement linéaire entre le flux de cette grandeur et le gradient.

Les deux premières notions sont souvent associées à l'équilibre d'un système.

Lois de conservation Soit $\Omega \subset \mathbb{R}^n$, $n = 1, 2$ ou 3 . Pour tout sous-domaine $\omega(t) \subset \Omega$, on définit, d'une manière générale, une grandeur $\mathcal{U}(t)$ à partir de sa densité spécifique $u(x, t)$ à l'aide de la masse volumique du milieu ρ par :

$$\mathcal{U}(t) = \int_{\omega(t)} \rho(x, t) u(x, t) dx \quad (6.32)$$

On suppose que la variation de cette grandeur $\mathcal{U}(t)$ ne peut se faire que par un apport extérieur volumique noté $\int_{\omega(t)} \varphi u dv$ et/ou un apport extérieur surfacique $\int_{\partial\omega(t)} J_{\mathcal{U}} ds$.

Le Lemme de Cauchy affirme alors l'existence d'un tenseur $j_{\mathcal{U}}$ tel que $J_{\mathcal{U}} = j_{\mathcal{U}} \cdot n$ et que l'on a la loi de conservation locale :

$$\frac{\partial}{\partial t}(\rho u) + \operatorname{div}(\rho u v) = \varphi u - \operatorname{div} j_{\mathcal{U}} \quad (6.33)$$

Cette équation de conservation local peut, en utilisant la conservation de la masse, se mettre sous la forme :

$$\rho \frac{Du}{Dt} = \varphi u - \operatorname{div} j_{\mathcal{U}} \quad (6.34)$$

Hypothèse de milieu immobile et de régime stationnaire Si le milieu est immobile, la vitesse eulérienne (particulaire), $v(x, t)$ est nulle et donc la dérivée particulaire se réduit à :

$$\rho \frac{Du}{Dt} = \rho \frac{\partial u}{\partial t} + \nabla u(x, t) \cdot v(x, t) = \rho \frac{\partial u}{\partial t} \quad (6.35)$$

L'équation de conservation précédente devient alors :

$$\rho \frac{\partial u}{\partial t} = \varphi u - \operatorname{div} j_{\mathcal{U}} \quad (6.36)$$

Si de plus, on est en régime stationnaire, i.e. $\rho \frac{\partial u}{\partial t} = 0$, alors :

$$\varphi u - \operatorname{div} j_{\mathcal{U}} = 0 \quad (6.37)$$

Hypothèse de loi de comportement linéaire On va maintenant introduire dans l'équation précédente la loi de comportement du milieu. Le choix le plus simple est celui liant de manière linéaire le flux et le gradient :

$$j_{\mathcal{U}}(x, t) = A(x, t) \nabla u(x, t) \quad (6.38)$$

En reportant cette loi dans l'équation de conservation, il vient $\operatorname{div} A(x, t) \nabla u(x, t) = \varphi u$, soit :

$$A \Delta u + \nabla A \nabla u = \varphi u \quad (6.39)$$

Cette équation est de type elliptique dès que l'opérateur $A(x)$ possède de bonnes propriétés de positivité.

Équilibre général

Pour résumer, les modèles elliptiques se mettent sous la forme :

— Équation de flux (ou loi de comportement dans le cas vectoriel) :

$$j = -A(x) \nabla u \quad (6.40)$$

— Équation d'équilibre ou de conservation :

$$\operatorname{div} j = f - c(x)u \quad (6.41)$$

où :

- la fonction scalaire inconnue $u : \Omega \in \mathbb{R}^n \rightarrow \mathbb{R}$ est appelée un potentiel ;
- la fonction vectorielle inconnue $j : \Omega \in \mathbb{R}^n \rightarrow \mathbb{R}^n$ est appelée un flux ;
- la fonction scalaire connue $f : \Omega \in \mathbb{R}^n \rightarrow \mathbb{R}$ correspond au termes sources du potentiel ;
- les fonctions scalaires connues $A(x)$ et $c(x)$ sont les données du problème.

Dans le cas vectoriel :

- u devient une fonction vectorielle $\Omega \in \mathbb{R}^n \rightarrow \mathbb{R}^n$;
- j un tenseur d'ordre 2 ;
- $A(x)$ un tenseur d'ordre 4 ;
- et $c(x)$ une fonction vectorielle.

Retour sur les exemples précédents

Si $u \rightarrow T$ est la température, $j \rightarrow q$ le flux de chaleur, f la source de chaleur, $A(x) \rightarrow \kappa(x)$ la conductivité du matériaux et $c(x) \equiv 0$, on retrouve l'équation de la chaleur.

Si $u \rightarrow V$ est le potentiel électrostatique, $\nabla u \rightarrow E$ le champ électrostatique, $j \rightarrow D$ le déplacement électrique ou flux de densité de courant, $f \rightarrow \rho$ la densité de charge, $A(x) \rightarrow \epsilon(x)$ le tenseur diélectrique et $c(x) \equiv 0$, on retrouve la loi de Faraday.

Si $u \rightarrow V$ est le potentiel électrique, $\nabla u \rightarrow E$ le champ électrique, $j \rightarrow J$ la densité de courant, $f \equiv 0$, $A(x) \rightarrow \sigma(x)$ le tenseur de conductivité et $c(x) \equiv 0$, on retrouve la loi d'Ohm.

Si $u \rightarrow C$ est la concentration, j le flux molaire, $f \rightarrow \rho$ le taux d'absorption ou le taux de réaction, $A(x) \rightarrow D(x)$ la constante de diffusion et $c(x)$ la coefficient de réaction, alors on retrouve le cas de la diffusion moléculaire.

Si u est le déplacement orthogonal, $\nabla u \rightarrow \epsilon$ la déformation, $j \rightarrow \sigma$ la contrainte dans le plan, f la force volumique extérieure, $A(x) \rightarrow H(x)$ le tenseur des rigidités et $c(x) = 0$, on se retrouve dans le cas de l'élasticité plane ($n = 2$) linéaire.

6.6.2 Formulation générale

En repartant du second principe de la dynamique, et en introduisant un terme de dissipation visqueuse, on écrira le problème général sous la forme :

$$\begin{cases} \operatorname{div} \sigma + f = \rho \ddot{u} + \mu \dot{u} & \text{dans } \Omega & \text{équation de la dynamique} \\ \sigma = H(\epsilon - \epsilon_{th}) & \text{dans } \Omega & \text{loi de comportement} \end{cases} \quad (6.42)$$

où f est une force de volume, ρ la masse volumique, μ est un facteur traduisant l'amortissement visqueux, ϵ_{th} est la déformation d'origine thermique et H traduit la relation de Hooke, i.e. la loi de comportement. Les inconnues sont donc les déplacements u , les déformations ϵ et les contraintes σ .

Nous allons maintenant faire quelques remarques concernant ces équations, notamment des simplifications correspondant à certains cas courant d'étude.

6.6.3 Dynamique / statique

La formulation générale est celle de la dynamique avec amortissement.

Une structure est sans amortissement si $\mu = 0$.

Le problème est statique s'il ne dépend pas du temps, i.e. s'il n'y a plus les termes \ddot{u} et \dot{u} .

6.6.4 Grands / petits déplacements

La relation entre déformation et déplacement est donnée par :

$$\varepsilon = \frac{1}{2} (\nabla u + (\nabla u)^T) \text{ soit pour chaque composante : } \varepsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} + \frac{\partial u_k}{\partial x_i} \frac{\partial u_k}{\partial x_j} \right) \quad (6.43)$$

Dans le cas des petits déplacements (termes d'ordre 2 négligés), elle devient :

$$\varepsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \text{ que l'on note } \varepsilon_{ij} = \frac{1}{2} (u_{i,j} + u_{j,i}) \quad (6.44)$$

Dans le cas de grandes déformations, on décompose les différentes grandeurs en deux parties : l'une symétrique, l'autre antisymétrique. On ne parle plus de contraintes de Cauchy (qui est l'application linéaire reliant le vecteur contrainte à la normale unitaire au point considéré), mais de contraintes nominales et de contraintes de Piola-Kirchhoff de seconde espèce (plus de contrainte corotationnelle pour les poutres ou les coques). Pour les déformations, on remarque qu'elles sont le produit d'une rotation pure par une déformation pure... Ces tenseurs sont présentés au paragraphe 20.1.1, quant aux grandes déformation, elles sont exposées au paragraphe 20.2.

6.6.5 Loi de comportement

La loi de comportement, reliant contrainte et déformation, peut être relativement compliquée selon les cas, et varie en fonction du niveau de chargement. On se reportera alors au paragraphe 20.3. Si elle est linéaire, on parle alors d'élasticité linéaire. La figure 6.1 présente quelques lois types de comportement pour différents matériaux.

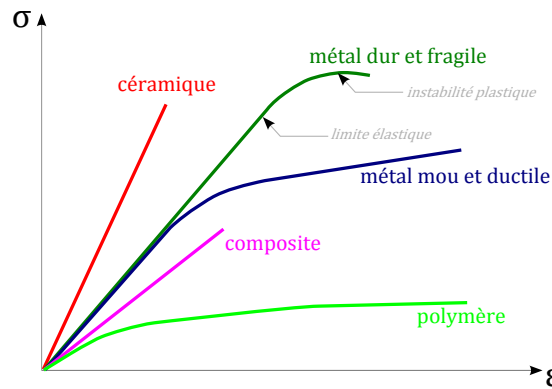


FIGURE 6.1: Quelques lois types de comportement pour différents matériaux.

Dans le cas de matériaux isotropes (i.e. possédant les mêmes propriétés matérielles quelque soit l'orientation), les relations constitutives s'écrivent à l'aide des coefficients de Lamé λ et μ :

$$\sigma = 2\mu\varepsilon + \lambda \operatorname{tr}(\varepsilon)I \quad (6.45)$$

i.e. $\sigma_{ij} = 2\mu\varepsilon_{ij} + \lambda\varepsilon_{kk}\delta_{ij}$. On rappelle également que :

$$\lambda = \frac{Ev}{(1-2\nu)(1+\nu)} \quad \text{et} \quad \mu = \frac{E}{2(1+\nu)} \quad (6.46)$$

avec E le module d'Young et ν le coefficient de Poisson, que l'on exprime, inversement :

$$E = \frac{\mu(3\lambda + 2\mu)}{\lambda + \mu} \quad \text{et} \quad \nu = \frac{\lambda}{2(\lambda + \mu)} \quad (6.47)$$

Pour assurer que la matrice H est définie positive, il faut que $0 < \nu < 0,5$.

Le cas $\nu = 0,5$ correspond au cas d'un matériau incompressible, comme par exemple le caoutchouc. Nous verrons que selon les formulations il est plus ou moins aisé de prendre en compte ces matériaux. De plus, la littérature ne manque pas sur le sujet.

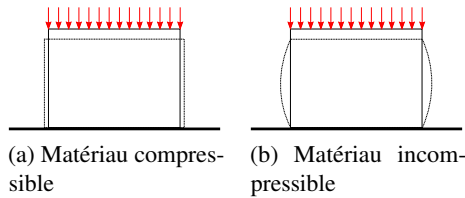


FIGURE 6.2: Déformation et coefficient de Poisson

La signification physique du coefficient de Poisson est simple : 2ν traduit le pourcentage de déformation dans le sens transverse à l'effort. En restant dans le cas plan, quand on fait subir à une pièce un effort de compression telle qu'elle subit une déformation (de compression) de 1, alors elle s'élargit de 2ν dans la direction transverse (ν de chaque côté par symétrie).

Ainsi, il devient évident qu'un matériau incompressible est caractérisé par $2\nu = 1$, i.e. lorsqu'il perd un volume 1 selon une direction, il reprend un volume 1 dans la direction transverse : en d'autres termes, il ne varie pas de volume, il est donc incompressible [ce qui ne veut pas dire « incompressible » !].

Attention, cette signification physique du coefficient de Poisson n'est exacte que dans le cas de matériaux non seulement isotropes mais pleins. Pour un assemblage de matériaux ou de géométries, cette relation est mise en défaut, tout comme pour des matériaux orthotropes.

Si l'on considère un matériau constitué d'alvéoles (de type hexagonal), mais dont les pointes sont rentrantes et non sortantes (figure 6.3), alors le coefficient de Poisson macroscopique de ce matériau est négatif : quand on le comprime, il ne gonfle pas mais rétrécit.

On parle alors d'un matériau auxétique.

On voit bien que dans un tel cas, la différence de comportement n'est due qu'à l'effet structural, qu'à l'assemblage. Le matériau constituant les alvéoles lui, peut (et est) tout à fait conventionnel : on trouve ce genre de structures en carton, en aluminium et en acier, qui sont tous des matériaux homogènes isotropes.

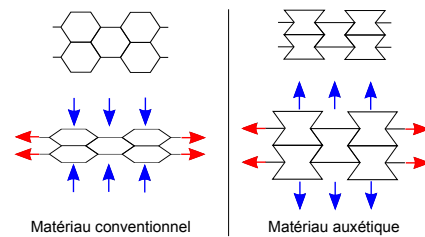


FIGURE 6.3: Matériau auxétique

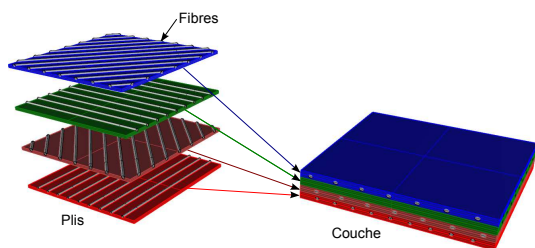


FIGURE 6.4: Matériau composite

Par exemple, la rigidité globale d'un assemblage de couches de composites renforcés de fibres, comme montré à la figure 6.4 peut conduire, selon les orientations des plis, à des valeurs du coefficient de Poisson très variables, d'autant plus que celles-ci sont fonctions du repère dans lequel elles sont données (repère global, repère des fibres...)

Un stratifié unidirectionnel T300 fibre de carbone/résine époxyde 934 a un coefficient de Poisson de seulement $\nu_{12} = 0,04$.

Un stratifié unidirectionnel T300 fibre de carbone/résine N5208 a un coefficient de Poisson de $\nu_{12} = 0,3$ dans le sens des fibres, mais de $\nu_{12} = -0,77$ à 45° .

En utilisant du verre ($E = 69 \text{ GPa}$, $\nu = 0,25$) et de la résine époxyde ($E = 3,5 \text{ GPa}$, $\nu = 0,4$), on peut réaliser un stratifié unidirectionnel verre-époxyde ayant pour propriétés $E_{11} = 45 \text{ GPa}$, $E_{tt} = 12 \text{ GPa}$, $G_{lt} = 4,5 \text{ GPa}$, $\nu_{lt} = 0,3$ pour une fraction volumique de verre de 60%, et un stratifié $[0/90]$ ayant pour propriétés $E_x = E_y = 20 \text{ GPa}$, $G_{xy} = 2,85 \text{ GPa}$ et $\nu_{xy} = 0,13$ pour une fraction volumique de verre de 50%.

Notons également que si $\mu \neq 0$ et $3\lambda + 2\mu \neq 0$, alors on peut inverser la relation contrainte/dé-

placement :

$$\varepsilon_{ij} = \frac{1}{2\mu} \left(\sigma_{ij} - \frac{\lambda}{3\lambda + 2\mu} \text{tr}(\sigma) \delta_{ij} \right) \quad (6.48)$$

Nous reviendrons plus tard sur ces aspects non linéaires au chapitre 20.

6.7 Équations de l'acoustique

L'acoustique, de manière la plus générale, a pour objet l'étude des sons et des ondes mécaniques. Elle fait appel aux phénomènes ondulatoires et à la mécanique vibratoire.

C'est là tout « le problème » de la vibro-acoustique : elle met en jeu des échelles d'énergies très différentes. Considérons l'étude vibro-acoustique d'un véhicule. Alors, on doit aussi bien prendre en compte les effets mécaniques ayant une forte énergie (la rotation du moteur fait bouger tout le véhicule), que des effets ayant une énergie comparativement très faible : la composante aérienne du bruit. Dans le premier cas, vous êtes face à quelque chose capable de faire « sauter » une voiture sur place, alors que dans le second vous êtes face au déplacement d'une feuille de papier lorsque vous criez devant elle !

Il est clair qu'il sera nécessaire d'étudier chaque aspect :

- solidien/solidien : il s'agit d'une étude mécanique « classique », bien que l'on se trouve dans le cas d'une sollicitation qui dépend du temps. Celle-ci est d'ailleurs généralement périodique, d'où l'utilisation d'outils appropriés (en fréquence). On est dans le cas de propagation solidienne d'onde ;
- aérien/aérien : il s'agit de la propagation d'une onde acoustique de sa source jusqu'au récepteur. On est en présence d'un modèle purement de propagation d'onde ;
- solidien/aérien : il s'agit du cas du rayonnement. Une vibration mécanique fait vibrer un élément qui se met à émettre une onde dans l'air ;
- aérien/solidien : dans le cas de sources acoustiques très puissantes, l'onde acoustique se propageant dans l'air peut arriver à déformer et à faire vibrer une surface physique (par exemple un tablier de voiture).

Les deux derniers cas sont typiquement des cas dits de couplage fluide/structure.

D'un point de vue de la formulation de ce problème sous forme d'équation aux dérivées partielles, tout a été dit précédemment. Nous verrons plus loin, aux chapitres 18 et surtout 19, comment prendre en compte tous ces aspects de manière plus appliquée.

6.8 Multiplicateurs de Lagrange

Les multiplicateurs de Lagrange ont une grande utilité, car leurs applications sont vraiment très larges. Pour rester très général, on va dire qu'ils permettent de « faire des liens » entre des grandeurs. Nous allons expliquer cela sur quelques exemples.

Si l'on reprend les équations de l'élasticité, on peut les réécrire en fonction uniquement des déplacements. L'avantage est que l'on dispose qu'un problème qui ne dépend que d'un unique champ inconnu, pour lequel les conditions aux limites correspondent à des choses très physiques : les déplacements imposés (appui simple, encastrement, contact).

Lorsque l'on résout un problème avec uniquement le champ de déplacements, alors évidemment, on obtient le champ de déplacements, mais uniquement lui ! Alors comment remonter aux autres grandeurs ? En utilisant les relations existantes... mais : entre déplacement et déformation, on a une dérivation, qui peut faire perdre en qualité d'approximation, d'autant plus lorsque l'on néglige les termes du second ordre. Ensuite on se sert des relations entre déformation et contraintes, mais encore faut-il être capable de mesurer ces grandeurs... et même si l'on sait le faire, que se passe-t-il à l'interface entre deux matériaux distincts, comme par exemple entre l'une des peaux et l'âme

d'un matériau sandwich ? (sur le sujet, abordé en plusieurs fois dans ce document, voir synthèse au paragraphe 12.3)

Le cas du calcul des contraintes à l'interface entre deux matériaux est un exemple typique d'application des multiplicateurs de Lagrange. Si l'on considère chacun des deux matériaux comme indépendant, alors au sein de chaque matériau, les relations fonctionnent. Si l'on impose maintenant la continuité des déplacements à l'interface entre les deux matériaux (en écrivant que λ (qui représente les multiplicateurs de Lagrange) multiplié par la différence entre les déplacements de part et d'autre de l'interface est nul : analogie immédiate avec les résidus pondérés – voir chapitre suivant), alors on s'aperçoit (ce sera détaillé dans le chapitre sur les formulations variationnelles) que les multiplicateurs de Lagrange s'interprètent comme les composantes des contraintes qui doivent être continues à l'interface (on parle de trace... comme défini au chapitre sur les espaces de Sobolev).

Si maintenant l'un des domaines est l'air et l'autre la structure considérée, alors on se doute que les multiplicateurs de Lagrange peuvent servir à « faire le lien » entre ces deux domaines, à les coupler en écrivant une équation liant la pression dans le fluide aux efforts de surface sur la structure par exemple... On voit comment on peut résoudre les problèmes évoqués dans le court paragraphe sur l'acoustique.

Mais les multiplicateurs de Lagrange peuvent également être utilisés directement afin d'améliorer les méthodes numériques : ils permettent par exemple de « passer » les informations entre des maillages incompatibles, permettant par exemple de circonscrire une zone de maillage raffiné (avec raffinement automatique en plus si l'on veut) et une zone de maillage plus léger...

Une illustration de leur utilisation pour imposer des conditions aux limites en déplacement est donnée en 12.6.

Les multiplicateurs de Lagrange peuvent être introduits partout où l'on veut créer un lien entre des grandeurs, et cela sans se soucier de leur signification physique. Toutefois, dans de nombreux cas, il se trouve que ces multiplicateurs représentent certaines grandeurs, permettant une analogie entre différentes formulations d'un même problème.

Les multiplicateurs de Lagrange seront également repris au chapitre 16 sur l'optimisation, où la notion de Lagrangien, abordée au paragraphe 7.6, sera beaucoup plus détaillée.

Chapitre 7

Formulations faible et variationnelle

Résumé — Un problème physique est généralement décrit par la donnée d'équations différentielles ou plus certainement aux dérivées partielles. Une telle formulation est appelée formulation forte du problème.

Nous allons voir qu'il est possible d'exprimer ces équations différentielles ou équations aux dérivées partielles d'une manière « moins contraignante » pour les solutions recherchées. Une telle formulation sera qualifiée de formulation faible, et ses solutions appelées solutions faibles.

Évidemment, une solution forte du problème d'origine est également solution de la formulation faible.

7.1 Principe des formulations faible et variationnelle

Quand on cherche la solution d'une équation aux dérivées partielles, on peut être confronté aux problèmes suivants :

- La solution ou les coefficients de l'équation aux dérivées partielles linéaires peuvent ne pas être assez réguliers pour donner un sens classique à l'équation ;
- La solution peut être régulière mais l'espace de régularité en question n'a pas les bonnes propriétés pour obtenir directement l'existence de la solution dans cet espace.

On va donc essayer de donner un sens plus faible à la notion de solution, sans perdre de vue si possible les notions les plus naturelles. Donner un sens faible à une équation aux dérivées partielles signifie :

1. Chercher une solution dans un espace de régularité plus faible que souhaité ;
2. Établir, en général à l'aide de fonction tests et d'intégrations par parties « formelles », une série d'équations que doit vérifier cette solution. Cette série d'équations doit être construite (si possible) de telle sorte que les éventuelles solutions régulières soient aussi solutions faibles et, *a contrario*, que des solutions faibles qui seraient régulières soient aussi solutions classiques du problème initial.

Les difficultés d'une telle approche sont nombreuses :

- Le choix de l'espace fonctionnel dans lequel on va chercher la solution est crucial, pas toujours unique et parfois non trivial.
- Si l'on affaiblit trop la notion de solution, il sera *a priori* plus facile de démontrer l'existence mais les propriétés d'unicité seront plus difficiles à obtenir.
- Si l'on n'affaiblit pas suffisamment l'équation, l'existence sera ardue à prouver mais l'unicité sera plus simple.

Il s'agit donc de trouver un juste équilibre...

Dans le cas des équations elliptiques, celles-ci ont pour propriété, en général, de mettre en jeu des dérivées d'ordre pair de la solution dans le terme principal de l'équation. L'idée première pour résoudre ces problèmes aussi bien d'un point de vue théorique que numérique va être de multiplier

formellement l'équation par une fonction test régulière, puis d'intégrer sur le domaine et enfin d'intégrer par parties un nombre suffisant de fois pour faire porter autant de dérivées sur la solution que sur les fonctions tests. On pourra ainsi envisager d'utiliser le même espace fonctionnel pour la solution et les fonctions tests.

Définition 54 — Formulations forte et faible. Étant donné un opérateur différentiel $A(\cdot)$ et une fonction f définie sur un domaine ouvert Ω , la formulation forte du problème est la suivante :

$$\text{Trouver la fonction } u, \text{ définie sur } \Omega, \text{ vérifiant } A(u) = f \text{ en tout point de } \Omega. \quad (7.1)$$

Une solution u du problème précédent est alors également solution du problème suivant appelé formulation faible (On appelle v une fonction test) :

$$\text{Trouver une fonction } u, \text{ définie sur } \Omega, \text{ vérifiant } \int_{\Omega} A(u)v = \int_{\Omega} f v \quad \forall v \text{ définie sur } \Omega. \quad (7.2)$$

Justification. Tout réside dans le fait de « multiplier par une fonction test et d'intégrer ». Pourquoi peut-on faire ça ?

Partons de l'équation forte :

$$A(u) = f, \quad \forall x \in \Omega \quad (7.3)$$

Nous pouvons multiplier les deux membres de l'équation par une fonction v à condition que cette fonction v ne soit pas identiquement nulle. Dans ce cas, la solution de

$$A(u)v = f \cdot v, \quad \forall x \in \Omega \quad (7.4)$$

est la même $\forall v$ suffisamment sympathique. Comme cette seconde équation est vraie en tout point de Ω , alors elle est encore vraie sous forme intégrale, i.e. sous forme faible, à condition que v soit suffisamment régulière pour pouvoir être intégrée.

On entrevoit alors clairement la suite des opérations : si en plus d'être régulière, la fonction v est différentiable, alors on va pouvoir réaliser sur le terme $\int A(u) \cdot v$ une intégration par parties pour diminuer les conditions de dérivabilité portant sur u , mais en augmentant celles portant sur v .

Si l'ordre de dérivation de u est paire (disons $2k$), alors on pourra faire en sorte par un nombre suffisant de manipulations que u et v aient à être différentiables à l'ordre k ... et on pourra essayer de faire en sorte que u et v appartiennent au même espace fonctionnel...

Remarque. Petit aparté sur d'autres manières de présenter « la chose » : en mécanique, on parle également du « principe du travail virtuel » (ou des puissances virtuelles), de « minimisation de l'énergie potentielle », ou encore de « méthode des résidus pondérés ». Toutes ces méthodes sont « équivalentes » à celle exposée ici, mais souvent leur démonstration se fait uniquement « par les mains », ce qui peut être frustrant.

Principe des puissances virtuelles Le Principe des travaux virtuels est un principe fondamental en mécanique, qui postule un équilibre de puissance dans un mouvement virtuel. Il s'agit d'une formulation duale du principe fondamental de la dynamique. On raisonne comme suit : si un solide est à l'équilibre (statique du solide), la somme des efforts est nulle. Donc si l'on fait faire un déplacement fictif (virtuel) à l'objet, la somme des puissances des forces et moments est nulle. Ce principe constitue la base d'une démarche de modélisation pour les milieux continus (théorie du premier gradient, théorie du second gradient). On parle parfois du principe des travaux virtuels qui est sensiblement identique.

Histoire

L'origine de ce principe revient à Jean Bernoulli, qui énonce en 1725 le principe des vitesses virtuelles, qui consiste à considérer la perturbation de l'équilibre d'un système mécanique par un mouvement infinitésimal respectant les conditions de liaison du système, un mouvement virtuel, et d'en déduire une égalité de puissance.

Ce principe a été par la suite généralisé par D'Alembert et Lagrange en ce qui est connu actuellement sous le nom de principe de D'Alembert (1743).

Le principe des puissances virtuelles est une synthèse de ces principes, ancrée dans un cadre beaucoup plus rigoureux et mathématique (on parle alors de « dualisation » et non plus de « perturbation » de l'équilibre ou du mouvement par un mouvement infinitésimal).



J. Bernoulli D'Alembert Lagrange

Minimisation de l'énergie potentielle *L'énergie potentielle d'un système physique est l'énergie liée à une interaction, qui a le « potentiel » de se transformer en énergie cinétique.*

Cette énergie est une fonction de ce système, dépendant des coordonnées d'espace, et éventuellement du temps, ayant la dimension d'une énergie et qui est associée à une force dite conservative dont l'expression s'en déduit par dérivation (ce qui veut dire au passage que dans les cas où toutes les forces ne sont pas conservatives, par exemple s'il y a du frottement, alors la méthode ne fonctionne pas et doit être « adaptée »). La différence entre les énergies potentielles associées à deux points de l'espace est égale à l'opposé du travail de la force concernée pour aller d'un point à l'autre, et ce quel que soit le chemin utilisé.

Méthode des résidus pondérés *Si l'on considère l'équation $A(u) = f, \forall x \in \Omega$ sur Ω , on appelle résidu des équations d'équilibre la quantité $R(x) = A(u) - f$. Aux bords, on a les conditions $B(u) = h, \forall x \in \Gamma = \partial\Omega$, et on appelle résidu des équations de bords la quantité $\bar{R}(x) = B(u) - h$. En écrivant le problème dans Ω et sur son contour Γ de manière intégrale, on retombe sur une formulation faible.*

La méthode des résidus pondérés consiste à calculer les composantes de la solution approchée par la projection sur des couples de fonctions de projection, appelées fonctions de pondération.

Selon les fonctions de pondération on retrouve les méthodes de collocation par sous-domaine (fonctions constante sur chaque sous-domaine), de collocation par point (fonctions non nulles en un seul point), des fonctions splines (en ajoutant des conditions des raccordements des fonctions et des dérivées), de Galerkin (mêmes fonctions pour les fonctions de forme et de pondération).

Selon la nature du problème (donc selon la nature de l'opérateur A), la formulation faible (7.2) peut être transformée (par exemple par intégration par parties), ce qui permet de faire apparaître une forme symétrique ayant la nature d'un produit scalaire (ou d'un produit hermitien dans le cas de fonctions complexes). Les deux fonctions u et v appartiennent alors à un même espace fonctionnel.

La formulation variationnelle d'un problème régi par des équations aux dérivées partielles correspond à une formulation faible de ces équations qui s'exprime en termes d'algèbre linéaire dans le cadre d'un espace de Hilbert.

Il n'y a donc fondamentalement pas de différence entre formulation faible et formulation variationnelle, sauf que cette dernière appellation est généralement réservée à des cas « qui vont bien »... et de manière très pragmatique aux cas où la formulation faible peut être transformée en un problème de minimisation : on dit alors qu'il existe une fonctionnelle Π dont la variation $\delta\Pi$ correspond au problème faible, d'où le terme de formulation variationnelle (d'un point de vue physique, la fonctionnelle représente l'énergie du système, et le lieu où sa variation est nulle $\delta\Pi = 0$ l'état recherché).

Par ailleurs, d'autres contraintes sur la frontière de Ω peuvent (doivent) être imposées à u (et à v). Ce sont les conditions aux limites. Les conditions aux limites types ont été présentées au paragraphe 6.2.

7.2 Théorème de représentation de Riesz-Fréchet

Le théorème le plus fondamental pour la méthode des éléments finis sera le théorème de Lax-Milgram, toutefois, nous aurons besoin du théorème de représentation de Riesz-Fréchet pour sa démonstration. Ce magnifique théorème fait lui-aussi partie des outils de base d'analyse fonctionnelle.

Pour tout vecteur y d'un espace de Hilbert H , la forme linéaire qui à x associe $\langle y, x \rangle$ est continue sur H (sa norme est égale à celle de y , d'après l'inégalité de Cauchy-Schwarz). Le théorème de Riesz énonce la réciproque : toute forme linéaire continue sur H s'obtient de cette façon.

7.2.1 Cas des formes linéaires

Théorème 38 — Théorème de représentation de Riesz-Fréchet. Soient H un espace de Hilbert (réel ou complexe) muni de son produit scalaire noté $\langle \cdot, \cdot \rangle$ et $f \in H'$ une forme linéaire continue sur H . Alors il existe un unique y dans H tel que pour tout x de H on ait $f(x) = \langle y, x \rangle$. En d'autres termes :

$$\exists ! y \in H, \quad \forall x \in H, \quad f(x) = \langle y, x \rangle \quad (7.5)$$

7.2.2 Extension aux formes bilinéaires

Si a est une forme bilinéaire continue sur un espace de Hilbert réel H (ou une forme sesquilinéaire complexe continue sur un Hilbert complexe), alors il existe une unique application A de H dans H telle que, pour tout $(u, v) \in H \times H$ on ait $a(u, v) = \langle Au, v \rangle$.

De plus, A est linéaire et continue, de norme égale à celle de a .

$$\exists ! A \in \mathcal{L}(H), \quad \forall (u, v) \in H \times H, \quad a(u, v) = \langle Au, v \rangle \quad (7.6)$$

7.3 Théorème de Lax-Milgram

Définition 55 — Continuité d'une forme linéaire. Une forme bilinéaire a est dite continue si elle vérifie :

$$\exists c > 0, \quad \forall (u, v) \in H^2, \quad |a(u, v)| \leq c \|u\| \|v\| \quad (7.7)$$

Définition 56 — Coercitivité d'une forme linéaire. Une forme bilinéaire a est dite coercitive si elle vérifie :

$$\exists \alpha > 0, \quad \forall u \in H, \quad a(u, u) \geq \alpha \|u\|^2 \quad (7.8)$$

Théorème 39 — Théorème de Lax-Milgram. Soit H un espace de Hilbert réel ou complexe muni de son produit scalaire noté $\langle \cdot, \cdot \rangle$, de norme associée $\| \cdot \|$.

Soit $a(\cdot, \cdot)$ une forme bilinéaire (ou une forme sesquilinéaire dans le cas complexe) continue et coercitive sur $H \times H$, et soit f une forme linéaire continue sur H (i.e. $f \in H'$).

Alors il existe un unique u dans H tel que l'équation $a(u, v) = f(v)$ soit vérifiée $\forall v \in H$.

$$\exists ! u \in H, \quad \forall v \in H, \quad a(u, v) = f(v) \quad (7.9)$$

Démonstration. La continuité et la coercitivité de a permettent de dire que a définit un produit scalaire $(\cdot, \cdot)_a$ dont la norme associée $\| \cdot \|_a$ est équivalente à la norme $\| \cdot \|$. Le problème devient : trouver u tel que $\forall v \in H, (u, v)_a = f(v)$.

Or $f(v) = \langle f, v \rangle$ dans la dualité $H H'$ et le problème devient : trouver u tel que $\forall v \in H, (u, v)_a = \langle f, v \rangle$.

Le théorème de représentation de Riesz-Fréchet permet de conclure à l'existence et à l'unicité de la solution. \square

Autre démonstration. Une autre voie consiste à étudier directement le problème de minimisation de la fonctionnelle définie par $J = \frac{1}{2}a(v, v) - f(v)$ (ce qui est une façon de montrer le théorème de Riesz) en considérant une suite minimisante et en montrant qu'elle est de Cauchy par l'identité du parallélogramme. \square

Ce théorème est l'un des fondements de la méthode des éléments finis dits « classiques » ou « en déplacement ». Dans le cas d'éléments finis plus « exotiques » tels que les éléments mixtes, hybrides, mixtes-hybrides et des multiplicateurs de Lagrange, on a besoin d'un peu plus... c'est ce que nous allons voir aux paragraphes suivants.

La forme variationnelle du théorème de Lax-Milgram a déjà été abordée dans la seconde « preuve » proposée du théorème de de Lax-Milgram. Il s'agit, dans les cas où cela est possible, d'aborder le problème sous l'angle de la minimisation d'une fonctionnelle.

Théorème 40 — Forme variationnelle du théorème de Lax-Milgram. Si de plus la forme bilinéaire a est symétrique, alors u est l'unique élément de H qui minimise la fonctionnelle $J : H \rightarrow \mathbb{R}$ définie par $J(v) = \frac{1}{2}a(v, v) - f(v)$ pour tout v de E :

$$\exists ! u \in H, \quad J(u) = \min_{v \in H} J(v) = \min_{v \in H} \left(\frac{1}{2}a(v, v) - f(v) \right) \quad (7.10)$$

Pour les problèmes issus de la mécanique des solides déformables, la quantité $\frac{1}{2}a(v, v)$ représente l'énergie de déformation du solide associé à un « déplacement virtuel » v , et $f(v)$ l'énergie potentielle des forces extérieures appliquées au solide pour le même déplacement virtuel v .

Histoire

Nous venons de voir le théorème de Lax-Milgram, fondement de la méthode des éléments finis, publié dans leur article *Parabolic equations* de 1954 (Contributions to the theory of partial differential equations, Annals of Mathematics Studies, no. 33, Princeton University Press, Princeton, N. J., 1954, pp. 167–190).

Plusieurs améliorations ont permis d'étendre encore la portée de celle-ci. En 1971, dans son article *Error-bounds for finite element method* (Numerische Mathematik 16 : 322–333), Babuška a généralisé le théorème de Lax-Milgram aux problèmes non nécessairement coercitifs (i.e. la forme bilinéaire n'est plus supposée coercitive), élargissant le champ d'application de la méthode des éléments finis.

Puis, en 1974, dans *On the existence, uniqueness and approximation of saddle-point problems arising from lagrangian multipliers*, Brezzi a traité le cas où le problème n'est plus décrit à l'aide d'une seule variable (u jusqu'à présent), mais à l'aide de plusieurs variables, ouvrant ainsi la porte aux formulations mixtes. Ce faisant, il a eu besoin d'imposer des conditions, connues sous le nom de conditions inf-sup, ou conditions BBL pour Babuška-Brezzi-Ladyženskaja (voir par exemple : O. A. Ladyženskaja, On integral estimates, convergence, approximate methods, and solution in functionals for elliptic operators. (Russian) Vestnik Leningrad. Univ. 13 1958 no. 7, 60–69).

Enfin, Brezzi a remplacé sa condition de V-ellipticité par des conditions de Babuška dans son théorème généralisé.



Lax

Milgram

Babuška

Brezzi

Ladyženskaja

7.4 Théorème de Babuška et condition inf-sup

Théorème 41 — Théorème de Babuška. Soient H et M deux espaces de Hilbert. Soient $a(\cdot, \cdot)$ une forme bilinéaire continue sur $H \times M$ et $f(\cdot)$ une forme linéaire continue sur M . On s'intéresse au problème : Existe-t-il $u \in H$ tel que $a(u, v) = f(v)$, $\forall v \in M$?

Si $a(\cdot, \cdot)$ vérifie :

$$\begin{cases} \inf_{\substack{u \in H \\ \|u\|_H=1}} \sup_{\substack{v \in M \\ \|v\|_M=1}} a(u, v) > 0 \\ \sup_{\substack{u \in H \\ \|u\|_H=1}} a(u, v) > 0, \quad \forall v \in M \setminus \{0\} \end{cases} \quad (7.11)$$

Alors $\forall f \in M'$, le problème admet une unique solution u .

La première condition s'écrit aussi :

$$\exists \beta > 0, \quad \sup_{\substack{v \in M \\ \|v\|_M=1}} a(u, v) \geq \beta \|u\|_H, \quad \forall u \in H \quad (7.12)$$

Alors que la seconde s'écrit :

$$\forall v \in M \setminus \{0\}, \quad \exists u \in H, \quad a(u, v) > 0 \quad (7.13)$$

Démonstration. La preuve du théorème de Babuška revient à montrer que l'opérateur A associé à la forme bilinéaire a est un homéomorphisme. \square

7.5 Théorèmes de Brezzi et condition BBL

Théorème 42 — Théorème de Brezzi (1974). Considérons le problème variationnel mixte suivant : Trouver (u, λ) dans $H \times M$ tels que :

$$\begin{cases} a(u, v) + b(v, \lambda) = \langle f, v \rangle & \forall v \in H, \\ b(u, \mu) = \langle g, \mu \rangle & \forall \mu \in M. \end{cases} \quad (7.14)$$

Si les conditions suivantes sont satisfaites :

- $a(\cdot, \cdot)$ est continue ;
- $a(\cdot, \cdot)$ est coercitive sur V ou V -elliptique (uniquement sur V et pas sur H tout entier) (sur $\ker B$) :

$$\exists \alpha > 0 \text{ tel que : } a(v, v) \geq \alpha \|v\|^2, \quad \forall v \in V = \{v \in H; b(v, \mu) = 0, \forall \mu \in M\} \quad (7.15)$$

- $b(\cdot, \cdot)$ est continue ;
- $b(\cdot, \cdot)$ vérifie la condition inf-sup suivante :

$$\inf_{\substack{\mu \in M \\ \|\mu\|_M=1}} \sup_{\substack{v \in H \\ \|v\|_H=1}} b(v, \mu) > 0 \quad (7.16)$$

Alors, $\forall (f, g) \in H' \times M'$, le problème précédent possède une unique solution $(u, \lambda) \in H \times M$.

Rappel : on parle de condition inf-sup ou condition de Brezzi, ou condition de Babuška-Brezzi-Ladyjenskaïa, ou condition BBL.

La condition de V -ellipticité s'écrit également :

$$\inf_{v \in V \setminus \{0\}} a(v, v) > 0 \quad (7.17)$$

et la condition BBL :

$$\exists \beta > 0, \quad \sup_{v \in H} \frac{b(v, \mu)}{\|v\|_H} \geq \beta \|\mu\|_M, \quad \forall \mu \in M \quad (7.18)$$

Nous allons écrire la démarche de la démonstration car elle est typique de la résolution de ce genre de problème.

Démonstration. On va se servir de $H = V \oplus V^\perp$ avec $V = \ker B$ où évidemment B est l'opérateur défini de H dans M' tel que : $\langle Bu, \mu \rangle_M = b(u, \mu)$, $\forall \mu \in M$. Pour le produit scalaire induit, V et V^\perp sont des Hilbert.

La solution cherchée est $u = u_0 + u_\perp$.

Les inconnues seront donc u_0 , u_\perp et λ .

On a les équations pour v_0 , v_\perp et μ . Pour v_\perp l'équation est :

$$\begin{cases} a(u_\perp, v_\perp) + a(u_0, v_\perp) + b(v_\perp, \lambda) &= \langle f, v_\perp \rangle & \forall v_\perp \in V^\perp \\ a(u_\perp, v_0) + a(u_0, v_0) + b(v_0, \lambda) &= \langle f, v_0 \rangle & \forall v_0 \in V \\ b(u_\perp, \mu) + b(u_0, \mu) &= \langle g, \mu \rangle & \forall \mu \in M \end{cases} \quad (7.19)$$

Or dans la seconde équation, $b(v_0, \lambda) = 0$ car $v_0 \in \ker b$ et dans l'équation 3, $b(u_0, \mu) = 0$ car $u_0 \in V$.

On est ramené à 3 sous-problèmes :

- sous-problème 1 : chercher $u_\perp \in V^\perp$ tel que $b(u_\perp, \mu) = g(\mu)$, $\forall \mu \in M$, ce qui est effectivement le cas d'après le théorème de Babuška ;
- sous-problème 2 : chercher $u_0 \in V$ tel que $a(u_0, v_0) = \langle f, v_0 \rangle - a(u_\perp, v_0)$, $\forall v_0 \in V$. Il suffit de vérifier que le second membre définit un opérateur tel que l'on a bien les hypothèses du théorème de Lax-Milgram ;
- sous-problème 3 : chercher $\lambda \in M$ tel que $b(v_\perp, \lambda) = \langle f, v_\perp \rangle - a(u_\perp, v_\perp)$, $\forall v_\perp \in V^\perp$. Il faut un peu plus travailler pour vérifier que l'on est dans le cas d'application du théorème de Babuška, mais ça se fait (une quinzaine de lignes en détail). \square

On peut remplacer les conditions de V-ellipticité sur $a(\cdot, \cdot)$ par des conditions de Babuška.

Théorème 43 — Théorème généralisé de Brezzi. Considérons le problème variationnel mixte suivant : Trouver (u, λ) dans $H \times M$ tels que :

$$\begin{cases} a(u, v) + b(v, \lambda) &= \langle f, v \rangle & \forall v \in H, \\ b(u, \mu) &= \langle g, \mu \rangle & \forall \mu \in M. \end{cases} \quad (7.20)$$

Si les conditions suivantes sont satisfaites :

- $a(\cdot, \cdot)$ est continue ;
- $a(\cdot, \cdot)$ vérifie les deux conditions suivantes :

$$\exists \alpha > 0 \text{ tel que : } \sup_{v \in V \setminus \{0\}} \frac{a(u, v)}{\|v\|} \geq \alpha \|u\|_H^2, \quad \forall u \in V = \ker B \quad (7.21)$$

$\forall v \in V \setminus \{0\}, \exists u \in V, a(u, v) \neq 0$ ou de manière équivalente $\forall v \in V \setminus \{0\}, \sup_u a(u, v) > 0$

- $b(\cdot, \cdot)$ est continue ;
- $b(\cdot, \cdot)$ vérifie la condition inf-sup suivante :

$$\inf_{\substack{\mu \in M \\ \|\mu\|_M=1}} \sup_{\substack{v \in H \\ \|v\|_H=1}} b(v, \mu) > 0 \quad (7.22)$$

Alors, $\forall (f, g) \in H' \times M'$, le problème précédent possède une unique solution $(u, \lambda) \in H \times M$.

Pour vérifier la condition LBB, il est souvent plus simple de vérifier le critère suivant : la condition inf-sup sur $b(\cdot, \cdot)$ est équivalente à :

$$\forall \mu \in M, \exists v \in H \text{ (uniquement dans } V^\perp) \text{ tel que : } b(v, \mu) = \|\mu\|^2 \text{ et } \|v\| \leq \frac{1}{\beta} \|\mu\| \quad (7.23)$$

7.6 Multiplicateurs de Lagrange

Nous avons déjà présenté les multiplicateurs de Lagrange d'un point de vue pratique au chapitre précédent. Regardons maintenant l'aspect mathématique.

Dans les paragraphes 7.3 et 7.4 sur les théorèmes de Lax-Milgram et de Babuška, nous avons traité le cas d'un problème décrit par un seul champ inconnu. Dans le paragraphe 7.5 sur le théorème de Brezzi, nous avons traité le cas d'un problème formulé avec deux champs inconnus (mais on pourrait l'étendre à autant de champs que nécessaire). Dans ce paragraphe, nous allons traiter un cas un peu « au milieu » : par exemple deux domaines possédant chacun leur formulation, mais couplés. Il est donc nécessaire dans ce cas d'introduire une « équation de couplage » entre ces formulations. C'est ce que l'on se propose de réaliser à l'aide des multiplicateurs de Lagrange.

Définition 57 — Formulation comme problème de minimisation sous contrainte. Soient H et M deux espaces de Hilbert. Soient $a(\cdot, \cdot) : H \times H \rightarrow \mathbb{R}$ une forme bilinéaire symétrique et continue et $b(\cdot, \cdot) : H \times M \rightarrow \mathbb{R}$ une forme bilinéaire continue. Soient enfin $f \in H'$ et $g \in M'$ deux formes linéaires. On notera $\langle \cdot, \cdot \rangle_H$ et $\langle \cdot, \cdot \rangle_M$ les produits de dualité entre H et H' et entre M et M' respectivement.

On considère le problème suivant :

$$\begin{aligned} &\text{Trouver } u \in H \text{ tel que } a(u, v) = f(v), \forall v \in H \\ &\text{et vérifiant les contraintes supplémentaires : } b(v, \mu) = g(\mu), \forall \mu \in M. \end{aligned} \quad (7.24)$$

Ce problème est équivalent à :

$$\begin{aligned} &\text{Trouver } u \in H \text{ tel que } J(u) = \min\{J(v), v \in H \text{ et } b(v, \mu) = \langle g, \mu \rangle_M, \mu \in M\} \\ &\text{avec } J(v) = \frac{1}{2}a(v, v) - \langle f, v \rangle_H \end{aligned} \quad (7.25)$$

On parle de problème de minimisation sous contrainte.

On introduit un Lagrangien : $\mathcal{L}(v, \mu) = J(v) + b(v, \mu) - \langle g, \mu \rangle_M$. Évidemment, $\forall \mu, \mathcal{L}(v, \mu) = J(v)$ si v vérifie les contraintes. L'idée est de considérer $\min_{v \in H} \mathcal{L}(v, \cdot)$. La question devient alors : existe-t-il un multiplicateur particulier $\lambda \in M$ tel que $\min_{v \in H} \mathcal{L}(v, \lambda)$ soit égale à $u \in H$ solution de (7.25) ? On appelle parfois λ pénalité ou fonction de pénalisation.

En notant que :

$$\left\langle \frac{\partial \mathcal{L}}{\partial v}(u, \lambda), v \right\rangle_H = \frac{d}{dt} \mathcal{L}(u + tv, \lambda)|_{t=0} = a(u, v) - \langle f, v \rangle_H + b(v, \lambda) \quad (7.26)$$

Alors $(u, \lambda) \in H \times M$ est solution du problème mixte (7.14) :

$$\begin{cases} a(u, v) + b(v, \lambda) = \langle f, v \rangle & \forall v \in H, \\ b(u, \mu) = \langle g, \mu \rangle & \forall \mu \in M. \end{cases} \quad (7.27)$$

et on se retrouve dans le cadre du théorème de Brezzi au paragraphe 7.5.

De plus, si $a(v, v) \geq 0, \forall v \in H$, alors (u, λ) est solution de (7.14) correspond à (u, λ) est un point-selle de \mathcal{L} , i.e. :

$$\forall \mu \in M, \mathcal{L}(u, \mu) \leq \mathcal{L}(u, \lambda) \leq \mathcal{L}(v, \lambda), \forall v \in H \quad (7.28)$$

Remarquons enfin que si l'on note :

$$A((u, \lambda), (v, \mu)) = a(u, v) + b(v, \lambda) + b(u, \mu) \quad (7.29)$$

$$L((v, \mu)) = f(v) + g(\mu) \quad (7.30)$$

et que l'on pose les nouvelles variables $\mathcal{U} = (u, \lambda)$ et $\mathcal{V} = (v, \mu)$, alors on revient à un problème de type Lax-Milgram : trouver $\mathcal{U} \in H \times M$ tel que pour tout $\mathcal{V} \in H \times M, A(\mathcal{V}, \mathcal{U}) = L(\mathcal{V})$.

Nous reviendrons sur cette approche au chapitre 16.

Comme nous l'avons déjà mentionné, Lagrange est le fondateur du calcul des variations avec Euler.

Il est également le fondateur de la théorie des formes quadratiques, et démontre le théorème de Wilson sur les nombres premiers et la conjecture de Bachet sur la décomposition d'un entier en quatre carrés (connu aussi sous le nom de théorème des quatre carrés de Lagrange). On lui doit un cas particulier du théorème auquel on donnera son nom en théorie des groupes, un autre sur les fractions continues, l'équation différentielle de Lagrange...



Lagrange

En physique, en précisant le principe de moindre action, avec le calcul des variations, vers 1756, il invente la fonction de Lagrange, qui vérifie les équations de Lagrange, puis développe la mécanique analytique, vers 1788, pour laquelle il introduit les multiplicateurs de Lagrange. Il entreprend aussi des recherches importantes sur le problème des trois corps en astronomie, un de ses résultats étant la mise en évidence des points de libration (dits points de Lagrange) en 1772.

En mécanique des fluides, il introduisit le concept de potentiel de vitesse en 1781, bien en avance sur son temps. Il démontra que le potentiel de vitesse existe pour tout écoulement de fluide réel, pour lequel la résultante des forces dérive d'un potentiel. Dans le même mémoire de 1781, il introduisit, en plus, deux notions fondamentales : le concept de la fonction de courant, pour un fluide incompressible, et le calcul de la célérité d'une petite onde dans un canal peu profond. Rétrospectivement, cet ouvrage marqua une étape décisive dans le développement de la mécanique des fluides moderne.

Chapitre 8

Problèmes physiques : formulations faibles et variationnelles

Résumé — Nous allons maintenant reprendre les problèmes types exposés sous forme forte dans le chapitre 6, pour leur appliquer les méthodes du chapitre précédent. Nous obtiendrons alors les formulations faibles et variationnelles des mêmes problèmes.

Rappel méthodologique : La formulation variationnelle d'un problème s'obtient en multipliant l'équation d'état du problème par une fonction test, en intégrant par parties, et en utilisant les conditions aux limites. (De plus, en mécanique, la formulation variationnelle est le principe des puissances virtuelles).

8.1 Phénomènes de propagation et de diffusion

8.1.1 Équations de Laplace et Poisson

Sur ce premier exemple, nous détaillerons plus la démarche et les étapes que dans la suite des cas présentés.

Laplacien de Dirichlet

On considère l'équation de Poisson (Laplace si $f = 0$) avec les conditions aux limites de Dirichlet, i.e. le problème suivant :

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ u = 0 & \text{sur } \Gamma = \partial\Omega \end{cases} \quad (8.1)$$

Multiplier par une fonction test v et intégrer par parties (ou utiliser la formule de Green) conduit à :

$$\int_{\Omega} \nabla u \cdot \nabla v - \int_{\Gamma} \frac{\partial u}{\partial n} v = \int_{\Omega} f v \quad (8.2)$$

Comme, rappelons le, le but est de faire en sorte que u et v jouent un rôle symétrique, on veut donc qu'ils aient même régularité et mêmes conditions aux limites. La régularité nécessaire est u et v sont $H^1(\Omega)$, et la prise en compte des conditions aux limites conduit finalement à chercher u et v dans $H_0^1(\Omega)$.

Le problème est donc :

$$\text{Trouver } u \in H_0^1(\Omega) \text{ tel que } \forall v \in H_0^1(\Omega), \quad \int_{\Omega} \nabla u \cdot \nabla v = \int_{\Omega} f v \quad (8.3)$$

dont le second membre n'a de sens que si $f \in L^2(\Omega)$, ce qui sera supposé. Notons que le problème est également bien défini dans le cas où $f \in H^{-1}(\Omega)$, car $v \in H_0^1(\Omega)$.

Avec cette formulation, le théorème de Lax-Milgram permet de conclure à l'existence et à l'unicité de la solution.

Équivalence des solutions Nous avons construit un problème variationnel dont on sait qu'il a une unique solution. Une question naturelle est alors de savoir en quel sens on a résolu le problème initial.

Il est facile de voir que l'équation $-\Delta u = f$ est vérifiée au sens des distributions. Or $f \in L^2(\Omega)$, donc $-\Delta u \in L^2(\Omega)$, et l'égalité a donc lieu presque partout. De la même manière, on a $u = 0$ presque partout sur Γ (pour la mesure surfacique sur Γ).

On comprend pourquoi, parmi différentes formulations faibles, on en retient une plutôt qu'une autre : afin d'obtenir une formulation variationnelle à partir de laquelle on raisonne rigoureusement, i.e. existence et unicité du résultat, puis retour au problème initial.

Une autre formulation variationnelle. Afin d'illustrer la remarque précédente, construisons une autre formulation variationnelle du même problème. Pour cela, intégrons une nouvelle fois par parties, afin de chercher une solution encore plus faible. Le problème devient :

$$\text{Trouver } u \in L^1(\Omega) \text{ tel que } \forall v \in C_c^\infty(\Omega), \quad - \int_{\Omega} u \Delta v = \int_{\Omega} f v \quad (8.4)$$

Tout d'abord, on peut remarquer que l'on ne peut plus appliquer le théorème de Lax-Milgram. Ensuite, rappelons une remarque déjà faite : plus on cherche des solutions en un sens faible, plus il est facile de prouver l'existence de solutions, mais plus il est difficile d'obtenir l'unicité de la solution.

En travaillant dans $L^1(\Omega)$, on ne pourrait pas, par exemple, donner de sens à $u = 0$ sur Γ .

Comme nous l'avons vu au chapitre précédent, l'intérêt des problèmes symétriques, c'est qu'ils peuvent s'interpréter en terme de minimisation.

La solution du problème initial est également la fonction u qui réalise le minimum de :

$$\min_{u \in H_0^1(\Omega)} \frac{1}{2} \int_{\Omega} |\nabla u|^2 - \int_{\Omega} f u \quad (8.5)$$

Laplacien de Dirichlet relevé

On considère l'équation de Laplace-Poisson avec les conditions aux limites suivantes :

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ u = g & \text{sur } \Gamma = \partial\Omega \end{cases} \quad (8.6)$$

où $g \in H^{1/2}(\Gamma)$ est une fonction donnée.

L'idée est évidemment de se ramener au cas précédent en réalisant un relèvement d'espace.

Pour cela, on considère la fonction $v = u - \bar{u}$ où $\bar{u} \in H^1(\Omega)$ est un relèvement de la condition au bord, i.e. tel que $\bar{u} = g$ sur Γ .

On utilise ensuite la surjectivité de l'application trace, γ_0 (ce qui devrait malgré tout rester compréhensible bien que la trace ait été définie de manière « un peu légère » au paragraphe 5.5.1).

La formulation variationnelle du problème est :

$$\text{Trouver } u \in \{v \in H^1(\Omega), v = g \text{ sur } \Gamma\} \text{ tel que } \forall v \in H_0^1(\Omega), \quad \int_{\Omega} \nabla u \cdot \nabla v = \int_{\Omega} f v \quad (8.7)$$

Laplacien de Neumann

On considère l'équation de Laplace-Poisson avec les conditions aux limites de Neumann, i.e. le problème suivant :

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ \frac{\partial u}{\partial n} = g & \text{sur } \Gamma = \partial\Omega \end{cases} \quad (8.8)$$

On obtient immédiatement la formulation variationnelle suivante :

$$\text{Trouver } u \in H^1(\Omega) \text{ tel que } \forall v \in H^1(\Omega), \quad \int_{\Omega} \nabla u \cdot \nabla v = \int_{\Omega} f v + \int_{\Gamma} g v \quad (8.9)$$

qui nécessite $f \in L^2(\Omega)$ et $g \in L^2(\Gamma)$.

On peut remarquer qu'ici une simple intégration par parties a suffi à intégrer les conditions aux limites dans la formulation variationnelle, i.e. les conditions aux limites n'ont pas eu à être introduites dans l'espace fonctionnel pour u , contrairement au cas du Laplacien de Dirichlet (en même temps, demander à une fonction de $H^1(\Omega)$ d'avoir une dérivée normale égale à une fonction g au bord n'a pas de sens).

Lorsque les conditions aux limites s'introduisent d'elles-mêmes comme dans le cas du Laplacien de Neumann, on dit que l'on tient compte des conditions aux limites de manière naturelle. Lorsqu'il faut les inclure, on dit qu'on en tient compte de manière essentielle.

Il faut remarquer que le problème variationnel ne peut avoir de solution que si f et g vérifient une condition de compatibilité, i.e. si :

$$\int_{\Omega} f + \int_{\Gamma} g = 0 \quad (8.10)$$

Il faut aussi remarquer que si u est une solution du problème, alors $u + c$, où c est une constante, est également solution. On ne peut donc pas espérer prouver qu'il existe une unique solution au problème sans restreindre l'espace fonctionnel pour u . Pour éliminer l'indétermination due à cette constante additive, on introduit, par exemple, $V = \{u \in H^1(\Omega), \int_{\Omega} u = 0\}$ et on pose le problème sous la forme :

$$\text{Trouver } u \in V \text{ tel que } \forall v \in V, \quad \int_{\Omega} \nabla u \cdot \nabla v = \int_{\Omega} f v + \int_{\Gamma} g v \quad (8.11)$$

En utilisant l'inégalité de Poincaré-Wirtinger, on montre que le problème est bien coercitif et donc qu'il admet une unique solution.

Remarquons que si la condition de compatibilité n'est pas vérifiée, alors la solution de la dernière formulation résout bien le problème initial avec un f ou un g modifié d'une constante additive, de manière à vérifier la condition de compatibilité.

Il y aurait lieu à quelques discussions sur l'interprétation des résultats, mais celle-ci serait beaucoup plus abstraite que dans le cas du Laplacien de Dirichlet. Le lecteur désireux de rentrer (à raison) dans ce genre de détails trouvera aisément un cours de M2 sur le sujet. Cela dépasse le cadre que nous nous sommes fixé dans ce document.

Laplacien avec conditions mixtes

On considère l'équation de Laplace-Poisson avec les conditions aux limites mixtes suivantes :

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ \frac{\partial u}{\partial n} + \alpha u = 0 & \text{sur } \Gamma = \partial\Omega \end{cases} \quad (8.12)$$

On obtient la formulation variationnelle :

$$\text{Trouver } u \in H^1(\Omega), \text{ tel que } \forall v \in H^1(\Omega), \quad \int_{\Omega} \nabla u \cdot \nabla v + \int_{\Gamma} \alpha u v = \int_{\Omega} f v + \int_{\Gamma} g v \quad (8.13)$$

Laplacien avec conditions mixtes en formulation mixte

On considère le problème précédent général comportant des conditions de Dirichlet et de Neumann sur les parties Γ_D et Γ_N respectivement. Ce problème est exprimé sous forme mixte, i.e. à l'aide de deux variables u et p , par :

$$\begin{cases} \nabla p + u = 0 & \text{dans } \Omega \\ \nabla \cdot u = u & \text{dans } \Omega \\ p = g_D & \text{sur } \Gamma_D \\ \frac{\partial u}{\partial n} = g_N \cdot n & \text{sur } \Gamma_N \end{cases} \quad (8.14)$$

On obtient la formulation variationnelle : Trouver u et p tels que :

$$\begin{cases} \forall v \in \mathbb{X}, \quad \int_{\Omega} p \nabla \cdot v = \int_{\Gamma_D} g_D \cdot n \\ \forall q \in L^2(\Omega), \quad \int_{\Omega} q \nabla \cdot u = \int_{\Omega} q f \\ \frac{\partial u}{\partial n} = g_N \cdot n \text{ sur } \Gamma_n \end{cases} \quad (8.15)$$

avec $\mathbb{X} = \{v \in H(\text{div}, \Omega); v \cdot n = 0 \text{ sur } \Gamma_N\}$.

8.1.2 Équation d'onde

Rappelons que l'équation des ondes correspond à un problème d'évolution hyperbolique linéaire. Différentes conditions aux limites peuvent être imposées. Dans le cadre de ce document, nous nous intéresserons plus particulièrement à l'acoustique. Ce paragraphe est donc traité un peu plus loin dans le document.

8.1.3 Équation de la chaleur

Après avoir présenté le problème elliptique du Laplacien, intéressons-nous maintenant au cas parabolique de l'équation de la chaleur. Pour cela, considérons le problème suivant : Trouver $u(x, t)$ tel que :

$$\begin{cases} \partial_t u - \Delta u = f & \text{dans } [0; T] \times \Omega \\ u(t, \cdot) = 0 & \text{sur } [0; T] \times \Gamma \\ u(0, \cdot) = u_0 & \text{sur } \Omega \end{cases} \quad (8.16)$$

On opère comme précédemment. Toutefois, nous faisons jouer à la variable de temps t et la variable d'espace x des rôles différents : on considère la fonction $u(t, x)$ comme une fonction du temps, à valeur dans un espace fonctionnel en x , et on prend comme fonctions tests des fonctions qui dépendent seulement de la variable x . On obtient alors facilement la formulation variationnelle suivante :

$$\text{Trouver } u \text{ tel que } \forall v \in H_0^1(\Omega), \quad \frac{d}{dt} \int_{\Omega} uv + \int_{\Omega} \nabla u \cdot \nabla v = \int_{\Omega} f v \quad (8.17)$$

Pour donner un sens à cette formulation variationnelle, une certaine régularité de u est requise :

$$u \in V = C^0(0, T; L^2(\Omega)) \cap L^2(0, T; H_0^1(\Omega)) \quad (8.18)$$

L'espace $C^0(0, T; L^2(\Omega))$ est l'espace des fonctions $v(t, x)$ telles que pour tout t , $v(t, \cdot)$ est une fonction de $L^2(\Omega)$; l'espace $L^2(0, T; H_0^1(\Omega))$ est l'espace des fonctions $v(t, x)$ telles que pour

presque tout t , $v(t, \cdot)$ est une fonction de $H_0^1(\Omega)$ et $\int_0^T \|v\|_{H_0^1(\Omega)}^2 < \infty$. Le premier est un espace de Banach, le second, un espace de Hilbert.

La formulation variationnelle est donc :

$$\text{Trouver } u \in V \text{ tel que } u(0, \cdot) = u_0 \text{ et tel que } \forall v \in H_0^1(\Omega), \quad \frac{d}{dt} \int_{\Omega} uv + \int_{\Omega} \nabla u \cdot \nabla v = \int_{\Omega} f v \quad (8.19)$$

Pour résoudre le problème, il faut utiliser une méthode numérique. On montre l'existence puis l'unicité par les estimateurs à priori. Nous détaillerons cela plus loin.

8.2 Mécanique des fluides

Pour la présentation des formulations faibles, nous commencerons par le cas de l'équation de Stokes avant de passer à celui de l'équation de Navier-Stokes (afin de nous permettre d'introduire les notations progressivement).

8.2.1 Équation de Stokes

Si l'on considère le problème de Stokes pour un fluide incompressible avec condition de Dirichlet, il nous faut résoudre le système :

$$\begin{cases} -\eta \Delta u + \nabla p = \rho f & \text{dans } \Omega \\ \operatorname{div} u = 0 & \text{dans } \Omega \\ u = 0 & \text{sur } \Gamma = \partial\Omega \end{cases} \quad (8.20)$$

La formulation mixte qui en découle est :

$$\eta \int_{\Omega} \nabla u : \nabla v - \int_{\Omega} p \operatorname{div} v = \int_{\Omega} f \cdot v \quad (8.21)$$

On introduit les formes bilinéaires :

$$\begin{aligned} a : H_0^1(\Omega)^n \times H_0^1(\Omega)^n &\rightarrow \mathbb{R} \\ (u, v) &\mapsto a(u, v) = \eta \int_{\Omega} \nabla u : \nabla v \end{aligned} \quad (8.22)$$

et

$$\begin{aligned} b : H_0^1(\Omega)^n \times L^2(\Omega) &\rightarrow \mathbb{R} \\ (u, q) &\mapsto b(u, q) = - \int_{\Omega} (\operatorname{div} u) q \end{aligned} \quad (8.23)$$

On introduit également l'espace $L_0^2(\Omega)$ des fonction $L^2(\Omega)$ à moyenne nulle :

$$L_0^2(\Omega) = \left\{ q \in L^2(\Omega), \quad \int_{\Omega} q = 0 \right\} \quad (8.24)$$

La formulation variationnelle mixte (vitesse/pression) est alors : Trouver $(u, p) \in H_0^1(\Omega) \times L_0^2(\Omega)$ tels que :

$$\begin{cases} a(u, v) + b(v, p) = (f, v) & \forall v \in H_0^1(\Omega)^n \\ b(u, q) = 0 & \forall q \in L_0^2(\Omega) \end{cases} \quad (8.25)$$

Notons que dans la seconde équation, la fonction test q pourrait de manière équivalente être cherchée dans $L^2(\Omega)$ (pas besoin de la moyenne nulle).

Par contre, il est important de chercher $p \in L_0^2(\Omega)$. C'est cette condition de pression nulle qui assure l'unicité de la pression (comme dans le cas précédent du Laplacien de Neumann).

Il est possible de formuler ce problème de manière non mixte, i.e. sous une forme compatible avec le théorème de Lax-Milgram. Considérons $V = \{u \in H_0^1(\Omega)^n / \operatorname{div} u = 0 \text{ dans } \Omega\}$. Alors u (la vitesse) est solution du problème :

$$\text{Trouver } u \in V \text{ tel que } a(u, v) = (f, v), \forall v \in V.$$

D'ailleurs c'est cette formulation qui permet (a est continue et coercitive sur V) de conclure à l'existence et à l'unicité de u . De là, il ne reste plus qu'à démontrer la même chose pour p dans la formulation mixte.

8.2.2 Équation de Navier-Stokes

Considérons le problème de Navier-Stokes, pour un fluide incompressible, et soumis à des conditions initiales en temps et sur le contour $\Gamma = \partial\Omega$:

$$\begin{cases} \frac{\partial u}{\partial t} + (u \cdot \nabla)u + \eta \Delta u + \nabla p = \rho f & \text{dans } \Omega \times \mathbb{R}^+ \\ \operatorname{div} u = 0 & \text{dans } \Omega \times \mathbb{R}^+ \\ u = g & \text{sur } \Gamma \times \mathbb{R}^+ \\ u(\cdot, 0) = u_0 & \text{dans } \Omega \end{cases} \quad (8.26)$$

La fonction $g = g(x, t)$ dans la condition de type Dirichlet sur le bord du domaine doit être à flux nul sur Γ , i.e. on suppose que :

$$\int_{\Gamma} g \cdot n = 0 \quad (8.27)$$

avec n la normale extérieure à Γ . Cette condition est nécessaire pour que le problème de Navier-Stokes admette une solution.

Introduisons également les espaces de fonctions à valeurs dans un espace de Banach B . Pour $q \geq 1$, on a :

$$L^q(0, T; B) = \left\{ v : [0; T] \rightarrow B; v \text{ est mesurable et } \int_0^T \|v(t)\|_B^q dt < +\infty \right\} \quad (8.28)$$

Ces espaces sont munis de la norme :

$$\|u\|_{L^q(0, T; B)} = \left(\int_0^T \|u(t)\|_B^q dt \right)^{1/q} \quad (8.29)$$

Notons que ces espaces ont été « présentés » au paragraphe précédent sur l'équation de la chaleur. La notation était un peu différente mais il s'agit bien des mêmes espaces. Nous avons voulu profiter de ces exemples pour introduire ces deux notations, différentes mais désignant bien la même chose.

Posons $X = H_0^1(\Omega)^n$ et $Y = L_0^2(\Omega)^n$. La formulation variationnelle mixte des équations de Navier-Stokes s'écrit : Soit $f \in L^2(0, T; L^2(\Omega)^n)$ et $u_0 \in L^2(\Omega)^n$. Trouver :

$$u \in W(0, T) = \left\{ v \in L^2(0, T; X), \frac{dv}{dt} \in L^2(0, T; X') \right\} \quad (8.30)$$

et $p \in L^2(0, T; Y)$, tels que p.p. $t \in (0, T)$, on ait :

$$\begin{cases} \left\langle \frac{du}{dt}(t), v \right\rangle_{X', X} + a(u(t), v) + c(u(t), u(t), v) + b(v, p(t)) = (f(t), v) & \forall v \in X \\ b(u(t), q) = 0 & \forall q \in Y \\ u(0) = u_0 \end{cases} \quad (8.31)$$

où $c : X \times X \times X \rightarrow \mathbb{R}$ est la forme trilinéaire définie par :

$$c(w, z, v) = \int_{\Omega} [(w \cdot \nabla)z] \cdot v = \sum_{i,j=1}^n \int_{\Omega} w_j \frac{\partial z_i}{\partial x_j} v_i \quad (8.32)$$

Cette forme trilinéaire a certaines propriétés, par exemple lorsque l'on permute des termes où lorsque l'on considère que certaines variables sont à divergence nulle...

Pour des raisons de stabilité, la forme c est remplacée par le forme \tilde{c} définie par :

$$\tilde{c} = \frac{1}{2} (c(w, z, v) - c(w, v, z)) \quad (8.33)$$

qui est antisymétrique et possède par conséquent la propriété $\tilde{c}(w, z, z) = 0, \forall w, x, z \in X$.

Nous ne rentrons pas plus avant dans la méthode, car il faudrait alors parler dès à présent du schéma numérique associé, ce qui sera fait plus tard.

8.2.3 Équation d'Euler

Nous avons déjà vu que sans le cas d'un nombre de Reynolds élevé, alors le terme de convection non-linéaire $(u \cdot \nabla)u$ devient prépondérant, et on obtient alors l'équation d'Euler. Pour un fluide incompressible, le problème à résoudre est donc :

$$\begin{cases} \frac{\partial u}{\partial t} + (u \cdot \nabla)u + \nabla p = \rho f & \text{dans } \Omega \times \mathbb{R}^+ \\ \operatorname{div} u = 0 & \text{dans } \Omega \times \mathbb{R}^+ \end{cases} \quad (8.34)$$

On considérera qu'il s'agit d'un cas simplifié de Navier-Stokes, et on le traitera comme tel.

8.3 Équations de la mécanique des milieux continus des solides

8.3.1 Formulation générale

La formulation la plus générale du problème est :

$$\begin{cases} \operatorname{div} \sigma + f = \rho \ddot{u} + \mu \dot{u} & \text{dans } \Omega \quad \text{équation de la dynamique} \\ \sigma = H(\varepsilon - \varepsilon_{th}) & \text{dans } \Omega \quad \text{loi de comportement} \end{cases} \quad (8.35)$$

à laquelle il est nécessaire d'ajouter les conditions aux limites suivantes :

$$\begin{cases} u = \bar{u} & \text{sur } \Gamma_D \quad \text{déplacement imposé} \\ \sigma \cdot n(x) = g_N & \text{sur } \Gamma_N \quad \text{condition sur le bord} \\ \dot{u}(t=0) = \dot{u}_0 & \text{condition initiale en vitesse} \\ u(t=0) = u_0 & \text{condition initiale en déplacement} \end{cases} \quad (8.36)$$

avec comme d'habitude, Γ_D et Γ_N une partition de la frontière $\Gamma = \partial\Omega$ du domaine Ω .

Pour obtenir une formulation faible, on applique les mêmes recettes que ci-dessus. Dans un premier temps, on ne traite que l'équation de la dynamique, la loi de comportement n'étant là

finalement que pour « simplifier » l'équation en permettant de lier des variables entre elles. Ainsi, la formulation faible générale est :

$$\int_{\Omega} \rho v \ddot{u} + \int_{\Omega} \mu v \dot{u} + \int_{\Omega} \varepsilon : \sigma - \int_{\Omega} f v - \int_{\Gamma_N} g_N v = 0, \quad \forall v \text{ tel que } v = \bar{u} \text{ sur } \Gamma_D \quad (8.37)$$

Nous sommes restés très prudents sur l'appartenance des différentes grandeurs à des espaces, car cela dépend du choix des variables.

8.3.2 Choix des variables

Dans la formulation variationnelle présentée, on ne manquera pas de remarquer que les variables sont : les déplacements (et leurs dérivées temporelles : vitesse et accélération), les déformations et les contraintes.

Il est possible de choisir de conserver toutes ces variables, i.e. d'avoir un problème dépendant de trois champs inconnus : u , ε et σ . Une telle formulation est qualifiée de mixte. On qualifie d'ailleurs de mixte, toute formulation ayant plus d'un seul champ inconnu.

Comme on connaît des relations entre les différents champs (relation déplacement / déformation et relation déformation / contrainte), il est possible d'obtenir des formulations ayant deux ou même un seul champ inconnu, comme nous l'avons déjà vu au paragraphe 7.5 (théorèmes de Brezzi).

Formulation classique en déplacements

L'intérêt de réduire le nombre de champ inconnu est évident. Faut-il donc opter définitivement pour une formulation à un seul champ ? Ce n'est pas évident. En effet, utiliser une formulation à k champs parmi n se traduira par n'obtenir que ces k champs comme solution directe du système. Les $n - k$ champs manquant doivent alors être obtenus à partir des k champs choisis... et cela peut poser certains problèmes, dont nous parlerons plus tard.

Il est à noter que la formulation dite en déplacements, également qualifiée de classique, est la plus utilisée. Comme son nom l'indique, elle ne comporte que le champ de déplacement comme inconnu. Il suffit de considérer, dans la formulation précédente, que les contraintes sont obtenues à partir des déformations, via la loi de comportement, et que les déformations sont obtenues à partir des déplacements par la relation idoïne (en petits ou grands déplacements).

Dans la formulation en déplacement, on suppose que les relations liant les déformations aux déplacements (opérateur de dérivation en petites ou grandes déformations) et les contraintes aux déplacements (en fait les contraintes aux déformations via la loi de comportement, puis les déformations aux déplacements comme ci-avant) sont connues et vérifiées :

$$\int_{\Omega} \rho v \ddot{u} + \int_{\Omega} \mu v \dot{u} + \int_{\Omega} \varepsilon(v) : \sigma(u) - \int_{\Omega} f v - \int_{\Gamma_N} g_N v = 0, \quad \forall v \text{ tel que } v = \bar{u} \text{ sur } \Gamma_D \quad (8.38)$$

De manière plus explicite, si on note $\varepsilon = \mathcal{L}u$ la relation entre déformations et déplacements et $\sigma = H\varepsilon = H\mathcal{L}u$ la relation entre contraintes et déplacement, alors il vient :

$$\int_{\Omega} \rho v \ddot{u} + \int_{\Omega} \mu v \dot{u} + \int_{\Omega} \mathcal{L}v : H\mathcal{L}u - \int_{\Omega} f v - \int_{\Gamma_N} g_N v = 0, \quad \forall v \text{ tel que } v = \bar{u} \text{ sur } \Gamma_D \quad (8.39)$$

où n'apparaissent effectivement plus que les déplacements comme inconnues.

On se contente toutefois généralement d'écrire $\varepsilon(u)$ et $\sigma(u)$ pour dire que l'inconnue est bien u , ce qui évite d'alourdir l'écriture.

Cette formulation « en déplacement » correspond au problème de minimisation de la fonctionnelle de l'énergie potentielle totale que l'on appelle également fonctionnelle primale :

$$J(u) = \frac{1}{2} \int_{\Omega} \varepsilon(v) : \sigma(u) - \int_{\Omega} f v - \int_{\Gamma_N} g_N v \quad (8.40)$$

où le champ de déplacements doit vérifier $v = \bar{u}$ sur Γ_D . Les relations entre déformations et déplacements et entre contraintes et déplacements sont supposées vérifiées de manière implicite dans Ω .

Illustration : formulation en déplacements de la statique des matériaux isotropes

En se plaçant dans le cas statique, et en écrivant « en déplacements », i.e. en supprimant le champ de contrainte des inconnues (et en utilisant, au passage, la symétrie de σ), la formulation variationnelle précédente devient :

$$\int_{\Omega} H_{ijkl} \varepsilon_{ij}(u) \varepsilon_{kl}(v) = \int_{\Omega} f v + \int_{\Gamma_N} g_N v, \quad \forall v \in H_D^1(\Omega) \quad (8.41)$$

Notons que pour prouver la coercitivité de la forme bilinéaire, il faut recourir à l'inégalité de Korn.

Si en plus le matériau est isotrope, cette expression se simplifie (λ et μ sont les coefficients de Lamé) en :

$$\int_{\Omega} \lambda \operatorname{div}(u) \operatorname{div}(v) + 2\mu \varepsilon(u) \cdot \varepsilon(v) = \int_{\Omega} f v + \int_{\Gamma_N} g_N v, \quad \forall v \in H_D^1(\Omega) \quad (8.42)$$

que l'on peut voir comme un problème de minimisation de la fonctionnelle :

$$J(u) = \int_{\Omega} \lambda (\operatorname{div} u)^2 + 2\mu \sum_{ij} \varepsilon_{ij}(u)^2 - \int_{\Omega} f v - \int_{\Gamma_N} g_N v \quad (8.43)$$

D'une manière générale, toute formulation faible, quelque soit le nombre de champs inconnus, s'écrit en mécanique sous la forme de minimisation d'une fonctionnelle, dont on sait trouver la signification physique.

Formulations mixte et hybride

La formulation la plus générale est celle comportant les trois champs de déplacements u , de déformations ε , et de contraintes σ comme inconnues. Cette formulation mixte est connue sous le nom de principe de Hu-Washizu (1975). Sous forme variationnelle, elle s'écrit :

$$J(u, \varepsilon, \sigma) = \frac{1}{2} \int_{\Omega} \varepsilon : \sigma + \int_{\Omega} (\nabla u) \sigma - \int_{\Omega} f v - \int_{\Gamma_N} g_N v - \int_{\Gamma_D} (u - \bar{u})(\sigma(v) \cdot n) \quad (8.44)$$

et ne comporte aucune condition (puisque elles sont toutes formellement présente dans la formulation).

Il est possible d'obtenir une formulation n'ayant que le champ de contraintes comme inconnue. Il s'agit de la fonctionnelle duale :

$$J(\sigma) = -\frac{1}{2} \int_{\Omega} \sigma S \sigma - \int_{\Omega} \sigma(v) S \sigma_0 + \int_{\Gamma_D} \bar{u} \sigma(v) \cdot n \quad (8.45)$$

où le champ de contrainte doit vérifier les conditions $\operatorname{div} \sigma + f = 0$ dans Ω et $\sigma n = g_N$ sur Γ_N . S désigne la loi de comportement donnée sous forme inverse (i.e. sous forme de souplesse, i.e. $S = H^{-1}$).

L'utilisation de multiplicateurs de Lagrange permet de construire des fonctionnelles multi-champs. La fonctionnelle de l'énergie complémentaire est la suivante :

$$J(\sigma, \lambda, \mu) = -\frac{1}{2} \int_{\Omega} \sigma S \sigma - \int_{\Omega} \lambda (\operatorname{div} \sigma + f) + \int_{\Gamma_D} \bar{u} \sigma \cdot n + \int_{\Gamma_N} \mu (\sigma \cdot n - g_N) \quad (8.46)$$

où λ et μ sont les multiplicateurs de Lagrange. Les multiplicateurs de Lagrange présentés ci-dessus ont une signification physique (que l'on trouve en écrivant la stationnarité de la fonctionnelle). En remplaçant les multiplicateurs par les quantités qu'ils représentent, i.e. en remplaçant λ par u dans Ω et μ par u sur Γ_N , on obtient la fonctionnelle d'Hellinger-Reissner sous sa première forme :

$$J(u, \sigma) = -\frac{1}{2} \int_{\Omega} \sigma S \sigma - \int_{\Omega} (\operatorname{div} \sigma + f)u + \int_{\Gamma_D} \bar{u} \sigma \cdot n - \int_{\Gamma_N} (\sigma \cdot n - g_N)v \quad (8.47)$$

En notant :

$$a(\sigma, \sigma) = - \int_{\Omega} \sigma S \sigma \quad (8.48)$$

$$b(\sigma, u) = - \int_{\Omega} \operatorname{div} \sigma u + \int_{\Gamma_D} \bar{u} \sigma \cdot n + \int_{\Gamma_N} u \sigma \cdot n \quad (8.49)$$

$$L(u) = \int_{\Omega} f u + \int_{\Gamma_N} g_N v \quad (8.50)$$

alors, la première fonctionnelle de Reissner se met sous la forme :

$$J(\sigma, u) = \frac{1}{2} a(\sigma, \sigma) + b(\sigma, u) - L(u) \quad (8.51)$$

En effectuant une intégration par parties du terme mixte $b(\sigma, u)$, on obtient la fonctionnelle d'Hellinger-Reissner sous sa deuxième forme :

$$J(u, \sigma) = -\frac{1}{2} \int_{\Omega} \sigma S \sigma + \int_{\Omega} \sigma : \varepsilon - \int_{\Omega} f v - \int_{\Gamma_N} g_N v \quad (8.52)$$

où le champ de déplacement doit vérifier $v = \bar{u}$ sur Γ_D , et où la relation entre déformations et déplacement est supposée vérifiée dans Ω . Cette fonctionnelle d'Hellinger-Reissner est également appelée fonctionnelle mixte. L'adjectif hybride peut être adjoint à l'une quelconque des formulations précédentes si le ou les champs sont interpolés d'une part sur tout le domaine Ω et d'autre part sur tout ou partie du contour Γ de façon indépendante.

Il est possible de modifier la fonctionnelle de l'énergie complémentaire de plusieurs manières. Nous ne présentons que celle conduisant à ce que l'on appelle souvent fonctionnelle hybride et qui est la fonctionnelle de Pian et Tong (1978) :

$$J(u, \varepsilon) = -\frac{1}{2} \int_{\Omega} \sigma S \sigma + \int_{\Gamma} u \sigma(v) \cdot n - \int_{\Gamma_N} g_N v \quad (8.53)$$

Sa variation :

$$\delta J(u, \varepsilon) = - \int_{\Omega} \sigma S \sigma + \int_{\Gamma} u \sigma(v) \cdot n + v \sigma(u) \cdot n - \int_{\Gamma_N} g_N v \quad (8.54)$$

a l'avantage de ne pas comporter de dérivée.

Au paragraphe 11.1, nous reprendrons ces formulations pour les exprimer, peut-être de manière plus « explicite » pour le lecteur sous forme matricielle.

8.3.3 Équation des plaques

Soit Ω un ouvert borné régulier de \mathbb{R}^n , l'équation des plaques est :

$$\Delta(\Delta u) = f \text{ dans } \Omega \quad (8.55)$$

Nous considérons les conditions aux limites suivantes :

$$\begin{cases} u = 0 & \text{sur } \Gamma = \partial\Omega \\ \frac{\partial u}{\partial n} = 0 & \text{sur } \Gamma \end{cases} \quad (8.56)$$

La formulation variationnelle correspondante est :

$$\int_{\Omega} \Delta u \Delta v = \int_{\Omega} f v, \quad \forall v \in V \quad (8.57)$$

où V est l'espace des fonction $C^2(\overline{\Omega})$ vérifiant les conditions aux limites, i.e. telles que $v = 0$ et $\partial v / \partial n = 0$ sur Γ (démonstration 1 ci-dessous), ou V est l'espace $H_0^2(\Omega)$ (démonstration 2 ci-dessous).

Démonstration. Nous allons faire l'effort ici d'écrire la démonstration.

À priori, une solution u de l'équation des plaques est une fonction de $C^4(\overline{\Omega})$.

Soit u une solution régulière de l'équation des plaques (8.55). Pour toute fonction $v \in V$, on a :

$$\int_{\Omega} \Delta(\Delta u) v = \int_{\Omega} f v$$

Par intégration par parties, il vient :

$$\int_{\Omega} \Delta(\Delta u) v = - \int_{\Omega} \nabla(\Delta u) \cdot \nabla v + \int_{\Gamma} \frac{\partial(\Delta u)}{\partial n} v$$

or $v = 0$ sur Γ , d'où :

$$\int_{\Omega} \Delta(\Delta u) v = - \int_{\Omega} \nabla(\Delta u) \cdot \nabla v$$

Une seconde intégration par parties conduit à :

$$\int_{\Omega} \Delta(\Delta u) v = \int_{\Omega} \Delta u \Delta v - \int_{\Gamma} \Delta u \frac{\partial v}{\partial n}$$

et comme $\partial v / \partial n = 0$, il reste :

$$\int_{\Omega} \Delta(\Delta u) v = \int_{\Omega} \Delta u \Delta v$$

Finalement, on obtient bien :

$$\int_{\Omega} \Delta u \Delta v = \int_{\Omega} f v, \quad \forall v \in V$$

Vérifions la réciproque. Supposons donc maintenant que u est une solution du problème (8.57). En effectuant deux intégration par parties successives, il vient :

$$\int_{\Omega} (\Delta(\Delta u) - f) v = 0$$

pour toute fonction $v \in V$. Or V est un sous-espace dense de $L^2(\Omega)$, et donc :

$$\Delta(\Delta u) - f = 0$$

□

Autre démonstration par application du théorème de Lax-Milgram. La formulation variationnelle associée à l'équation des plaques consiste à déterminer $u \in H_0^2(\Omega)$ tel que :

$$a(u, v) = b(v) \quad \forall v \in H_0^2(\Omega)$$

avec :

$$a(u, v) = \int_{\Omega} \Delta u \Delta v \quad \text{et} \quad b(v) = \int_{\Omega} f v$$

La seule hypothèse non triviale à vérifier afin de pouvoir appliquer le théorème de Lax-Milgram est la coercitivité de la forme bilinéaire $a(., .)$. Pour tout $u \in H_0^2(\Omega)$, deux intégrations par parties successives conduisent à :

$$a(u, u) = \sum_{i,j} \int_{\Omega} \left| \frac{\partial^2 u}{\partial x_i \partial x_j} \right|^2 = \|\nabla^2 u\|_{L^2(\Omega)}^2$$

L'application de l'inégalité de Poincaré deux fois permet de conclure à l'existence de deux constantes positives C et C' telles que, $\forall u \in H_0^2(\Omega)$ on ait :

$$\|u\|_{L^2(\Omega)}^2 \leq C \|\nabla u\|_{L^2(\Omega)}^2 \leq C' \|\nabla^2 u\|_{L^2(\Omega)}^2 = C' a(u, u)$$

et donc, il existe une constante positive C telle que $\forall u \in H_0^1(\Omega)$:

$$\|u\|_{H^2(\Omega)}^2 \leq C a(u, u)$$

La coercitivité de $a(., .)$ est établie, et le théorème de Lax-Milgram nous permet de conclure à l'existence et à l'unicité de la solution.

Vérifions maintenant que, réciproquement, la solution du problème variationnel est solution du problème avec conditions aux limites.

Pour cela, considérons ω un ouvert inclus dans le compact Ω . Il existe une fonction $\chi \in C_c^\infty(\Omega)$ telle que $\chi = 1$ sur ω . Pour toute fonction $v \in C_c^\infty(\Omega)$ de support inclus dans ω :

$$\int_{\Omega} \chi \Delta u \Delta v = \int_{\Omega} \Delta u \Delta v = \int_{\Omega} f v$$

Or on sait (enfin dira-ton, sinon on admettra) que : si w et f sont des fonctions de $L^2(\Omega)$ vérifiant pour toute fonction $v \in C_c^\infty(\Omega)$:

$$-\int_{\Omega} w \Delta v = \int_{\Omega} f v$$

alors $(\chi w) \in H^2(\Omega)$ quelle que soit la fonction $\chi \in C_c^\infty(\Omega)$.

D'où, l'on déduit que $\chi \Delta u \in H^2(\Omega)$. Il est par conséquent licite d'effectuer deux intégrations par parties successives, et l'on obtient :

$$\int_{\Omega} \Delta(\chi \Delta u) v = \int_{\Omega} f v$$

pour toute fonction $v \in C_c^\infty(\Omega)$ de support inclus dans ω . On en déduit que pour presque tout $x \in \omega$:

$$\Delta(\Delta u)(x) = f(x)$$

En considérant une suite ω_n de compacts dont l'union est Ω , on en déduit que cette relation reste valable pour presque tout $x \in \Omega$.

Enfin, comme $u \in H_0^2(\Omega)$, la solution du problème variationnel vérifie automatiquement les conditions aux limites. \square

8.4 Équations de l'acoustique

On s'intéresse à la propagation et à la réflexion d'ondes de pression dans un fluide parfait non pesant. On supposera le mouvement harmonique autour d'un état moyen (ambiance) au repos (ce qui est le lien « physique » avec l'introduction de l'espace $L_0^2(\Omega)$ présenté pour l'équation de Stokes).

On rappelle que l'équation des ondes acoustiques de célérité c dans un milieu est :

$$\Delta p' - \frac{1}{c^2} \ddot{p}' = 0 \tag{8.58}$$

8.4.1 Équation de Helmholtz

Il s'agit de la formulation générale d'un problème acoustique linéaire. Elle est obtenue en cherchant la solution harmonique de l'équation des ondes, i.e. une solution sous la forme : $p'(x, y, z, t) = p(x, y, z)e^{i\omega t}$. (Au passage, les physiciens utilisent plutôt j au lieu de i pour gagner en clarté sans doute... comme s'ils passaient leur temps à parler d'intensité, qu'en plus ils notent plutôt I !)

On obtient alors l'équation de Helmholtz :

$$\Delta p + k^2 p = 0 \quad (8.59)$$

où k est le nombre d'onde :

$$k = \frac{\omega}{c} = \frac{2\pi}{\lambda} \quad (8.60)$$

8.4.2 Conditions aux limites en acoustique

Soit $\Gamma = \partial\Omega$ la frontière du domaine concerné. On considère que Γ est partitionnée en trois zones (i.e. dont l'union vaut Γ et dont aucune n'intersecte les autres) qui portent chacune une condition aux limites différente : Γ_D de Dirichlet, Γ_N de Neumann et Γ_R de Robin. On a donc :

- Dirichlet : $p = \bar{p}$ sur Γ_D ;
- Neumann : $\partial_n p = -i\rho\omega\bar{V}_n$ sur Γ_N ;
- Robin : $\partial_n p = -i\rho\omega A_n p$ sur Γ_R .

Remarques.

Condition de Dirichlet : La condition est à comprendre comme « la pression a une valeur prescrite sur la frontière ». Sa traduction mathématique est alors plutôt :

$$\frac{\partial p}{\partial n} = \bar{p} \quad \text{sur } \Gamma_D$$

Condition de Neumann : \bar{V}_n est la valeur imposée de la composante normale de la vitesse. Cette condition limite modélise la vibration d'un panneau, ce qui en fait la condition la plus souvent utilisée lorsque l'on a un couplage faible entre la structure et le fluide. On effectue alors une étude vibratoire de la structure, puis grâce à celle-ci une étude acoustique du fluide sans se préoccuper des interactions fluide-structure.

Condition de Robin : A_n est appelé coefficient d'admittance, il vaut l'inverse de l'impédance Z_n : $A_n \cdot Z_n = 1$. Physiquement, ce coefficient représente l'absorption et la réflexion de l'onde par le matériau environnant dont la frontière est modélisée par Γ_R .

Nous aurons l'occasion de reparler de cela au chapitre 19 qui portera spécifiquement sur l'acoustique.

8.4.3 Formulation faible

En notant $H_D^1(\Omega) = \{p \in H^1(\Omega), p = \bar{p} \text{ sur } \Gamma_D\}$ et $H_D^0(\Omega) = \{v \in H^1(\Omega), v = 0 \text{ sur } \Gamma_D\}$, les espaces qui sont des sous-espaces de $H^1(\Omega)$, on trouve finalement :

$$a(p, \tilde{w}) = b(\tilde{w}), \quad \forall w \in H_D^0(\Omega) \quad (8.61)$$

avec $\tilde{\cdot}$ la conjugaison complexe, $a(p, q) : H_D^1 \times H_D^1 \rightarrow \mathbb{C}$ la forme sesquilinéaire définie par :

$$a(p, q) = \int_{\Omega} \partial_{x_i} p \partial_{x_i} \tilde{q} - k^2 p \tilde{q} + \int_{\Gamma_R} i\rho c k A_n \tilde{q} \quad (8.62)$$

et $b(a) : H_D^1 \rightarrow \mathbb{C}$ définie par :

$$b(q) = - \int_{\Gamma_N} i\rho c k \bar{V}_n q \quad (8.63)$$

Et comme on peut vérifier que l'on est dans un cas qui va bien, on définit l'énergie potentielle totale du domaine par :

$$W_p = \frac{1}{2}a(p, \tilde{p}) - b(\tilde{p}) \quad (8.64)$$

et la résolution du problème devient :

$$\text{Trouver } p \in H_D^1(\Omega) \text{ tel que } W_p \text{ soit minimal} \quad (8.65)$$

ou de manière équivalente :

$$\text{Trouver } p \in H_D^1(\Omega) \text{ tel que } \delta W_p = 0, \quad \forall \delta p \in H_0^1(\Omega) \quad (8.66)$$

En fonction des conditions aux limites et des conditions initiales du problème, la variation de pression solution du problème peut être réelle, i.e. être de la forme $p'(x, y, z, t) = \Re(p(x, y, z)e^{i\omega t})$. Une telle solution est appelée monochromatique. On pourra alors résoudre le problème en restant dans \mathbb{R} au lieu de passer dans \mathbb{C} . Cela se produit lorsque l'on résout l'équation d'Helmholtz (8.59) assortie uniquement de conditions aux limites de Dirichlet.

Chapitre 9

Exemple de formulation variationnelle d'un problème de Neumann

Résumé — Dans cet exemple, portant sur formulation d'un problème de Neumann, nous allons essayer de montrer comment la réflexion mathématique se fait et évolue « au fil de l'eau » pour transformer un problème initial donné et obtenir les bonnes conditions d'existence et d'unicité de la solution sur les « espaces qui vont bien » (et qui eux, feront ensuite l'objet d'une discrétisation numérique).

Supposons que nos « investigations » sur un problème nous aient conduit à sa formulation sous forme d'un problème de Neumann (avec $f \in L^2(\Omega)$, ce qui n'est pas une condition si forte finalement, et c'est ce que l'on aimerait avoir) :

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ \frac{\partial u}{\partial n} = 0 & \text{sur } \Gamma = \partial\Omega \end{cases} \quad (9.1)$$

Nous allons donc nous poser la question de l'existence et de l'unicité de la solution de ce problème.

9.1 Étude directe de l'existence et l'unicité de la solution

Sans être magicien, il doit sembler à peu près évident qu'une condition nécessaire d'existence doit être que $f \in L_0^2(\Omega)$, et que les conditions imposées semblent un peu courtes pour assurer l'unicité.

On voit en effet que l'existence implique que

$$\int_{\Omega} -\Delta u = \int_{\Gamma} \frac{\partial u}{\partial n} = 0 \quad (9.2)$$

qui conduit à :

$$\int_{\Omega} f = 0 \quad (9.3)$$

i.e. $f \in L_0^2(\Omega)$. Donc $f \in L_0^2(\Omega)$ est bien une condition nécessaire d'existence.

De même, d'après ce qui précède, on voit que si u est solution, alors $u + c$ est solution, ce qui prouve qu'il n'y a pas unicité de la solution.

9.2 Formulation variationnelle

Nous allons écrire la formulation variationnelle de notre problème dans $H^1(\Omega)$. Nous regarderons si nous retrouvons la même condition d'existence. Nous nous demanderons ensuite si, en nous restreignant à un certain espace, il n'est pas possible d'obtenir l'unicité de la solution.

Nous commençons donc par écrire notre problème sous la forme :

$$\int_{\Omega} -\Delta u \cdot v = \int_{\Omega} \nabla u \cdot \nabla v - \int_{\Gamma} \frac{\partial u}{\partial n} v \quad (9.4)$$

et comme le dernier terme est nul (condition aux limites), il reste :

$$\int_{\Omega} -\Delta u \cdot v = \int_{\Omega} \nabla u \cdot \nabla v \quad (9.5)$$

Comme nous voulons la formulation variationnelle dans $H^1(\Omega)$ (i.e. $u \in H^1(\Omega)$), il vient :

$$\int_{\Omega} \nabla u \cdot \nabla v = \int_{\Omega} f v, \quad \forall v \in H^1(\Omega) \quad (9.6)$$

Si nous considérons $u = u_0 + c$ et $v = v_0 + d$, alors il vient :

$$\int_{\Omega} \nabla u_0 \cdot \nabla v_0 = \int_{\Omega} f v_0 + d \int_{\Omega} f, \quad \forall v_0 \in H^1(\Omega), \quad \forall d \quad (9.7)$$

ce qui implique, pour que ce soit vrai pour tout d que $\int_{\Omega} f = 0$, et on retrouve bien que $f \in L_0^2(\Omega)$ est une condition nécessaire d'existence.

Méزالor, il est naturel de considérer $u_0 \in H^1(\Omega) \cap L_0^2(\Omega)$ (car on veut que u et v vivent dans le même espace pour pouvoir appliquer nos « méthodes »), et la formulation précédente donne :

$$\int_{\Omega} \nabla u_0 \cdot \nabla v_0 = \int_{\Omega} f v_0, \quad \forall v_0 \in H^1(\Omega) \cap L_0^2(\Omega) \quad (9.8)$$

i.e. nous sommes amenés à considérer l'espace $V = H^1(\Omega) \cap L_0^2(\Omega)$.

L'espace V est un espace de Hilbert normé par $\|\cdot\|_{H^1(\Omega)}$.

- $v \in V \Rightarrow v \in H^1(\Omega)$ tel que $\int_{\Omega} v = 0$ et l'application $v \in H^1(\Omega) \mapsto \int_{\Omega} v$ est continue sur $H^1(\Omega)$, donc V est un sous-espace fermé de H^1 , donc est un Hilbert. Et donc maintenant que l'on sait que V est un Hilbert, on aimerait bien appliquer Lax-Milgram (voir 7.3), c'est pourquoi il faut s'intéresser à la forme bilinéaire définissant le problème. C'est ce qui suit.
- Si l'on considère la forme bilinéaire :

$$a(u_0, v_0) = \int_{\Omega} \nabla u_0 \cdot \nabla v_0 \quad (9.9)$$

alors on voit que

$$a(u_0, v_0) \leq |u_0|_1 |\overline{v_0}|_1 \leq \|u_0\|_{H^1} \|v_0\|_{H^1} \quad (9.10)$$

et donc a est **continue**.

- $(v_0, v_0) = |v_0|_1^2$. Or on a l'inégalité de type Poincaré suivante : $\exists c(\Omega) > 0$ tel que :

$$|\varphi|_{0,\Omega} \leq c(\Omega) |\varphi|_{1,\Omega}, \quad \forall \varphi \in V \quad (9.11)$$

Cela conduit à :

$$\frac{|v_0|_1^2}{|v_0|_1^2 + |v_0|_0^2} \geq \frac{1}{1 + c^2(\Omega)}, \quad \forall v \in V \quad (9.12)$$

Donc a est **V-elliptique**.

Le théorème de Lax-Milgram permet donc de conclure à l'existence et à l'unicité de la solution $u_0 \in V$.

9.3 Formulation mixte duale

On va regarder maintenant ce qui se passe si l'on exprime le problème sous forme mixte, i.e. en se servant de deux variables. C'est une voie fréquente pour « simplifier » les conditions aux limites.

En effet, plutôt que considérer $\frac{\partial u}{\partial n} = 0$ sur $\Gamma = \partial\Omega$, on peut interpréter cela comme une pression nulle sur les bords, i.e. $p \cdot n = 0$ sur Γ .

Le problème considéré se formule alors de la manière suivante :

$$\begin{cases} -\Delta u + p = 0 & \text{loi de comportement dans } \Omega \\ \operatorname{div} p = -f & \text{loi de conservation dans } \Omega \\ p \cdot n = 0 & \text{condition aux limites sur } \Gamma \end{cases} \quad (9.13)$$

On peut alors écrire la loi de conservation :

$$\int_{\Omega} (\operatorname{div} p) v = - \int_{\Omega} f v, \quad \forall v \in L^2(\Omega) \text{ (à priori)} \quad (9.14)$$

et la loi de comportement :

$$0 = \int_{\Omega} (p - \nabla u) \cdot q = \int_{\Omega} p \cdot q + \int_{\Omega} u \operatorname{div} q = \langle u, q \cdot n \rangle \quad (9.15)$$

On n'a pas d'information sur u . Par contre, on a un terme $p \cdot n = 0$. Et il serait bien de disposer de son « symétrique », i.e. d'un terme $q \cdot n = 0$ qui serait imposé par le choix de l'espace dans lequel vivrait u . Ce que l'on a en tête c'est bien sûr de pouvoir utiliser le théorème de Brezzi.

En correspondance avec les notations du paragraphe 7.5, on choisit alors les espaces : $H = \{q \in H(\operatorname{div}; \Omega); q \cdot n|_{\Gamma} = 0\}$ et $M = L^2(\Omega)$. Le problème est donné, dans la formulation de Brezzi, par les deux formes bilinéaires

$$a(p, q) = \int_{\Omega} p \cdot q \quad \text{et} \quad b(v, q) = \int_{\Omega} v \operatorname{div} q \quad (9.16)$$

Quant aux formes linéaires définissant le problème, la première est nulle et la seconde vaut :

$$- \int_{\Omega} f v \quad (9.17)$$

Maintenant que nous nous sommes rapprochés du cadre du théorème de Brezzi, il va nous falloir vérifier si toutes les conditions sont vérifiées. Toutefois, d'après ce qui a été vu dans les paragraphes précédents, il semble naturel de se demander s'il faudra bien conserver $L^2(\Omega)$, ou s'il faudra passer à $L_0^2(\Omega)$, et ce qu'il se passe quand $u = u_0 + c$.

Pour l'instant, on a :

$$\int_{\Omega} v \operatorname{div} p = - \int_{\Omega} f v, \quad \forall v \in L^2(\Omega) \quad (9.18)$$

Si l'on prend $v = 1$, alors il vient :

$$\int_{\Omega} \operatorname{div} p = - \int_{\Gamma} p \cdot n = 0 \quad (9.19)$$

ce qui implique que $\int_{\Omega} f = 0$, soit $f \in L_0^2(\Omega)$. On a également :

$$\int_{\Omega} p \cdot q + \int_{\Omega} u \operatorname{div} q = 0, \quad \forall q \in H \quad (9.20)$$

Or $q \in H$ implique $\int_{\Omega} \operatorname{div} q = 0$, ce qui implique aussi que si u est solution, alors $u + c$ l'est aussi.

Encore une fois il n'y a pas unicité de la solution. Pour l'imposer, il faut donc une condition sur u , comme par exemple $\int_{\Omega} u = 0$, i.e. $u \in L_0^2(\Omega)$. Si de plus $\operatorname{div} p \in L_0^2(\Omega)$ (et avec $f \in L_0^2(\Omega)$ comme vu au dessus), alors ce implique que :

$$\int_{\Omega} (\operatorname{div} p + f)(v_{L_0^2} + c) = \int_{\Omega} (\operatorname{div} p + f)v_{L_0^2} \quad (9.21)$$

Donc le choix $M = L_0^2(\Omega)$ semble un choix convenable à priori.

Il reste alors maintenant à **vérifier que toutes les conditions du théorème de Brezzi** sont bien satisfaites pour pouvoir conclure à l'existence et à l'unicité de la solution. Ces conditions sont :

- H et M sont des Hilbert ;
- a , b et f sont continues ;
- a est V-elliptique ;
- condition inf-sup sur b .

H est un Hilbert Pour montrer que H est un Hilbert, on s'appuie sur le fait que l'application : $\varphi : q \in H(\operatorname{div}; \Omega) \mapsto q \cdot n \in H^{-1/2}(\Gamma)$ est continue (car $\|q, 0\|_{-1/2, \Gamma} \leq \|q\|_{H(\operatorname{div}; \Omega)}$). Il en résulte alors que $H = \varphi^{-1}(0)$ est fermé et que $H(\operatorname{div}; \Omega)$ étant un Hilbert, c'est également un Hilbert pour la norme de $H(\operatorname{div})$.

M est un Hilbert Pour montrer que M est un Hilbert, on s'appuie sur le fait que l'application : $\varphi : v \in L^2(\Omega) \mapsto \int_{\Omega} v = 0$ est continue car $|\int_{\Omega} v| \leq \sqrt{|\Omega|} \|v\|_{L^2(\Omega)}$. Il en résulte que $M = \varphi^{-1}(0)$ est fermé et que $L^2(\Omega)$ étant un espace de Hilbert, c'est également un espace de Hilbert pour la norme de $L^2(\Omega)$.

a est continue

$$a(p, q) = \int_{\Omega} p \cdot q \leq |p|_0 |q|_0 \leq \|p\|_{H(\operatorname{div}; \Omega)} \|q\|_{H(\operatorname{div}; \Omega)} \quad (9.22)$$

et donc a est continue et $\|a\| \leq 1$.

b est continue

$$b(v, q) = \int_{\Omega} v \operatorname{div} q \leq |v|_0 |\operatorname{div} q|_0 \leq |v|_0 \|q\|_{H(\operatorname{div}; \Omega)} \quad (9.23)$$

et donc b est continue et $\|b\| \leq 1$.

f est continue

$$\int_{\Omega} f v \leq |f|_0 |v|_0 \quad (9.24)$$

ce qui implique que la norme de la forme linéaire est $\leq |f|_0$.

a est V-elliptique On a :

$$V = \left\{ q \in H; \int_{\Omega} v \operatorname{div} q = 0, \quad \forall v \in L_0^2(\Omega) \right\} \quad (9.25)$$

$q \in H$ implique que $\int_{\Omega} (v + c) \operatorname{div} q = 0$, et donc $\int_{\Omega} w \operatorname{div} q = 0, \forall w \in L^2(\Omega)$, ce qui finalement implique que $\operatorname{div} q = 0$ dans $L^2(\Omega)$.

On a donc $V = \{q \in H; \operatorname{div} q = 0\}$.

On a alors :

$$a(q, q) = \int_{\Omega} |q|^2 = \|q\|_{H(\operatorname{div}; \Omega)}^2 \quad (9.26)$$

si $q \in V$, et donc la V-ellipticité de a est démontrée.

Condition inf-sup Pour montrer la conditions inf-sup, nous construirons q à partir d'un φ vérifiant $\frac{\partial \varphi}{\partial n} = 0$ sur Γ , i.e. :

$$\forall v \in M, \quad \exists q \in H; \quad \int_{\Omega} v \operatorname{div} q = |v|_0^2 \quad \text{et} \quad \|q\|_{H(\operatorname{div})} \leq \frac{1}{\beta} |v|_0 \quad (9.27)$$

$q \in H$, $\operatorname{div} q = v$ et $q = -\nabla \varphi$, $q \cdot n = 0$ donc $\frac{\partial \varphi}{\partial n} = 0$, et :

$$\begin{cases} -\Delta \varphi = v \\ \frac{\partial \varphi}{\partial n} = 0 \end{cases} \quad (9.28)$$

Comme au paragraphe sur la formulation variationnelle, $\exists! \varphi \in H^1(\Omega) \cap L_0^2(\Omega)$ car $v \in L_0^2(\Omega)$:

$$|q|_0^2 = \int_{\Omega} |\nabla \varphi|^2 = |\varphi|_1^2 = \int_{\Omega} v \varphi \leq |v|_0 |\varphi|_0 \leq c(\Omega) |V|_0 |\varphi|_1 \quad (9.29)$$

d'où $|a|_0 \leq c(\Omega) |V|_0$. De plus :

$$|q|_0^2 + |\operatorname{div} q|_0^2 = |q|_0^2 + |v|_0^2 \leq (C^2(\Omega) + 1) |v|_0^2 \quad (9.30)$$

d'où :

$$\beta = \frac{1}{\sqrt{1 + c^2(\Omega)}} \quad (9.31)$$

On peut donc appliquer le théorème de Brezzi et conclure à l'existence et à l'unicité de la solution.



ÉLÉMENTS FINIS

Introduction

En introduction à cette partie, il nous semblait important d'en exposer sa structure, car elle peut sembler un peu décousue à la simple lecture de la table des matières.

Après les pré-requis exposés dans les deux premières parties, la partie III va s'ouvrir « tout naturellement » sur une présentation générale de la méthode des éléments finis au chapitre 10.

Immédiatement après, compte tenu du public visé, le chapitre 11 essaiera de mettre en avant les spécificités et surtout la complexité de la mécanique comme champ d'application de la méthode des éléments finis.

Cette mise en garde, au regard de l'expérience, nous semble importante : on a tendance souvent à considérer que la mécanique est quelque chose de très bien maîtrisé, et ce n'est pas le cas. Bien des sujets restent pointus, voire ouverts. Il convient donc de rester prudent, surtout pour ceux ayant une expérience de calcul importante qui les porte parfois à sous estimer les difficultés.

Le chapitre 12, qui fait suite logiquement en terme de présentation de la méthode des éléments finis, au chapitre 10, est souvent le mieux maîtrisé par le public d'ingénieurs, au moins concernant les éléments de Lagrange. Nous l'avons complété de remarques sur les modèles à plusieurs champs (qui peut faire pour partie écho au chapitre 11). Encore une fois, pour le public visé, c'est surtout le paragraphe 12.4 sur la validation pratique des éléments finis qui aura sans doute le plus de valeur ajoutée. Le contenu de ce paragraphe fait souvent partie des choses oubliées.

Comme nous en serons sur des choses un peu oubliées, nous en profiterons au chapitre 13 pour continuer dans le même sillon et rappeler quelques méthodes d'amélioration de la performance du calcul. Nous y présentons des choses qui sont utilisées plus ou moins fréquemment par notre public. Seuls les paragraphes 13.6 et 13.7 (sur les méthodes de réanalyse et aux dérivées d'ordre élevé) sont généralement moins bien connus. Nous avons voulu les introduire ici plutôt qu'au chapitre 23 car ils sont vraiment en lien avec les préoccupations directes de notre public.

Le chapitre 14 permettra une petite pause en exposant brièvement les stratégies de maillage, et plus particulièrement la triangulation de Delaunay.

Pour continuer sur les choses pouvant avoir un réel intérêt pratique pour le public d'ingénieurs mécaniciens (ou acousticiens), nous aborderons au chapitre suivant 15 les méthodes d'homogénéisation. Si les méthodes les plus « physiques » sont bien connues, l'approche mathématique (généralisante) est bien souvent quasi totalement inconnue.

S'en suivra au chapitre 16 une introduction à l'optimisation, où nous aurons le loisir de revenir sur les multiplicateurs de Lagrange. Nous y aborderons l'optimisation de forme, dont la version « optimisation topologique » recourra aux techniques d'homogénéisation vues juste avant.

À ce niveau du document, il nous semble que nous aurons parcouru bon nombre des applications typiques de la méthode des éléments finis, surtout dans le cadre de la mécanique... mais uniquement sous l'angle statique !

Il sera donc temps d'aborder « le temps », i.e. les problèmes non stationnaires, au chapitre 17. La propagation des ondes, quant à elle, ne sera traitée qu'au chapitre 18. C'est d'ailleurs dans ce chapitre que seront abordés également les modes propres, dont nous n'aurons pas parlé jusqu'alors (à quelques exception près lors de remarques diverses et variées... mais rien de sérieux). Le chapitre 19 continuera la spécification en se focalisant sur l'acoustique.

Dès lors, on pourra considérer qu'une présentation assez complète de la méthode des éléments

finis a été faite. Toutefois, ce qui était encore de l'ordre de la recherche il y a une décennie fait aujourd'hui partie des phénomènes que notre public doit prendre en compte de plus en plus souvent. Ces phénomènes, un peu plus complexes, un peu plus exotiques, sont souvent liés à ce que l'on appelle la ou plutôt les non linéarités. C'est ce qui sera abordé au chapitre 20.

Parmi toutes les non linéarités, c'est celle des lois de comportement qu'il nous est demandé d'exposer en priorité. Les problèmes de contact ou de grands déplacements, qui seront abordés dans ce chapitre, nous sont bien moins demandés car souvent traités par des personnes très au fait des méthodes et même souvent en lien avec la recherche dans le domaine.

Quant au traitement de la modélisation de la rupture en mécanique, le chapitre 21 y sera consacré. Il s'agit néanmoins d'une affaire de spécialistes, et nous ne ferons qu'aborder le sujet.

Le chapitre 22 présentera brièvement quelques méthodes probabilistes pour la prise en compte des aléas dans les équations considérées. C'est encore une fois sous l'angle de la « mécanique aléatoire » que nous illustrerons cela, en incluant une présentation de l'indice de fiabilité d'une structure.

Le chapitre 23 permet de mettre l'accent sur le fait que même si la méthode des éléments finis est une méthode extrêmement générale, performante et répandue, elle n'est qu'une méthode numérique parmi une multitude d'autres méthodes. Ce chapitre n'a pas vocation bien évidemment à approfondir aucune des méthodes abordées, mais uniquement à fournir un petit complément culturel qui nous semblait indispensable dans le monde actuel.

Le chapitre 24 clôt cette partie sur un petit rappel lié aux singularités, juste comme une petite piqure de rappel sur des problèmes dont les ingénieurs pratiquant le calcul numérique restent assez conscients. Il nous a semblé que passer les singularités sous silence pouvait laisser à penser que ce problème n'était peut-être pas si important, ce qui n'est pas le cas bien évidemment.

Chapitre 10

La méthode des éléments finis

Résumé — Une fois le travail précédent accompli, i.e. une fois que l'on dispose d'une formulation faible, « il n'y a plus qu'à » calculer la solution ! La méthode des éléments finis est l'un des outils numérique développé pour cela.

La méthode des éléments finis se propose de mettre en place, sur la base de formulations faibles, un algorithme discret (discrétisation) permettant de rechercher une solution approchée d'un problème aux dérivées partielles sur un domaine compact avec conditions aux bords et/ou dans l'intérieur du compact.

Il s'agit donc de répondre aux questions d'existence et d'unicité de la solution, de stabilité et de convergence des méthodes numériques, ainsi que d'apprécier l'erreur entre la solution exacte et la solution approchée (indicateurs et estimateurs d'erreur, *a priori* et *a posteriori*).

10.1 Introduction

Les chapitres des parties précédentes ont eu pour but de nous permettre de décrire un problème physique à partir d'équations aux dérivées partielles, mais à aucun moment nous n'avons encore parlé de la manière de résoudre ces équations (que ce soit la formulation forte ou la faible). Ce sera le but de ce chapitre, dans lequel la mise en œuvre de la méthode des éléments finis va être exposée.

Nous avons vu dans la partie précédente, comment passer d'une formulation forte à une formulation faible. Dans de nombreux cas, nous avons montré qu'il y a équivalence (existence et unicité de la solution) entre l'étude des deux formulations. Il faut toutefois bien garder en tête que dans certains cas, l'équivalence entre les deux formulations n'est pas évidente.

Une condition qui n'a pas été vraiment évoquée jusque là est le cas où le bord du domaine n'est pas suffisamment régulier, par exemple s'il possède des points singuliers (par exemple point d'inflexion ou de rebroussement, ou dans le cas des fissures en MMC).

Cela peut également se produire si l'on ne peut pas assurer que la fonction f (celle agissant dans tout le domaine) n'est pas suffisamment dérivable. Comme en général on suppose dans les problèmes physiques que la solution est C^∞ , on n'est pas confronté à ce genre de problème.

L'idée de la méthode des éléments finis est de décomposer (on dit discrétiser) le domaine Ω en un certain nombre de sous-domaines (les éléments). Les éléments recouvriront l'intégralité du domaine (de la frontière pour les éléments de frontière qui est une autre méthode) et sans chevauchement entre eux (les éléments peuvent se chevaucher dans la méthode des volumes finis). De plus, on va chercher la fonction solution u comme étant interpolée par des « bouts » de solutions définis sur chaque élément.

Le problème étant interpolé sur les éléments, on se doute que le nombre d'éléments va jouer sur la qualité de cette approximation de la solution. On se doute également que, comme on résout un problème comportant des dérivées, c'est plutôt dans les endroits où la solution va varier vite qu'il sera nécessaire de « resserrer » le maillage.

On veillera également à ce que les éléments ne soient pas trop distordus. Des critères dits « de forme » existent. Néanmoins, dans certains cas il est possible de violer allègrement ces critères tout en obtenant une très bonne approximation de la solution.

Histoire

C'est l'ingénieur américain Ray William Clough qui, semble-t-il, a utilisé le terme de méthode des éléments finis le premier dans un article de 1960 intitulé *The Finite Element Method in Plane Stress Analysis*. Le mot rigidité (Stiffness) apparaissait dans le titre de son article *Stiffness and Deflection Analysis of Complex Structures* datant de 1956 (et coécrit avec M. Turner, H. C. Martin et L. J. Topp).

Si on veut replacer très brièvement cela dans un contexte plus global, on peut dire que l'analyse des structures est née vers 1850.

La RdM, recourant au calcul manuel, était développée par Maxwell, Castigliano, Mohr.

Le concept d'éléments finis est né vers 1940, avec des figures comme Newmark, Hrennikoff (1941), Mc Henry, Courant (1942)...

Son réel essor ne commence toutefois que dans les années 60 avec le développement du calcul numérique sur ordinateur.

La méthode des éléments finis (MEF) prend ses origines dans le besoin de résoudre des problèmes complexes d'élasticité et d'analyse de structures en ingénierie civile et aéronautique. Son développement remonte aux travaux d'Alexander Hrennikoff (1941) et de Richard Courant (1942). Bien qu'utilisant des approches différentes, ces deux pionniers partagent la même caractéristique essentielle à savoir la discrétisation par maillage du domaine continu en sous-domaines discrets, que l'on appelle éléments. C'est Olgierd Zienkiewicz de l'Imperial College qui synthétisa ces deux méthodes en ce que l'on peut appeler la méthode des éléments finis et qui fit la première formalisation mathématique de la méthode.



Clough Courant Zienkiewicz



Rayleigh Ritz Galerkin

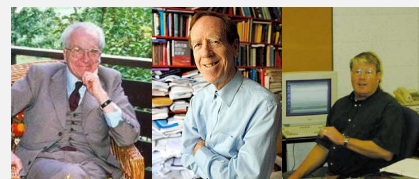
Dans ses travaux, Hrennikoff discrétise le domaine en utilisant une analogie avec les treillis, tandis que l'approche de Courant divise le domaine en sous-régions finies triangulaires pour résoudre les équations aux dérivées partielles elliptiques du second ordre issues du problème de la torsion d'un cylindre. On peut dire que la contribution de Courant était une évolution s'appuyant sur un vaste corpus de résultats antérieurs pour les équations aux dérivées

partielles développés par Rayleigh, Ritz et Galerkin.

Le développement de la méthode des éléments finis a véritablement commencé au milieu de années 1950 pour l'analyse structurale et aéronautique, et prit de l'ampleur à l'Université de Stuttgart grâce au travail de John Argyris et à Berkeley grâce au travail de Ray W. Clough [26]. Ray Clough est également l'un des pionniers du génie para sismique et s'est vu décerné en 2008, à la World Conference of Earthquake Engineering en Chine, le titre de « légende du génie para sismique » ("Legend of Earthquake Engineering").

À la fin des années 50, les concepts clés de matrice de rigidité et d'assemblage d'éléments existaient quasiment sous la forme actuelle. La NASA publia une demande de propositions pour le développement du logiciel d'éléments finis NASTRAN en 1965.

La base mathématique rigoureuse de la méthode des éléments finis a été consolidée en 1973 avec la publication de Strang et Fix de *An Analysis of The Finite Element Method*. Elle a depuis été intégrée comme une branche des mathématiques appliquées à la modélisation numérique des systèmes physiques dans une large variété de disciplines. Pour une discussion plus approfondie des apports et contributions relatives des différents pionniers de cette méthode, on pourra se référer à [38].



Argyris Strang Fix

Bien que la méthode des éléments finis soit théoriquement généralisable à toutes les dimensions d'espace et à tous les ordres de dérivation, dans la pratique, l'augmentation de ces paramètres

accroît de manière dramatique la difficulté et on se contente de résoudre des problèmes limités à trois dimensions en espace et deux ordres de dérivation.

Par exemple, un problème à trois dimensions d'espace et une dimension de temps, qui pourrait être discrétisé par une formulation à quatre dimensions (les éléments finis espace-temps existent), sera traité par une discrétisation éléments finis à trois dimensions d'espace couplé à un schéma en différences finis en temps.

Pour des problèmes avec des ordres de dérivations supérieurs (des techniques dites d'ordres élevés existent), comme par exemple en statique des poutres où l'on a une dérivation partielle d'ordre 4, on introduit des variables supplémentaires afin de diminuer les ordres des dérivées (déformations, contraintes...). cela a été évoqué en parlant du « choix des variables » en mécanique, nous détaillerons un peu plus loin la formulation éléments finis de ces formulations mixtes).

10.2 Problèmes de la modélisation « réelle »

Pour faire suite aux dernières remarques faites au paragraphe précédent, il faut se rendre à l'évidence que les équations qui modélisent les principales théories de la physique ont été présentées (et développées) dans un cadre idéalisé. Dans les problèmes réels, il est souvent difficile de rester dans ce cadre à cause de nombreuses difficultés telles que les problèmes géométriques, d'échelle et du couplage de différents modèles.

Toutefois, cela ne signifie pas qu'il est alors impossible de simuler le problème, mais seulement qu'il est nécessaire d'ajouter encore un peu plus d'intelligence et d'astuce. En effet, ces difficultés peuvent souvent être découplées dans un calcul numérique par des algorithmes adéquats et l'on se ramène alors à la résolution itérée de problèmes standard (d'où la nécessité de disposer de méthodes sûres et performantes pour calculer des solutions numériques de ces problèmes standard).

Nous allons présenter quelques exemples de problèmes pouvant survenir dans une modélisation, sachant qu'un certain nombre seront traités dans des chapitres ultérieurs (mais pas tous).

10.2.1 Problèmes géométriques

Jusqu'à présent, nous avons supposé que le domaine Ω dans lequel le problème était considéré était fixe. Or, dans de nombreux problèmes ce domaine est variable, on parle de problèmes à surface libre, dont voici quelques exemples :

- Le problème le plus évident pour le lecteur, et qui a déjà été évoqué, est celui de la grande déformation d'un solide.
- L'étude des mouvements d'un liquide, notamment les vagues ou le mouvement de l'eau dans un réservoir. Quand la variation de la forme du domaine n'est pas trop grande on peut définir le domaine inconnu comme l'image d'un domaine fixe par une certaine fonction. Cette fonction devient alors une inconnue du problème qui se ramène à une équation plus complexe que l'équation initiale mais sur un domaine fixe. Mais, parfois, le domaine peut être amené à se fragmenter comme dans le cas de la formation de gouttes...
- La fusion de la glace dans l'eau : la frontière entre la glace et l'eau est alors une inconnue du problème.

10.2.2 Problèmes d'échelle

La formulation d'un problème peut dépendre de l'échelle à laquelle on regarde ce problème, comme cela a été mentionné précédemment. Mais il se peut également que différentes échelles interagissent... Quelques problèmes classiques d'échelle sont :

- Nous avons déjà évoqué un peu cela en mécanique des fluides : dans l'écoulement d'un fluide la turbulence est un phénomène qui fait apparaître des mouvements à très petite échelle. La complexité de ces mouvements rend nécessaire, dans un calcul numérique, le remplacement

des valeurs exactes des champs inconnus par leur moyenne, en un sens à préciser. Cela conduit à des modèles de turbulence qui se différencient par les hypothèses supplémentaires qui sont faites.

- Dans un champ plus proche des considérations du lecteur, on pensera évidemment aux ondes. Une excitation à très haute fréquence crée une onde de longueur très petite. Or une longueur d'onde très petite ne peut pas être prise en compte dans un calcul numérique à grande échelle. La prise en compte de ce phénomène dans l'équation des ondes conduit à des modèles asymptotiques dans lesquels l'étude des ondes se ramène à la théorie de l'optique géométrique plus ou moins enrichie pour tenir compte de phénomènes comme la diffraction. Mais si la longueur d'onde est proche des longueurs des variations géométriques du bord du domaine d'étude il faut revenir à l'équation des ondes pour étudier l'effet de ces variations géométriques. On peut donc être amené à coupler l'optique géométrique et une étude directe de l'équation des ondes.

- Un autre champ lui-aussi déjà mentionné est celui des hétérogénéités d'un milieu continu, qui, quand elles sont à très petite échelle, peuvent empêcher la prise en compte exacte de celles-ci, au moins dans une approximation numérique.

C'est le cas des solides formés de matériaux composites, des fluides comme l'air chargé de gouttelettes d'eau ou de l'eau contenant des bulles de vapeur. On est conduit à définir des modèles limites dits homogénéisés dans lesquels on remplace les grandeurs usuelles (vitesse, masse volumique), qui ont une forte variation locale, par leur valeur moyenne.

Les équations obtenues peuvent avoir la même forme que les équations initiales comme dans le cas de la théorie élastique des matériaux composites (le seul problème est de calculer les propriétés matérielles du matériau homogénéisé, mais nous présenterons cela plus loin) ou bien, quand les paramètres des hétérogénéités (la densité de bulles par exemple) dépendent de la solution, ces paramètres s'ajoutent aux grandeurs inconnues initiales et le nombre d'équations peut donc s'accroître.

10.2.3 Couplage géométrique

De nombreux problèmes impliquent la prise en compte de plusieurs modèles (on dit parfois également « plusieurs physiques »).

C'est le cas de l'interaction d'un fluide et d'une structure : l'écoulement d'un fluide par exemple peut faire vibrer une structure qui en retour fait vibrer le fluide. Une des difficultés vient de ce que les inconnues utilisées dans la modélisation de chacun des milieux sont de natures différentes : les déplacements dans le solide et les vitesses dans le fluide.

Comme illustration de ce couplage, nous proposons le cas d'un poisson, issu de l'intéressante étude de San Martín et al. [62]. Le déplacement d'un poisson dans son milieu, illustré à la figure 10.1.a, n'est pas imposé mais résulte de l'interaction entre la déformation propre du poisson (donnée à la figure 10.1.b, et qui elle est imposée) et le fluide.

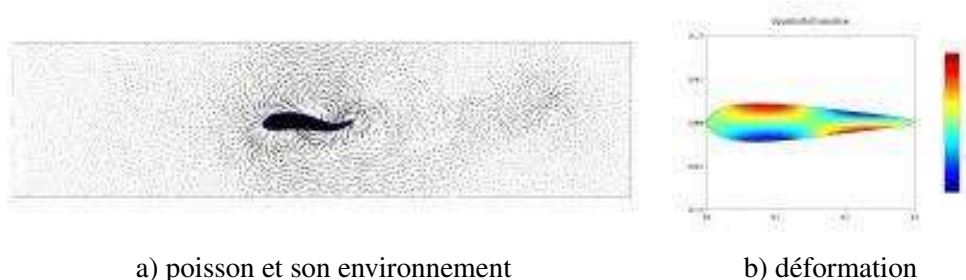


FIGURE 10.1: Interaction fluide/structure : déplacement d'un poisson

On retrouve aussi les problèmes d'interaction fluide/structure dans la modélisation de la circulation sanguine : le sang est un liquide très faiblement compressible, alors que les vaisseaux sanguins sont déformables et susceptibles de grands déplacements

Le contact entre deux solides peut également être placé dans cette catégorie.

10.2.4 Couplage intrinsèque

L'étude de la convection naturelle d'un fluide, pour la météorologie par exemple, se modélise par le couplage des équations de la dynamique des fluides d'une part, de la diffusion et du transport de la chaleur d'autre part : ce sont en effet les variations de température qui créent les variations de densité responsables du mouvement de l'air, mais ce mouvement lui-même entraîne un transport de chaleur responsable de variations de la température. Même si l'on considère un modèle simplifié linéarisé pour modéliser l'écoulement du fluide, le couplage fait apparaître une non linéarité.

L'étude des plasmas oblige à coupler les équations de dynamiques des fluides et les équations de Maxwell puisque ce sont les forces électromagnétiques qui font se mouvoir les particules chargées mais leur mouvement est lui même la cause d'un champ électromagnétique induit.

10.3 Principe de la méthode : résolution d'un système matriciel

Soit Ω un domaine ouvert de \mathbb{R}^n ($n = 1, 2$ ou 3 en pratique), de frontière $\Gamma = \partial\Omega$ et sur lequel on cherche à résoudre une équation aux dérivées partielles munie de conditions aux limites.

Ce problème, mis sous forme variationnelle, est :

$$\text{Trouver } u \in V \text{ tel que } a(u, v) = f(v), \quad \forall v \in V \quad (10.1)$$

où V est un espace de Hilbert. On supposera également que l'équation de départ a de bonnes propriétés, i.e. que l'on est dans les hypothèses vues précédemment permettant d'affirmer que le problème admet une solution unique u .

La méthode des éléments finis se propose de discrétiser le problème considéré. La discrétisation intervient à plusieurs niveaux :

discrétisation : il est nécessaire de disposer d'une description du domaine Ω sur lequel on souhaite travailler. Cette description va se faire en l'approchant par un maillage, qui sera constitué d'éléments.

interpolation : il est ensuite nécessaire de disposer d'une manière de représenter le ou les champs inconnus. Ce que se propose de faire la méthode des éléments finis, c'est d'approcher ces champs par des fonctions plus simples (disons polynomiales de degré un ou deux) définies sur chacun des éléments du maillage (le champ est approché par des bouts de fonctions qui, elles, ne sont définies chacune que sur un seul élément).

approximation : selon le type d'approximation, on remplace non seulement l'espace V (qui correspond, selon les notations utilisées au chapitre sur la formulation faible, aux espaces H , V , M ou même $H \times M$), de dimension infinie, par des approximations V_h de dimension finie, mais parfois également les formes bilinéaires et linéaires définissant le problème (nous expliquerons plus tard les motivations).

De manière classique, on notera \mathcal{T}_h le maillage de Ω considéré. Il est caractérisé par les dimensions géométriques représentatives que sont le diamètre maximum des éléments, h et le facteur de forme du maillage, σ (qui caractérise l'aplatissement du maillage).

Le maillage est constitué d'éléments, généralement notés K . On note h_K le diamètre de l'élément K , i.e. le maximum des distances (euclidiennes) entre deux points de K , et ρ_K la rondeur de l'élément K , i.e. le diamètre maximum des sphères incluses dans K .

On a donc évidemment $h_K \leq h$, $\forall K \in \mathcal{T}_h$ puisque par définition $h = \max_{K \in \mathcal{T}_h} h_K$, et $\frac{h_K}{\rho_K} \leq \sigma$, $\forall K \in \mathcal{T}_h$.

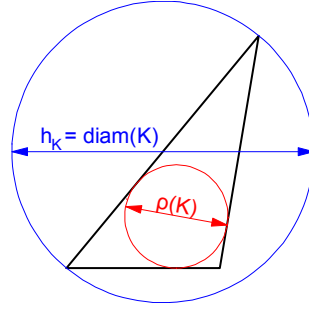


FIGURE 10.2: Dimensions géométriques caractéristiques d'un élément.

On notera \hat{K} l'élément de référence correspondant à l'élément K du maillage \mathcal{T}_h , et d'une manière générale on ajoutera ce « chapeau » $\hat{\cdot}$ sur toute quantité relative à un élément de référence (voir plus loin le chapitre sur la formulation pratique des éléments finis).

Les types d'approximation du problème (P) donné par la relation (10.1), qui seront détaillés un peu plus loin, sont donnés au tableau 10.1.

type d'approximation	espaces	formes	problème (P) \Rightarrow problème approché (P_h)
conforme	$V_h \subset V$	inchangées	Trouver $u_h \in V_h$ tq $a(u_h, v_h) = b(v_h), \forall v_h \in V_h$
non conforme	$V_h \subset V$	approchées	Trouver $u_h \in V_h$ tq $a_h(u_h, v_h) = b_h(v_h), \forall v_h \in V_h$
non conforme	$V_h \not\subset V$	approchées	Trouver $u_h \in V_h$ tq $a_h(u_h, v_h) = b_h(v_h), \forall v_h \in V_h$

Tableau 10.1: Relations approximation et espaces

En fait, on expose généralement les méthodes à partir de la méthode dite conforme selon le tableau ci-dessus, alors qu'en pratique on travaille souvent dans le second cas, i.e. approximation conforme en espace ($V_h \subset V$) mais non conforme concernant les formes, car au moins les intégrations sont réalisées numériquement.

Toujours est-il que dans tous les cas, l'espace V est remplacé par un espace V_h , de dimension finie N_h , dont (e_1, \dots, e_{N_h}) en est une base. L'approximation u_h de u dans cette base s'écrit :

$$u_h = \sum_{i=1}^{N_h} q_i e_i \quad (10.2)$$

Le problème de la méthode des éléments finis devient donc (on écrit ici en utilisant a_h , mais on pourrait utiliser a) :

$$\text{Trouver } q_1, \dots, q_{N_h} \text{ tels que } \sum_{i=1}^{N_h} q_i a_h(e_i, v_h) = f_h(v_h) \quad , \forall v_h \in V_h \quad (10.3)$$

soit, en exploitant les linéarités de $a_h(\cdot, \cdot)$ et $f_h(\cdot)$ et en décomposant v_h :

$$\text{Trouver } q_1, \dots, q_{N_h} \text{ tels que } \sum_{i=1}^{N_h} q_i a_h(e_i, e_j) = f_h(e_j) \quad , \forall j = 1, \dots, N_h \quad (10.4)$$

Il s'agit donc de résoudre le système linéaire :

$$\begin{bmatrix} a_h(e_1, e_1) & \dots & a_h(e_{N_h}, e_1) \\ \vdots & & \vdots \\ a_h(e_1, e_{N_h}) & \dots & a_h(e_{N_h}, e_{N_h}) \end{bmatrix} \begin{pmatrix} q_1 \\ \vdots \\ q_{N_h} \end{pmatrix} = \begin{pmatrix} f_h(e_1) \\ \vdots \\ f_h(e_{N_h}) \end{pmatrix} \quad (10.5)$$

que l'on note (dans la tradition mécanicienne) :

$$\mathbf{K}\mathbf{q} = \mathbf{F} \quad \text{ou encore} \quad [\mathbf{K}](\mathbf{q}) = (\mathbf{F}) \quad (10.6)$$

Telle que formulée, la matrice \mathbf{K} semble pleine à priori. L'astuce consiste à choisir des fonctions de base e_i dont le support sera petit, i.e. chaque fonction e_i sera nulle partout sauf sur quelques mailles. Ainsi les termes $a_h(e_i, e_j)$ seront le plus souvent nuls et la matrice sera creuse. De plus, on ordonnera les e_i de sorte que \mathbf{K} soit à structure bande, avec une largeur de bande la plus faible possible. Il existe un moyen avantageux de stocker une matrice creuse qui s'appelle le stockage en ligne de ciel ou skyline.

Les difficultés majeures en pratique sont de trouver les e_i et de les manipuler pour les calculs d'intégrales nécessaires à la construction de \mathbf{K} . Indiquons d'ores et déjà que la plupart de ces difficultés seront levées grâce à trois idées principales qui seront détaillées au chapitre sur la formulation pratique des éléments finis :

Principe d'unisolvance : On s'attachera à trouver des degrés de liberté (ou ddl) tels que la donnée de ces degrés de liberté détermine de façon univoque toute fonction de V_h . Il pourra s'agir par exemple des valeurs de la fonction en quelques points. Déterminer une fonction reviendra alors à déterminer ses valeurs sur ces degrés de liberté.

Définition des e_i : On définira les fonctions de base par $e_i = 1$ sur le $i^{\text{ème}}$ degré de liberté, et $e_i = 0$ sur les autres degrés de liberté. La manipulation des e_i sera alors simplifiée, et les e_i auront par ailleurs un support réduit à quelques mailles.

Notion de « famille affine d'éléments » : Le maillage sera tel que toutes les mailles soient identiques à une transformation affine près. De ce fait, tous les calculs d'intégrales pourront se ramener à des calculs sur une seule maille de référence, par un simple changement de variable.

Notons que la matrice \mathbf{K} est appelée matrice de rigidité par analogie avec la mécanique des solides.

Si la forme bilinéaire a est coercive, alors \mathbf{K} est symétrique, définie positive donc inversible. On obtient donc l'existence et l'unicité de $\mathbf{q} = \mathbf{K}^{-1}\mathbf{F}$.

De nombreuses méthodes permettent de résoudre le système matriciel (d'inverser la matrice de rigidité) : Gauß, mais également toutes sortes de décomposition de la matrice de rigidité comme LU, LDLT, LLT (Cholesky). Lorsque \mathbf{K} est symétrique définie-positive, la méthode de Cholesky est la meilleure. Elle consiste à décomposer la matrice \mathbf{K} en le produit $\mathbf{L}^T\mathbf{L}$, où la matrice \mathbf{L} est une matrice triangulaire inférieure. Les factorisations LU et de Cholesky conservent la structure bande des matrices. De plus, pour une matrice bande d'ordre n et de demi largeur de bande p , la factorisation LU demande de l'ordre de $\mathcal{O}(np^2/3)$ opérations, alors que celle de Cholesky n'en demande que $\mathcal{O}(np^2/6)$.

10.4 Approximation conforme et lemme de Céa

Dans le cas d'une approximation conforme (dite aussi approximation interne), on se propose de construire un espace V_h , de dimension N_h , comme étant un sous-espace vectoriel de V .

10.4.1 Cas Lax-Milgram

On se place dans le cas d'une formulation relevant du théorème de Lax-Milgram.

V_h étant de dimension finie, c'est un fermé de V , et V étant un espace de Hilbert, V_h l'est donc aussi. D'où l'existence et l'unicité de u_h , en application du théorème de Lax-Milgram.

L'espace V_h sera en pratique construit à partir d'un maillage du domaine Ω , et l'indice h désignera la taille typique des mailles. Lorsque l'on construit des maillages de plus en plus fins, la suite de sous-espaces $(V_h)_h$ formera une approximation interne de V , i.e. pour tout élément φ de V , il existe une suite de $\varphi_h \in V_h$ telle que $\|\varphi - \varphi_h\| \rightarrow 0$ quand $h \rightarrow 0$.

Cette méthode d'approximation interne est appelée méthode de Galerkin (Galerkin ou Galerkin).

Signification de u_h . On a $a(u, v) = f(v)$, $\forall v \in V$, donc en particulier $a(u, v_h) = f(v_h)$, $\forall v_h \in V_h$, car $V_h \subset V$. Par ailleurs, $a(u_h, v_h) = f(v_h)$, $\forall v_h \in V_h$. Par différence, il vient $a(u - u_h, v_h) = 0$, $\forall v_h \in V_h$.

Dans le cas où $a(\cdot, \cdot)$ est symétrique, il s'agit d'un produit scalaire sur V . u_h peut alors être interprété comme la projection orthogonale de u sur V_h au sens de $a(\cdot, \cdot)$.

Lemme 1 — Lemme de C  a. La forme bilin  aire $a(\cdot, \cdot)$   tant continue, de constante de majoration M , et coercitive, de constante de minoration α , il est ais   d'obtenir la majoration de l'erreur appel  e lemme de C  a :

$$\|u - u_h\| \leq \frac{M}{\alpha} \|u - v_h\|, \forall v_h \in V_h \quad \text{c'est-  -dire} \quad \|u - u_h\| \leq \frac{M}{\alpha} d(u, V_h) \quad (10.7)$$

o   d est la distance induite par la norme $\|\cdot\|$.

Cette majoration donn  e par le lemme de C  a ram  ne l'  tude de l'erreur d'approximation $\|u - u_h\|$    celle de l'erreur d'interpolation $d(u, V_h)$.

10.4.2 Cas Babu  ska

On se place cette fois dans le cas d'un probl  me relevant du th  or  me de Babu  ska. Le probl  me est (v est dans M et non plus V) :

$$\text{Trouver } u \in V \text{ tel que } a(u, v) = f(v), \forall v \in M \quad (10.8)$$

L'approximation conforme du probl  me (10.8) est le probl  me approch   :

$$\text{Trouver } u_h \in V_h \text{ tel que } a(u_h, v_h) = f(v_h), \forall v_h \in M_h \quad (10.9)$$

o   $V_h \subset V$ et $M_h \subset M$. Attention : rien ne garantit *a priori* que la condition inf-sup discr  te sera v  rifi  e, m  me si la condition inf-sup est v  rifi  e. Ces propri  t  s doivent donc   tre d  montr  es pour chaque probl  me. V_h et M_h   tant de dimension finie, il faut n  cessairement que $\dim V_h = \dim M_h$. Le probl  me (10.9) est bien pos  , i.e. admet une unique solution, si :

$$\begin{cases} \exists \alpha_h > 0 \text{ tel que } \inf_{v_h \in V_h} \sup_{w_h \in M_h} \frac{a(v_h, w_h)}{\|v_h\|_{V_h} \|w_h\|_{M_h}} \geq \alpha_h \\ \forall w_h \in M_h, \quad (a(v_h, w_h) = 0) \Rightarrow (w_h = 0) \end{cases} \quad (10.10)$$

ce qui est   quivalent aux conditions :

$$\begin{cases} \exists \alpha_h > 0 \text{ tel que } \inf_{v_h \in V_h} \sup_{w_h \in M_h} \frac{a(v_h, w_h)}{\|v_h\|_{V_h} \|w_h\|_{M_h}} \geq \alpha_h \\ \dim V_h = \dim M_h \end{cases} \quad (10.11)$$

La premi  re relation est appel  e condition inf-sup discr  te.

On dispose de plus de la majoration d'erreur suivante :

$$\|u - u_h\| \leq \left(1 + \frac{M}{\alpha_h}\right) \inf_{v_h \in V_h} \{\|u - v_h\|\} \quad (10.12)$$

10.4.3 Cas Brezzi

On se place cette fois dans le cas d'un problème relevant du théorème de Brezzi (problème mixte). Le problème est de trouver (u, λ) dans $H \times M$ tels que :

$$\begin{cases} a(u, v) + b(v, \lambda) = \langle f, v \rangle & \forall v \in H, \\ b(u, \mu) = \langle g, \mu \rangle & \forall \mu \in M. \end{cases} \quad (10.13)$$

Dans ce cas, on va faire une hypothèse forte d'injection continue... Le problème discrétisé correspondant à l'équation (10.13) est bien posé, si le problème défini par la relation (10.13) vérifie les conditions de Brezzi, et si $(H_h \hookrightarrow H, M_h \hookrightarrow M)$ vérifie les conditions de Brezzi (de constantes α^* et β^*). De plus, on dispose alors de la majoration d'erreur :

$$\|u - u_h\|_H + \|\lambda - \lambda_h\|_M \leq C \left[\inf_{v_h \in H_h} \|u - v_h\| + \inf_{\mu_h \in M_h} \|\lambda - \mu_h\| \right] \quad (10.14)$$

où la constante C ne dépend que de $\|a\|$, α^* et β^* .

10.5 Approximations non conformes et lemmes de Strang

Lorsque les formes bilinéaires et linéaires définissant les problèmes doivent être approchées, on se trouve alors en présence d'une approximation non conforme. Comme mentionné, il existe deux types d'approximations non conformes selon que $V_h \subset V$ ou $V_h \not\subset V$.

10.5.1 Approximation interne

Le problème est donc :

$$\text{Trouver } u_h \in V_h \text{ tel que } a_h(u_h, v_h) = f_h(v_h), \forall v_h \in V_h \quad (10.15)$$

où $V_h \subset V$ est de dimension finie, et a_h et f_h sont des approximations de a et f définies sur $V_h \times V_h$ et V_h respectivement. Dans la pratique, a_h et f_h sont définies sur un espace plus grand que V_h , mais pas sur V .

Motivation. Considérons la forme :

$$a(u, v) = \int_{\Omega} (a \nabla u, \nabla v)(x) dx \quad (10.16)$$

Comme $a = a(x)$, le problème est : comment intégrer ? On prend des formules d'intégrations, donc on ne calcule pas $a(\cdot, \cdot)$, mais $a_h(\cdot, \cdot)$ et $f_h(\cdot)$.

Une majoration de l'erreur est donnée par le lemme 1 de Strang.

Lemme 2 — Lemme 1 de Strang. S'il existe une constante $\alpha^* > 0$ indépendante de h telle que :

$$\forall h, \quad a_h(v_h, v_h) \geq \alpha^* \|v_h\|^2, \quad \forall v_h \in V_h$$

alors il existe une constante C indépendante de h telle que :

$$\begin{aligned} \|u - u_h\| \leq C & \left[\inf_{v_h \in V_h} \left\{ \|u - v_h\| + \sup_{w_h \in V_h} \frac{|a(v_h, w_h) - a_h(v_h, w_h)|}{\|w_h\|} \right\} \dots \right. \\ & \left. \dots + \sup_{w_h \in V_h} \left\{ \frac{|f(w_h) - f_h(w_h)|}{\|w_h\|} \right\} \right] \end{aligned} \quad (10.17)$$

Le premier terme $\|u - v_h\|$ est appelé erreur d'approximation, le reste erreur de consistance.

10.5.2 Approximation externe

Le problème est encore :

$$\text{Trouver } u_h \in V_h \text{ tel que } a_h(u_h, v_h) = f_h(v_h), \quad \forall v_h \in V_h \quad (10.18)$$

mais $V_h \not\subset V$ est de dimension finie, et a_h et f_h sont des approximations de a et f définies sur $V_h \times V_h$ et V_h respectivement.

Motivation. Considérons le problème du Laplacien de Dirichlet. L'espace V est $H_0^1(\Omega)$.

On considère l'espace des éléments finis triangulaires engendrés par l'élément ayant une pression constante par élément et les nœuds au centre des arêtes du triangle. Nous notons cet espace V_h ; alors $V_h \not\subset H^1(\Omega)$. Nous prenons maintenant l'espace $V_{h,0}$ qui tient compte des conditions aux limites de Dirichlet. Si on prend :

$$a(u, v) = \int_{\Omega} \nabla u \cdot \nabla v \quad (10.19)$$

alors, dans Ω , il y a des sauts et l'intégrale n'a pas de sens. Par contre, si l'on considère :

$$a_h(u_h, v_h) = \sum_{K \in \mathcal{T}_h} \int_K \nabla u_h \cdot \nabla v_h \quad (10.20)$$

l'intégrale sur chaque élément a un sens.

Une majoration de l'erreur est donnée par le lemme 2 de Strang.

Lemme 3 — Lemme 2 de Strang. Si la norme $\|\cdot\|_h$ est définie sur $V_h + V$, et s'il existe deux constantes indépendantes de h , $\alpha^* > 0$ et $M^* > 0$ telles que :

$$\begin{cases} a_h(v, v) \geq \alpha^* \|v\|_h^2, & \forall v \in V_h \\ |a_h(u, v)| \leq M^* \|u\|_h \|v\|_h, & \forall u \in V_h + V, \forall v \in V_h \end{cases}$$

alors il existe une constante C indépendante de h telle que :

$$\|u - u_h\| \leq C \left[\inf_{v_h \in V_h} \{\|u - v_h\|\} + \sup_{w_h \in V_h} \left\{ \frac{|a(u, w_h) - f_h(w_h)|}{\|w_h\|_h} \right\} \right] \quad (10.21)$$

Le premier terme $\|u - v_h\|_h$ est appelé erreur d'approximation, le second erreur de consistance.

10.6 Convergence de la méthode des éléments finis en approximation conforme interne

Reprenons un cadre général, i.e. considérons Ω notre domaine de \mathbb{R}^n , et notre problème que nous résolvons de manière approchée par la méthode des éléments finis, lorsque $V_h \subset V$. Il s'agit maintenant de fournir une estimation de l'erreur $\|u - u_h\|_m$ dans H^m (souvent $m = 0, 1$ ou 2). La régularité de u et de u_h (et donc les valeurs possibles pour m) dépendent du problème continu considéré ainsi que du type d'éléments finis choisis pour sa résolution.

On rappelle que \mathcal{T}_h est le maillage de Ω considéré. On supposera de plus le domaine Ω polygonal, ce qui permet de le recouvrir exactement par le maillage. Si ce n'est pas le cas, il faut prendre en compte l'écart entre le domaine couvert par le maillage et le domaine réel.

10.6.1 Calcul de la majoration d'erreur

Les étapes du calcul de la majoration de l'erreur sont les suivantes :

- $\|u - u_h\|_m \leq C \|u - \pi_h u\|_m$: l'erreur d'approximation est bornée par l'erreur d'interpolation ;
- $\|u - \pi_h u\|_m^2 = \sum_{K \in \mathcal{T}_h} \|u - \pi_h u\|_{m,K}^2$: décomposition en majorations locales sur chaque élément ;
- $\|u - \pi_h u\|_{m,K} \leq C(K) \|\hat{u} - \hat{\pi} \hat{u}\|_{m,\hat{K}}$: passage à l'élément de référence ;
- $\|\hat{u} - \hat{\pi} \hat{u}\|_{m,\hat{K}} \leq \hat{C} |\hat{u}|_{k+1,\hat{K}}$: majoration sur l'élément de référence ;
- $\|u - \pi_h u\|_m \leq C' h^{k+1-m} |u|_{k+1}$: assemblage des majorations locales.

Majoration par l'erreur d'interpolation

En appliquant le lemme de Céa à $v_h = \pi_h u$, il vient :

$$\|u - u_h\| \leq \frac{M}{\alpha} \|u - \pi_h u\| \quad (10.22)$$

Décomposition en majorations locales sur chaque élément

$$\|u - \pi_h u\|_m^2 = \sum_{K \in \mathcal{T}_h} \|u - \pi_h u\|_{m,K}^2 = \sum_{K \in \mathcal{T}_h} \sum_{l=0}^m |u - \pi_h u|_{l,K}^2 \quad (10.23)$$

et le calcul est ramené à un calcul sur chaque élément, pour toutes les semi-normes $\|\cdot\|_{l,K}$ pour $l = 0, \dots, m$.

Passage à l'élément de référence

On se sert du théorème suivant :

Théorème 44 Soit K un élément quelconque de \mathcal{T}_h , et \hat{K} l'élément de référence. Soit F la transformation affine de \hat{K} vers K : $F(\hat{x}) = B\hat{x} + b$, avec B inversible. On a :

$$\forall v \in H^l(K), \quad |\hat{v}|_{l,K} \leq C(l, n) \|B\|_2^l |\det B|^{-1/2} |v|_{l,\hat{K}} \quad (10.24)$$

Ce résultat n'est rien d'autre qu'un résultat de changement de variable dans une intégrale.

On a même :

$$\forall v \in H^l(K), \quad |\hat{v}|_{l,K} \leq C(l, n) \|B^{-1}\|_2^l |\det B|^{1/2} |v|_{l,\hat{K}} \quad (10.25)$$

Et avec les données géométriques de l'élément :

$$\|B\| \leq \frac{h_K}{\hat{\rho}} \quad \text{et} \quad \|B^{-1}\| \leq \frac{\hat{h}}{\rho_K} \quad (10.26)$$

Majoration sur l'élément de référence

On a le théorème suivant :

Théorème 45 soient l et k deux entiers tels que $0 \leq l \leq k + 1$. Si $\hat{\pi} \in \mathcal{L}(H^{k+1}(\hat{K}), J^l(\hat{K}))$ laisse $P_k(\hat{K})$ invariant, i.e. si $\forall \hat{p} \in P_k(\hat{K}), \hat{\pi} \hat{p} = \hat{p}$ alors :

$$\exists C(\hat{K}, \hat{\pi}), \quad \forall \hat{v} \in H^{k+1}(\hat{K}), \quad |\hat{v} - \hat{\pi} \hat{v}|_{j,\hat{K}} \leq C |\hat{v}|_{k+1,\hat{K}} \quad (10.27)$$

Majoration sur un élément quelconque

De ce qui précède, il vient :

$$|v - \pi_K v|_{l,K} \leq \hat{C}(\hat{\pi}, \hat{K}, l, k, n) \frac{h_K^{k+1}}{\rho_K^l} |v|_{k+1,K} \quad (10.28)$$

où il faut remarquer que la constante \hat{C} est indépendante de l'élément K .

Assemblage des majorations locales

Il ne reste plus qu'à assembler les résultats précédents pour tous les éléments K du maillage \mathcal{T}_h et pour toutes les valeurs des semi-normes, i.e. pour $l = 0, \dots, m$.

Par majoration et sommation sur les éléments, on obtient :

$$\|v - \pi_h v\|_m \leq C(\mathcal{T}_h, m, k, n) h^{k+1-m} |v|_{k+1} \quad (10.29)$$

10.6.2 Majoration de l'erreur

On obtient au final, le résultat classique de majoration d'erreur :

$$\|u - u_h\|_m \leq C h^{k+1-m} |u|_{k+1} \quad (10.30)$$

Quelques remarques :

- On rappelle que la formule précédente a été obtenue pour un domaine Ω polygonal. Si ce n'est pas le cas, elle n'est plus valable. Les éléments linéaires conduisent à une erreur en $O(h)$. Utiliser des éléments plus raffinés (de degré 2 au lieu d'éléments linéaires) permet, même si la géométrie n'est également ici pas décrite de manière exacte, d'obtenir une erreur asymptotique en $O(h^2)$. Quelque soit le degré de l'approximation, si le domaine Ω n'est pas représenté de manière exacte, alors cela revient à une modification des conditions aux limites.
- Les calculs, et plus précisément les intégrations, ont été réalisées sans erreur. Si celles-ci sont faites de manière numérique, il convient d'introduire encore un terme correctif supplémentaire, appelé erreur de consistance due au remplacement des formes par leur approximation ($a(\cdot, \cdot)$ par $a_h(\cdot, \cdot)$, $f(\cdot)$ par $f_h(\cdot)$...). Toutefois, cette erreur supplémentaire peut être estimée en $O(h^k)$ selon la précision du schéma d'intégration utilisé.
- Le résultat de majoration d'erreur est souvent utilisé dans le cas $m = 1$. Comme l'espace des polynômes $P_k(\hat{K}) \subset H^1(\hat{K})$, alors si $\hat{\pi}$ est bien défini sur $H^{k+1}(\hat{K})$, on a :

$$\text{si } u \in H^{k+1}(\Omega), \quad \|u - u_h\|_1 \leq C h^k |u|_{k+1} \quad (10.31)$$

Chapitre 11

Choix d'un Modèle

Résumé — L'intégralité de la méthode des éléments finis a été présentée au chapitre précédent.

Avant de présenter plus en détail la formulation d'éléments finis, nous tenions à ajouter un court chapitre comme mise en garde en modélisation. Ce document étant essentiellement destiné à un public d'ingénieurs mécaniciens, nous allons illustrer notre propos en mécanique.

Histoire

La mécanique est sans doute aussi vieille que l'homme. Aussi bien pour des aspects pratiques (faire des outils pour chasser...), que pour des aspects plus philosophiques et spirituels visant notamment à expliquer les mouvements des astres...

Archimède, outre ces travaux en mathématiques, pourrait sans conteste être considéré comme le saint patron de la mécanique. Il est tout au moins indubitablement le père de la mécanique statique.

Il s'intéressa aussi bien aux aspects « théoriques » portant sur le principe du levier et la recherche de centre de gravité dans *De l'équilibre des figures planes*, sur le principe d'Archimède pour les corps plongés dans un liquide dans *Des corps flottants* ; qu'aux aspects « pratiques » au travers de nombreuses inventions : machines de traction (où il démontre qu'à l'aide de poulies, de palans et de leviers, l'homme peut soulever bien plus que son poids), machines de guerre (principe de la meurtrière, catapultes, bras mécaniques utilisés dans le combat naval), l'odomètre (appareil à mesurer les distances), la vis sans fin et la vis d'Archimède, le principe de la roue dentée...

Le siège de Syracuse, les miroirs d'Archimède et sa mort ne font qu'ajouter à sa légende.

Bien qu'Aristote posa le premier (avant Archimède) les bases d'une véritable théorie mécanique (alors encore très imparfaite), les fondements de la mécanique, en tant que science et au sens moderne du terme, sont posées par Galilée en 1632 dans *les Dialogues* et en 1638 dans *les Discours*.

La mécanique n'est alors pas dissociée des arts mécaniques, i.e. des techniques de construction des machines. La distinction entre la mécanique en tant que science et la mécanique en tant que technique ne se fera qu'au XIX^e siècle.



Aristote

Archimède

Galilée



Newton

Halley

En 1677, Newton reprend ses travaux sur la mécanique, i.e. sur la gravitation et ses effets sur les orbites des planètes. En novembre 1684, il fit parvenir à Halley un petit traité de neuf pages intitulé *De motu corporum in gyrum* (Mouvement des corps en rotation) montrant la loi en carré inverse, la force centripète, ainsi que les prémices des lois du mouvement.

Son ouvrage *Philosophiæ Naturalis Principia Mathematica* (aujourd'hui connu sous le nom de *Principia* ou *Principia Mathematica*), écrit en 1686 et publié le 5 juillet 1687, est considéré comme une œuvre majeure dans l'histoire de la science. Il y décrit la gravitation universelle, formule les trois lois du mouvement et jette les bases de la mécanique classique ou mécanique newtonienne. Ces trois lois universelles du mouvement resteront inchangées, sans aucune amélioration, durant plus de deux siècles, jusqu'à l'arrivée des mécaniques relativiste et quantique.

Rappelons que ces trois lois sont : 1) le principe d'inertie ; 2) le principe fondamental de la dynamique ; et 3) le principe des actions réciproques.

La mécanique classique sera ensuite complétée et mathématisée pour devenir la mécanique analytique. Cette dernière, initiée dès le XVIII^e siècle, regroupe, en plus de la mécanique newtonienne, les mécaniques de Hamilton et de Lagrange. Toutes ces mécaniques ont en commun l'application initiale d'un principe variationnel, et avec lui l'utilisation du calcul variationnel... ce qui fait le lien avec le présent document.

11.1 La mécanique, un problème à plusieurs champs

Dans ce paragraphe, nous reprenons les formulations présentées au paragraphe 8.3.2, mais en essayant de les éclairer par un discours plus pragmatique. C'est pourquoi nous changerons quelque peu les notations, afin de retomber sur des choses peut-être plus familières pour des ingénieurs mécaniciens.

Bien qu'étant un sujet ancien, la mécanique n'en est pas pour autant un problème simple. En mécanique, les champs inconnus sont les déplacements, déformations et contraintes (et d'autres si besoin, comme la température...). De plus, on dispose de relations entre ces champs.

Il est possible de n'exprimer le problème qu'à l'aide des seuls déplacements. Les déformations sont alors calculées à partir des déplacements (par combinaison linéaire des dérivées, obtenues de manière approchée), puis les contraintes sont obtenues à partir des déformations par la loi de comportement (linéaire ou non, isotrope ou non...).

Il est également tout à fait possible d'exprimer le problème en utilisant les déplacements et les contraintes comme inconnues. Cela permet de prendre en compte certaines continuités des contraintes en certains lieux (interface entre deux matériaux par exemple, voir paragraphe 12.3 pour une synthèse) de la structure, et d'imposer les conditions aux limites de nullité des contraintes aux lieux le nécessitant (par exemple bord libre).

Par contre, le champ de contraintes peut également être « trop » continu en certains endroits selon le type de problème. Cette continuité étant liée à la constitution de l'élément fini choisi, on peut ne pas disposer d'éléments capables de modéliser correctement ce phénomène...

Notons qu'il serait tout aussi possible d'utiliser une formulation ayant les trois champs comme inconnues... Enfin bref, tout est possible, mais il faut veiller à ce que la modélisation de chaque champ soit cohérente avec le phénomène physique à modéliser. En d'autres termes, on ne choisit pas un élément au hasard, juste parce qu'il a le bon nombre d'inconnues, il est évidemment nécessaire de se demander comment ces inconnues sont interpolées, ce que cela implique sur la régularité des solutions et donc si cela est compatible avec le phénomène physique que l'on souhaite simuler. Un autre exemple typique est le cas des éléments « déplacement-pression » utilisés par exemple pour la modélisation des solides incompressibles (nous avons évoqué le problème précédemment).

Ce document s'adressant à des personnes connaissant déjà la méthode des éléments finis, nous allons présenter très brièvement, sous forme d'aparte, la discrétisation multi-champs, de manière un peu découplée du reste du document, juste pour fixer les idées.

Si l'on note u le champ de déplacements, celui-ci peut être approché à partir des déplacements nodaux (sous forme de vecteur) \mathbf{q} par l'intermédiaire de fonctions de formes rangées dans la matrice \mathbf{N}_u . Compte tenu des notations utilisées, les interpolations des différents champs se font, sur chaque élément, de la manière suivante :

$$u = \mathbf{N}_u \mathbf{q} \quad (11.1)$$

De la même manière, si les déformations ε sont choisies comme champ inconnu, elles seront approchées à partir des déformations nodales (sous forme de vecteur avec la convention de l'ingénieur) γ par l'intermédiaire de fonctions de formes rangées dans la matrice \mathbf{N}_ε .

$$\varepsilon = \mathbf{N}_\varepsilon \gamma \quad (11.2)$$

Cela vaut également pour les contraintes σ qui, si elles sont choisies comme champ inconnu, seront approchées à partir des contraintes nodales (sous forme de vecteur avec la convention de l'ingénieur) τ par l'intermédiaire de fonctions de formes rangées dans la matrice \mathbf{N}_σ .

$$\sigma = \mathbf{N}_\sigma \tau \quad (11.3)$$

Notons que dans la pratique, rien n'empêche de prendre les mêmes fonctions de forme pour les différents champs. De manière analogue, le vecteur λ des multiplicateurs de Lagrange sera interpolé de la façon suivante :

$$\lambda = \mathbf{N}_\lambda \mathbf{L} \quad (11.4)$$

On aura donc :

$$\begin{aligned} \varepsilon &= \mathbf{N}_\varepsilon \gamma && \text{approximation en déformations} \\ &= \mathcal{L} u && \text{relation déformations déplacements} \\ &= \mathcal{L} \mathbf{N}_u \mathbf{q} && \text{approximation en déplacements} \\ &= \mathbf{S} \sigma && \text{loi de Hooke inverse} \\ &= \mathbf{S} \mathbf{N}_\sigma \tau && \text{approximation en contraintes} \end{aligned} \quad (11.5)$$

et, de façon inverse :

$$\begin{aligned} \sigma &= \mathbf{N}_\sigma \tau && \text{approximation en contraintes} \\ &= \mathbf{H} \varepsilon && \text{loi de Hooke généralisée} \\ &= \mathbf{H} \mathbf{N}_\varepsilon \gamma && \text{approximation en déformations} \\ &= \mathbf{H} \mathcal{L} u && \text{loi de Hooke en élasticité linéaire} \\ &= \mathbf{H} \mathcal{L} \mathbf{N}_u \mathbf{q} && \text{approximation en déplacements} \end{aligned} \quad (11.6)$$

avec, en petites déformations :

$$\mathcal{L} = \begin{bmatrix} \frac{\partial}{\partial x} & 0 \\ 0 & \frac{\partial}{\partial y} \\ \frac{\partial}{\partial y} & \frac{\partial}{\partial x} \end{bmatrix} \text{ en 2D, et } \mathcal{L} = \begin{bmatrix} \frac{\partial}{\partial x} & 0 & 0 \\ 0 & \frac{\partial}{\partial y} & 0 \\ 0 & 0 & \frac{\partial}{\partial z} \\ 0 & \frac{\partial}{\partial z} & \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} & 0 & \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} & \frac{\partial}{\partial x} & 0 \end{bmatrix} \text{ en 3D} \quad (11.7)$$

La méthode des éléments finis *classique*, ou *en déplacements*, n'utilise que le champ de déplacements comme variable. Elle est basée sur le principe du travail virtuel :

$$\delta \Pi_{TV} = \int_{\Omega} \delta \varepsilon(u) \mathbf{H} \varepsilon(u) - \delta u \bar{f}_{\Omega} \, d\Omega - \int_{\Gamma_{\sigma}} \delta u \bar{T} \, d\Gamma \quad (11.8)$$

avec \bar{f}_Ω les forces imposées dans Ω et \bar{T} les forces imposées sur Γ_σ , où Γ_σ et Γ_u forment une partition de $\Gamma = \partial\Omega$. Ce principe est obtenu comme variation de la fonctionnelle de l'énergie potentielle totale exprimée en déplacements :

$$\Pi_{TV} = \int_{\Omega} \frac{1}{2} {}^T \varepsilon(\mathbf{u}) \mathbf{H} \varepsilon(\mathbf{u}) - {}^T \mathbf{q} \bar{\mathbf{f}} \, d\Omega - \int_{\Gamma_\sigma} {}^T \mathbf{q} \bar{\mathbf{T}} \, d\Gamma \quad (11.9)$$

où les conditions subsidiaires sur les déplacements sont prises en compte directement par l'espace dans lequel les déplacements sont recherchés.

La fonctionnelle d'Hellinger-Reissner est sans doute la plus connue des fonctionnelles mixtes. Elle utilise les champs de déplacements et de contraintes comme variables indépendantes. Son expression est la suivante, \bar{U} désignant les déplacements imposés :

$$\Pi_{HR} = \int_{\Omega} -\frac{1}{2} \sigma S \sigma - \sigma_{ij,j} u + \bar{f}_\Omega u \, d\Omega - \int_{\Gamma_\sigma} (\bar{T} - T) u \, d\Gamma - \int_{\Gamma_u} \bar{U} T \, d\Gamma \quad (11.10)$$

bien que l'on puisse la trouver sous une autre forme, obtenue par intégration par parties de celle-ci. Elle n'a à satisfaire à aucune condition subsidiaire. La stationnarité de cette fonctionnelle conduit aux équations d'équilibre, à la loi de comportement, et aux conditions aux limites en forces et déplacements.

Cette fonctionnelle conduit à un élément ayant les champs de déplacements et de contraintes comme inconnues nodales. Il s'en suit que toutes les composantes de ces champs sont continues. Elle conduit à la résolution d'un système du type mixte (Brezzi) :

$$\begin{bmatrix} -\mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{0} \end{bmatrix} \begin{pmatrix} \tau \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{F} \end{pmatrix} \quad (11.11)$$

où la matrice de rigidité est symétrique et non définie-positive, et :

$$\mathbf{A} = + \int_{\Omega} {}^T \mathbf{N}_\sigma \mathbf{S} \mathbf{N}_\sigma \, d\Omega \quad \mathbf{B} = + \int_{\Omega} {}^T \mathbf{N}_\sigma \mathcal{L} \mathbf{N}_u \, d\Omega \quad \mathbf{F} = + \int_{\Omega} {}^T \mathbf{N}_u \bar{\mathbf{f}} \, d\Omega \quad (11.12)$$

L'un des moyens pour obtenir une fonctionnelle hybride est d'introduire une condition sur une partie du contour par l'intermédiaire de multiplicateurs de Lagrange comme présenté un peu plus loin.

Tout comme la fonctionnelle d'Hellinger-Reissner est associée à la méthode mixte, celle de Pian et Tong est indissociable de l'adjectif hybride, même si, en toute rigueur, elle est une méthode mixte (deux champs) hybride (différents domaines d'interpolation) :

$$\Pi_{PT} = -\frac{1}{2} \int_{\Omega} \sigma S \sigma \, d\Omega + \int_{\Gamma} T u \, d\Gamma - \int_{\Gamma_\sigma} \bar{T} u \, d\Gamma \quad (11.13)$$

Sa variation est :

$$\delta \Pi_{PT} = - \int_{\Omega} \sigma S \sigma \, d\Omega + \int_{\Gamma} \delta T u + \delta u T \, d\Gamma - \int_{\Gamma_\sigma} \delta u \bar{T} \, d\Gamma_\sigma \quad (11.14)$$

dont l'avantage principal est de ne pas comporter de dérivée.

Cette variation peut être transformée en :

$$\int_{\Omega} \delta \sigma (\varepsilon - S \sigma) \, d\Omega + \int_{\Gamma} \delta u T \, d\Gamma - \int_{\Gamma_\sigma} \delta u \bar{T} \, d\Gamma$$

La stationnarité de cette fonctionnelle conduit à la loi de comportement, et aux conditions aux limites en contraintes. Les conditions aux limites en déplacements seront imposées par l'espace dans lequel on cherche u .

Le système à résoudre est de la forme mixte donné précédemment à la relation (11.11) avec :

$$\mathbf{A} = + \int_{\Omega} {}^T \mathbf{N}_{\sigma} \mathbf{S} \mathbf{N}_{\sigma} d\Omega \quad \mathbf{B} = + \int_{\Gamma} {}^T \mathbf{N}_{\sigma} {}^T \mathbf{M} \mathbf{N}_u d\Gamma \quad \mathbf{F} = + \int_{\Gamma_{\sigma}} {}^T \mathbf{N}_u \bar{\mathbf{T}} d\Gamma \quad (11.15)$$

\mathbf{M} étant la matrice des cosinus directeurs donnée par les relations :

$$\mathbf{M} = \begin{bmatrix} n_1 & 0 & n_2 \\ 0 & n_2 & n_1 \end{bmatrix} \text{ dans le plan, et } \mathbf{M} = \begin{bmatrix} n_1 & 0 & 0 & 0 & n_3 & n_2 \\ 0 & n_2 & 0 & n_3 & 0 & n_1 \\ 0 & 0 & n_3 & n_2 & n_1 & 0 \end{bmatrix} \text{ dans l'espace.} \quad (11.16)$$

Toutefois, le champ de contraintes n'étant défini que dans Ω , il est possible d'effectuer une condensation statique du champ de contraintes, i.e. de transformer le système (11.11) en écrivant :

$$\tau = \mathbf{A}^{-1} \mathbf{B} \mathbf{q} \quad (11.17)$$

et le champ de déplacements est solution d'un système de type classique $\mathbf{K} \mathbf{q} = \mathbf{F}$, où la matrice de rigidité \mathbf{K} est remplacée par la matrice de rigidité équivalente suivante :

$$\mathbf{K}_{eq} = {}^T \mathbf{B} \mathbf{A}^{-1} \mathbf{B} \quad (11.18)$$

Sur le plan du déroulement du calcul, on commence par résoudre une « forme classique » mais en utilisant la matrice de rigidité équivalente \mathbf{K}_{eq} , puis le calcul des contraintes dans chaque élément se fait par la relation (11.17).

Les multiplicateurs de Lagrange peuvent être utilisés pour introduire des conditions supplémentaires directement à la fonctionnelle. Ces conditions sont généralement imposées sur Γ_X , tout ou partie de $\Gamma = \partial\Omega$. Pour ce faire, il suffit d'ajouter à la fonctionnelle Π du problème le terme :

$$\pm \int_{\Gamma_X} {}^T \lambda_{condition} \quad (11.19)$$

où les λ_i sont les multiplicateurs de Lagrange.

Les multiplicateurs de Lagrange sont généralement utilisés pour :

- introduire des variables physiques additionnelles comme inconnues ;
- obtenir des conditions de dérivabilité ou des conditions aux limites moins sévères sur les inconnues.

Notons qu'un autre de leurs avantages est de permettre l'utilisation d'une formulation différente dans chaque élément, en assurant les continuités nécessaires aux interfaces (voir synthèse de ce problème d'interface au paragraphe 12.3). Bien sûr, apparaissant comme inconnues, les multiplicateurs de Lagrange viennent grossir la taille du système à résoudre.

11.2 Plusieurs modélisations d'un même problème

La qualité de l'approximation des différents champs est non seulement liée au nombre de champs inconnus choisis, mais également au choix de l'élément. Ce dernier peut traduire une « simplification » du problème physique : il s'agit du problème du choix du type de modélisation, en une, deux ou trois dimensions. Selon le type d'information recherché, une modélisation « plus simple » qu'une autre peut fournir les résultats escomptés.

Dans ce paragraphe, nous nous proposons d'étudier un exemple où seul le champ de déplacement apparaît comme inconnue nodale. Toutefois, si ce n'est plus le nombre de champ qui peut modifier la formulation, nous allons voir que la mécanique offre moult théories qui toutes ont leurs avantages et inconvénients, mais qui surtout ne sont pas toutes équivalentes.

La paternité de la théorie des poutres est attribuée à Galilée, mais des études récentes indiquent que Léonard de Vinci l'aurait précédé. De Vinci avait supposé que la déformation variait de manière linéaire en s'éloignant de la surface neutre, le coefficient de proportionnalité étant la courbure, mais il ne put finaliser ses calculs car il n'avait pas imaginé la loi de Hooke^a. De son côté, Galilée était parti sur une hypothèse incorrecte (il supposait que la contrainte était répartie uniformément en flexion), et c'est Antoine Parent qui obtint la distribution correcte.

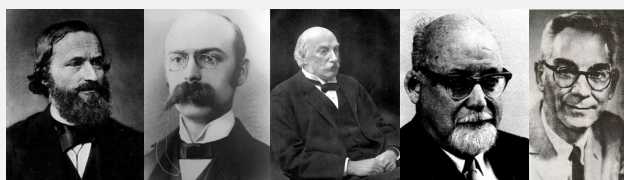
Ce sont Leonhard Euler et Jacques Bernoulli qui é mirent la première théorie utile vers 1750, tandis que Daniel Bernoulli, le neveu du précédent (et le fils de Jean Bernoulli), écrivit l'équation différentielle pour l'analyse vibratoire. À cette époque, le génie mécanique n'était pas considéré comme une science, et l'on ne considérait pas que les travaux d'une académie des mathématiques puissent avoir des applications pratiques... On continua donc à bâtir les ponts et les bâtiments de manière empirique. Ce n'est qu'au XIX^e siècle, avec la Tour Eiffel et les grandes roues, que l'on démontra la validité de la théorie à grande échelle.



Euler J. Bernoulli D. Bernoulli Timoshenko

La théorie des poutres est une simplification unidimensionnelle. On distingue :

- la théorie d'Euler-Bernoulli, qui néglige l'influence du cisaillement ;
- la théorie de Timoshenko qui prend en compte l'effet du cisaillement.



Kirchhoff Love Rayleigh Reissner Mindlin

En 1888, Love utilise les hypothèses de Gustav Kirchhoff, elles-mêmes inspirées des hypothèses d'Euler-Bernoulli pour les poutres, pour fonder une théorie des plaques minces.

La théorie des plaques épaisses a été consolidée par Mindlin à partir des travaux de Rayleigh (1877), Timoshenko (1921), Reissner (1945) et Uflyand (1948).

La théorie des plaques minces, ou théorie de Love-Kirchhoff, suppose que :

- le plan moyen (équivalent de la courbe moyenne des poutres) est initialement plan ;
- le feuillet moyen (équivalent de la fibre neutre des poutres) ne subit pas de déformation dans son plan ; on ne considère que le déplacement transversal w des points du feuillet moyen ;
- modèle de Kirchhoff : les sections normales au feuillet moyen restent normales lors de la déformation ; en conséquence, on peut négliger le cisaillement ;
- l'épaisseur est faible ; en conséquence, les contraintes dans le sens de l'épaisseur sont supposées nulles ;
- on reste en petites déformations.

Notons que cette théorie permet de déterminer la propagation des ondes dans les plaques, ainsi que l'étude des ondes stationnaires et des modes vibratoires.

Dans la théorie des plaques épaisses, ou théorie de Reissner et Mindlin, la fibre normale reste toujours rectiligne, mais n'est plus nécessairement perpendiculaire au plan moyen. On ne peut donc plus négliger le cisaillement.

^a. Robert Hooke, qui désirait obtenir une théorie des ressorts en soumettant ces derniers à des forces croissantes successives, énonça en 1678, à partir d'expériences datant de 1675, la loi de comportement suivante : « ut tensio sic vis » ce qui signifie « telle extension, telle force », ou bien en termes modernes « l'allongement est proportionnel à la force ».

La figure 11.1 propose trois modélisations d'un même problème. Il s'agit d'une poutre encastree à une extrémité et soumise à une force décentrée à l'autre. Vaut-il mieux modéliser l'intégralité du volume de la poutre, la traiter comme une plaque, ou peut-on se contenter d'un modèle de poutre ?

L'étude de la contrainte axiale σ_{xx} est donnée à la figure 11.2. Si les cartographies présentent bien la même répartition, seul le modèle 3D permet de mettre en évidence la concentration de contrainte due à la force ponctuelle.

Une comparaison plus fine des trois modélisations concernant cette même contrainte normale

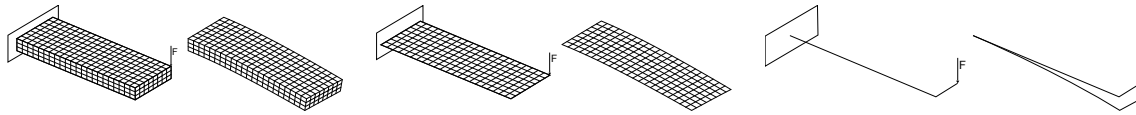


FIGURE 11.1: Trois modélisations d'un même problème.

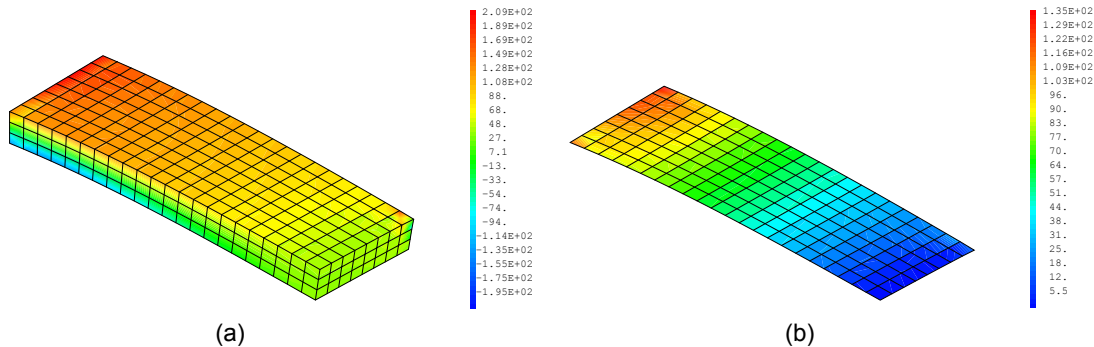


FIGURE 11.2: Contrainte axiale sur la peau supérieure : (a) modèle 3D, (b) modèle 2D.

le long de trois lignes est reportée à la figure 11.3. Les différences de résultats proviennent des

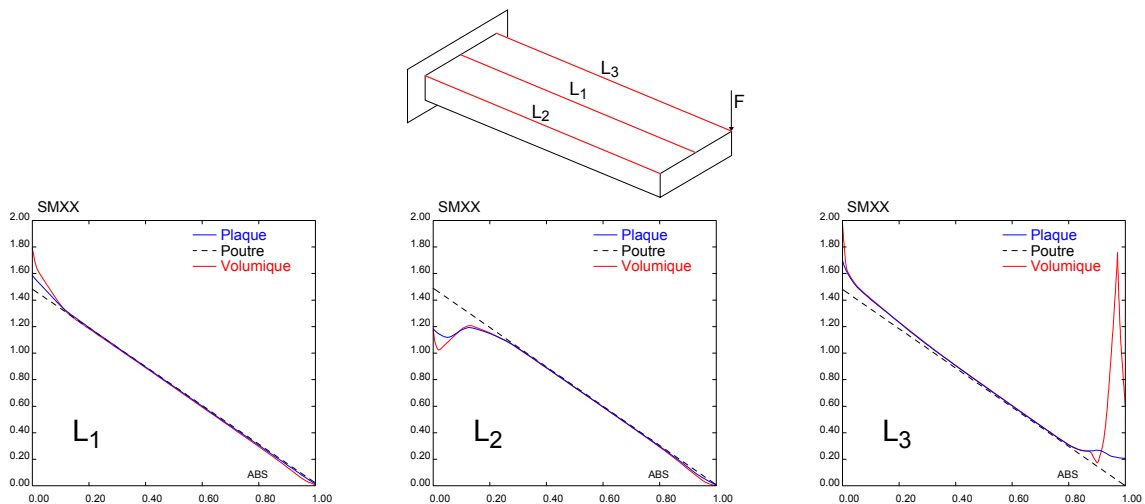


FIGURE 11.3: Contrainte axiale sur la peau supérieure en différentes sections

hypothèses cinématiques et des composantes accessibles dans les différentes théories utilisées. On y voit que loin des extrémités, les trois solutions concordent parfaitement, conformément au principe de Saint-Venant.¹ Les différences proviennent des effets de bord, i.e. des variations locales des contraintes et déformations au voisinage des conditions aux limites.

Les hypothèses cinématiques sont :

- la **théorie des poutres** (Figure 11.4a) suppose que chaque section droite suit un mouvement de solide rigide. Les sections ne peuvent donc pas se déformer, elles peuvent uniquement se translater et tourner dans l'espace (sans forcément rester perpendiculaires à la ligne moyenne, car dans ce cas on considère une théorie avec cisaillement transverse) ;
- la **théorie des plaques** (Figure 11.4b), moins restrictive, suppose que chaque segment perpendiculaire au plan moyen de la plaque suit un mouvement de solide rigide (là encore, sans

1. Le **principe de Saint-Venant** précise que le comportement en un point quelconque de la poutre, pourvu que ce point soit suffisamment éloigné des zones d'applications des forces et des liaisons, est indépendant de la façon dont sont appliquées les forces et de la façon dont sont physiquement réalisées les liaisons ; le comportement dépend alors uniquement du torseur des forces internes en ce point.

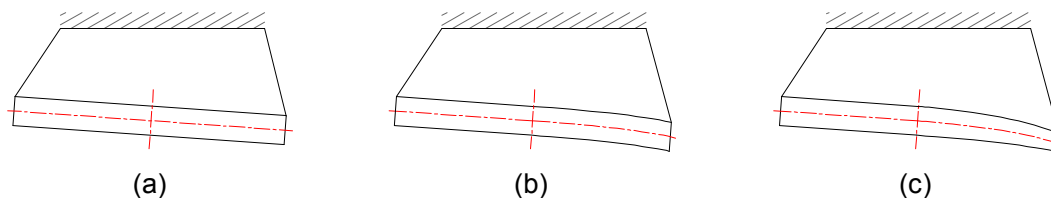


FIGURE 11.4: Mouvements d'une section droite : (a) théorie des poutres ; (b) théorie des plaques ou coques ; (c) théorie volumique

forcément rester perpendiculaire au plan moyen). Elle permet donc de modéliser certaines formes de déformations des sections, planes (flexion et cisaillement dans le plan de la section, traction dans le sens de la largeur) ou hors plan (certains types de gauchissement). Cependant, les segments ne pouvant pas changer de longueur, elle ne permet pas de modéliser l'écrasement de l'épaisseur ;

- enfin, la **théorie 3D** (Figure 11.4c) ne comporte aucune de ces restrictions et peut modéliser n'importe quelle forme de gauchissement ou d'écrasement, à condition que le maillage employé soit suffisamment fin.

Ici, la comparaison des résultats « plaque » et « 3D » montre que la contribution de l'écrasement de la section à la flèche semble négligeable (0,7% d'écart). Ce n'est pas forcément vrai car l'écrasement est un phénomène localisé sous la charge et lié au contact, qui a été modélisé très grossièrement... Cette très forte concentration de contraintes sous la charge est caractéristique d'une singularité. La valeur obtenue est peu fiable. On peut même dire que physiquement, une force ponctuelle n'existe pas. Elle est forcément distribuée sur une petite surface...

De même, la comparaison des résultats « poutre » et « plaque » montre que la flexion et le cisaillement de la section contribuent davantage à la flèche que l'écrasement (6,4% d'écart) : la largeur de la pièce est visiblement suffisante pour que les déformations des sections droites provoquent un déplacement vertical non négligeable sous la charge.

Il est essentiel de noter que toutes les théories ne permettent pas d'accéder à toutes les composantes du champ des contraintes : de manière générale, seuls les efforts qui travaillent dans les déplacements permis par la théorie sont accessibles (les théories sont ainsi faites afin de pouvoir respecter le premier principe de la thermodynamique dont l'écriture nécessite de calculer le travail de tous les efforts permis par la théorie). Ainsi :

- la **théorie des poutres** ne permet pas de calculer les contraintes dans le plan transversal (σ_{yy} , σ_{zz} et σ_{yz}) du fait de l'indéformabilité des sections droites ;
- la **théorie des plaques** ne permet pas de calculer la contrainte normale au plan de la plaque σ_{zz} , du fait de l'indéformabilité des segments perpendiculaires à ce plan ;
- enfin, la **théorie 3D** permet de calculer les six composantes du tenseur des contraintes.

À ces limitations théoriques peuvent s'ajouter des limitations techniques propres à chaque logiciel, susceptibles de rendre d'autres grandeurs physiques inaccessibles. Avant d'effectuer une modélisation par éléments finis, il est donc indispensable de s'assurer que le logiciel utilisé et son cadre théorique permettent bien d'accéder au résultat voulu (en plus d'être pertinents vis-à-vis de la géométrie du produit, de son environnement et du comportement attendu).

De plus, si certains logiciels peuvent avoir des « limitations techniques », ils peuvent également disposer de méthodes de post-traitement susceptibles d'améliorer ou de permettre d'accéder à certaines composantes (sous certaines hypothèses). Cela aussi doit être pris en compte.

11.3 Exemple : retour sur le calcul de poutre du paragraphe 11.1 avec CAST3M

Nous allons reprendre pour partie l'exemple de la poutre encastree traitée au paragraphe précédent et présenter le listing CAST3M correspondant.

11.3.1 Modélisation 2D

Nous commençons par définir les données du problème : longueur, largeur, épaisseur, nombre d'éléments selon chacune de ces directions, et force appliquée :

```
1 * DONNEES
2 * geometrie
3 long1=22.0;
4 larg1=8.;
5 ep1=4;
6 * maillage
7 nlong1=22;
8 nlarg1=8;
9 * effort
10 Forc1=-41.;
```

Puis nous définissons la dimension du problème (modèle tridimensionnel), et le type de découpage :

```
11 OPTION DIME 3 ELEM QUA4;
```

Nous définissons les points k_i (dénommés ainsi pour rappel de la syntaxe ANSYS des keypoints k, i, \dots), puis les lignes L_i , et la surface S_1 .

```
12 k1 = 0. 0. 0.;
13 k2 = long1 0. 0.;
14 k3 = long1 larg1 0.;
15 k4 = 0. larg1 0.;
16 *
17 L1 = DROI nlong1 k1 k2;
18 L2 = DROI nlarg1 k2 k3;
19 L3 = DROI nlong1 k3 k4;
20 L4 = DROI nlarg1 k4 k1;
21 *
22 S1 = DALLER L1 L2 L3 L4;
```

Le modèle est un modèle de mécanique élastique isotrope utilisant l'élément de coque coq4 pour le maillage S_1 :

```
1 Model1 = MODL S1 MECANIQUE ELASTIQUE ISOTROPE COQ4;
```

Enfin on résout le problème après avoir fourni les données matérielles et les conditions aux limites :

```
2 Mater1 = MATER Model1 YOUNG 70000.0 NU 0.33 RHO 2700.0;
3 Car1 = CARAC Model1 EPAI ep1;
4 Mater1 = Mater1 ET Car1;
5 MR1 = RIGID Model1 Mater1;
6 CL1 = BLOQ DEPL L4;
7 CL2 = BLOQ ROTA L4;
8 FOR1 = FORC(0. 0. Forc1) k3;
9 MTOT1 = MR1 ET CL1 ET CL2;
10 Depl1 = RESO MTOT1 FOR1 ;
```

On peut post-traiter les résultats et les afficher.

Les composantes de $S_i g_1$ dans le repère LOCAL sont :

- les efforts normaux : N11, N22 ;
- l'effort de cisaillement plan : N12 ;
- les moments de flexion : M11, M22 ;
- le moment de cisaillement plan : M12 ;
- les efforts tranchants V1, V2.

- Les composantes de Eps_1 dans le repère LOCAL sont :
- les élongations normales dans le plan : EPSS, EPTT ;
 - les cisssions dans le plan et transverses : GAST, GASN, GATN ;
 - les courbures : RTSS, RTTT, RTST.

```

11 * DEPLACEMENTS
12 UZ1 = EXCO 'UZ' dep1;
13 *TRAC UZ1 S1;
14 *
15 * DEFORMEE
16 def0 = DEFO S1 Dep1 0.0 BLEU;
17 def1 = DEFO S1 Dep1;
18 *TRAC (def0 ET def1);
19 *
20 * CONTRAINTES:
21 Sig1 = SIGM Model1 Mater1 Dep1;
22 Siigg1 = EXCO M11 Sig1;
23 TRAC Siigg1 Model1 def1;
24 *
25 * DEFORMATIONS
26 Eps1 = EPSI Model1 Mater1 Dep1;
27 Eppss1 = EXCO RTSS Eps1;
28 *TRAC Eppss1 Model1;
29 *
30 fin;

```

11.3.2 Modèle 3D

Nous allons maintenant construire le modèle tridimensionnel. Comme nous voulons travailler à l'économie, nous repartons du fichier précédent que nous adaptons.

```

1 * DONNEES
2 * geometrie
3 long1=22.0;
4 larg1=8.;
5 ep1=2;
6 * maillage
7 nlong1=22;
8 nlarg1=8;
9 nep1=3;
10 * effort
11 Forc1=-41.;

```

Cette fois, nous nous servons de l'élément volumique CUB8 au lieu de l'élément surfacique QUA4.

```

1 OPTION DIME 3 ELEM CUB8;
2 *
3 k1 = 0. 0. 0.;
4 k2 = long1 0. 0.;
5 k3 = long1 larg1 0.;
6 k4 = 0. larg1 0.;
7 *
8 L1 = DROI nlong1 k1 k2;
9 L2 = DROI nlarg1 k2 k3;
10 L3 = DROI nlong1 k3 k4;
11 L4 = DROI nlarg1 k4 k1;
12 *
13 S1 = DALLER L1 L2 L3 L4;

```

À partir de la surface S_1 , qui est la même que précédemment, nous allons créer le volume V_1 par translation selon le vecteur $Vec1$.

Nous en profitons également pour créer $Face_1$ sur laquelle porterons les conditions aux limites. Notons par exemple que la ligne L_4 appartient bien au maillage V_1 , puisque ce dernier est construit dessus. Par contre, la surface $face_1$ n'appartient pas à V_1 , il est donc nécessaire de la lier à V_1 en utilisant la commande ELIM.

```

1 Vec1=0. 0. (-1.0*ep1);
2 V1=S1 VOLU nep1 TRAN Vec1;
3 Face1 = L4 TRAN nep1 Vec1;
4 ELIM 0.0000001 V1 Face1;

```

Cette fois, le modèle correspond au maillage V_1 et utilise l'élément CUB8.

```
1 Model1 = MODL V1 MECANIQUE ELASTIQUE ISOTROPE CUB8;
```

On résout, puis on post-traite.

```
2 Mater1 = MATER Model1 YOUNG 70000.0 NU 0.33 RHO 2700.0;
3 MR1 = RIGID Model1 Mater1;
4 CL1 = BLOQ DEPL Face1;
5 CL2 = BLOQ ROTA Face1;
6 FOR1 = FORC(0. 0. Forc1) k3;
7 MR1 = MR1 ET CL1 ET CL2;
8 Depl1 = RESO MR1 FOR1 ;
9 *
10 * DEPLACEMENTS
11 UZ1 = EXCO 'UZ' depl1;
12 *TRAC CACH UZ1 V1;
13 * DEFORMEE
14 def0 = DEFO V1 Depl1 0.0 BLEU;
15 def1 = DEFO V1 Depl1;
16 *TRAC CACH (def0 ET def1);
17 *
18 * CONTRAINTES:
19 Sig1 = SIGM Model1 Mater1 Depl1;
20 * Les composantes de Sig1 sont: VONMISES, SMXX, SMYY, SMZZ, SMXY, SMXZ, SMYZ
21 Siigg1 = EXCO SMXX Sig1;
22 * on trace sur la geometrie deformee, c'est plus beau
23 TRAC CACH Siigg1 Model1 def1;
24 *
25 * DEFORMATIONS
26 *Eps1 = EPSI Model1 Mater1 Depl1;
27 *TRAC CACH Eps1 Model1;
28 *
29 fin;
```

11.4 Interpolation des champs et de la géométrie

Un élément est dit isoparamétrique si on prend les mêmes fonctions d'interpolation pour le déplacement et la géométrie.

De manière évidente, on définit également les éléments super et sous-paramétriques, comme illustré à la figure 11.5.

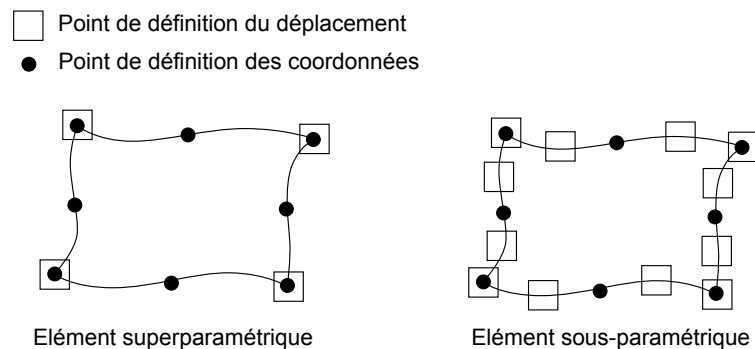


FIGURE 11.5: éléments finis non isoparamétriques

Ceci est juste une remarque en passant, pour définir le vocabulaire en somme, nous n'en reparlerons plus dans la suite du document.

Chapitre 12

Formulation pratique d'éléments finis

Résumé — L'intégralité de la méthode des éléments finis a été présentée au chapitre 10.

Dans ce chapitre et dans les suivants, nous allons détailler certains aspects. Nous proposons dans ce chapitre d'exposer un peu plus complètement les notions d'interpolation sur un élément, ainsi que le lien entre approximation locale (sur un élément) et approximation globale (construction de la base de V_h).

Nous avons dit vouloir interpoler le problème sur chaque élément. Pour ce faire, il faut prendre une base sur chaque élément. Plusieurs choix sont possibles, mais en général, les fonctions de base utilisées pour les éléments finis sont dites interpolantes, c'est-à-dire que les valeurs nodales sont les valeurs des grandeurs inconnues aux nœuds, et que c'est à partir de ces valeurs que l'on effectue l'interpolation.

La méthode la plus simple consiste à utiliser les polynômes de Lagrange. Dans cette méthode les fonctions de base valent 1 à un nœud du maillage et 0 à tous les autres. La fonction de base i est alors la fonction valant 1 au nœud i et 0 sur les autres nœuds et qui est polynomiale sur chaque élément. Il y a autant de fonctions de base par élément que de nombre de nœuds. On appelle élément la donnée d'une géométrie (souvent polygonale en deux dimensions, polyédrique en trois dimensions) et de fonctions de base associées à cette géométrie.

D'autres solutions peuvent exister pour les fonctions de base. Par exemple, les éléments finis d'Hermite ont la particularité d'avoir deux fonctions de base associées à chaque nœud. La valeur de la solution est alors ajustée avec la première fonction alors que la deuxième permet d'ajuster la valeur de la dérivée. Ce type de fonctions de base peut avoir un intérêt pour la résolution de certaines équations aux dérivées partielles (telle que l'équation des plaques en Mécanique des Milieux Continus), même si elle nécessite d'avoir deux fois plus de fonctions pour un maillage donné.

12.1 Éléments de Lagrange

Les éléments de Lagrange sont les éléments finis les plus simples.

12.1.1 Unisolvance

Définition 58 — Unisolvance. Soit $\Sigma = \{a_1, \dots, a_N\}$ un ensemble de N points distincts de \mathbb{R}^n . Soit P un espace vectoriel de dimension finie de fonctions de \mathbb{R}^n à valeurs dans \mathbb{R} . On dit que Σ est P -unisolvant si et seulement si pour tous réels $\alpha_1, \dots, \alpha_N$, il existe un unique élément p de P tel que $\forall i = 1, \dots, N, p(a_i) = \alpha_i$.

Cela revient à dire que la fonction de P dans \mathbb{R}^N qui à p fait correspondre $(p(a_1), \dots, p(a_N)) = (\alpha_1, \dots, \alpha_N)$ est bijective.

12.1.2 Éléments finis de Lagrange

Définition 59 — Éléments finis de Lagrange. Un élément fini de Lagrange est un triplet (K, Σ, P) tel que :

- K est un élément géométrique de \mathbb{R}^n , compact, connexe, et d'intérieur non vide ;
- $\Sigma = \{a_1, \dots, a_N\}$ est un ensemble de N points distincts de \mathbb{R}^n ;
- P est un espace vectoriel de dimension finie de fonctions réelles définies sur K , et tel que Σ soit P -unisolvant (donc $\dim P = N$).

Les fonctions de bases locales de l'élément fini de Lagrange (K, Σ, P) sont les N fonctions de P telles que $p_i(a_j) = \delta_{ij}$ pour $1 \leq i, j \leq N$.

Remarque. (p_1, \dots, p_N) est une base de P .

Définition 60 — Opérateur de P -interpolation. Un opérateur de P -interpolation sur Σ est un opérateur π_K qui à toute fonction v définie sur K associe la fonction $\pi_K v$ de P définie par :

$$\pi_K v = \sum_{i=1}^N v(a_i) p_i \quad (12.1)$$

Théorème 46 $\pi_K v$ est l'unique élément de P qui prend les mêmes valeurs que v sur les points de Σ .

On notera P_k l'espace vectoriel des polynômes de degré total inférieur ou égal à k .

- sur \mathbb{R} , $P_k = \text{vect}\{1, X, \dots, X^k\}$ et $\dim P_k = k + 1$;
- sur \mathbb{R}^2 , $P_k = \text{vect}\{X^i Y^j, 0 \leq i + j \leq k\}$ et $\dim P_k = \frac{(k+1)(k+2)}{2}$;
- sur \mathbb{R}^3 , $P_k = \text{vect}\{X^i Y^j Z^l, 0 \leq i + j + l \leq k\}$ et $\dim P_k = \frac{(k+1)(k+2)(k+3)}{6}$.

On notera Q_k l'espace vectoriel des polynômes de degré inférieur ou égal à k par rapport à chaque variable.

- sur \mathbb{R} , $Q_k = P_k$;
- sur \mathbb{R}^2 , $Q_k = \text{vect}\{X^i Y^j, 0 \leq i, j \leq k\}$ et $\dim Q_k = (k + 1)^2$;
- sur \mathbb{R}^3 , $Q_k = \text{vect}\{X^i Y^j Z^l, 0 \leq i, j, l \leq k\}$ et $\dim Q_k = (k + 1)^3$.

Éléments finis unidimensionnels

On discrétise le segment $[a; b]$ avec des polynômes de degrés 1 à m . On obtient les éléments du tableau 12.1.

Élément	P_1	P_2	...	P_m
K	$[a; b]$	$[a; b]$...	$[a; b]$
Σ	$\{a, b\}$	$\{a, \frac{a+b}{2}, b\}$...	$\{a + i \frac{b-a}{m}, i = 0, \dots, m\}$
P	P_1	P_2	...	P_m

Tableau 12.1: Éléments de Lagrange unidimensionnels de degrés 1 à m

Éléments finis bidimensionnels triangulaires

On discrétise le triangle de sommets $\{a_1, a_2, a_3\}$ avec, le long de chaque arête, une interpolation polynomiale de degré 1 à m . On obtient les éléments du tableau 12.2.

Élément	P_1	P_2
K	triangle de sommets $\{a_1, a_2, a_3\}$	triangle de sommets $\{a_1, a_2, a_3\}$
Σ	$\{a_1, a_2, a_3\}$	$\{a_{ij} = \frac{a_i + a_j}{2}, 1 \leq i, j \leq 3\}$
P	P_1	P_2

Tableau 12.2: Éléments de Lagrange bidimensionnels triangulaires de degrés 1 et 2

Remarque. Les fonctions de base pour l'élément P_1 sont définies par $p_i(a_j) = \delta_{ij}$. Ce sont les coordonnées barycentriques : $p_i = \lambda_i$.

Éléments finis bidimensionnels rectangulaires

On discrétise le rectangle de sommets $\{a_1, a_2, a_3, a_4\}$ de côtés parallèles aux axes. La formulation est décrite dans le tableau 12.3.

Élément	Q_1
K	rectangle de sommets $\{a_1, a_2, a_3, a_4\}$ de côtés parallèles aux axes.
Σ	$\{a_1, a_2, a_3, a_4\}$
P	Q_1

Tableau 12.3: Élément de Lagrange bidimensionnel rectangulaire de degré 1

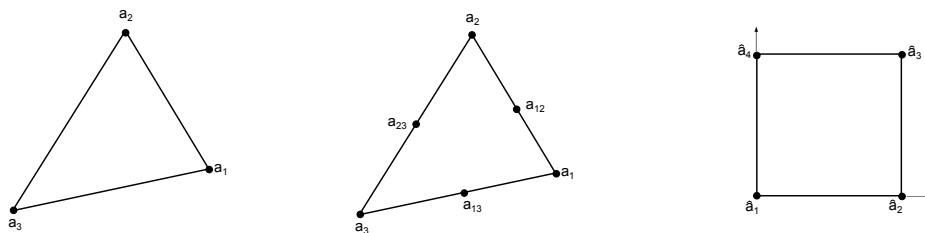


FIGURE 12.1: Éléments finis de Lagrange 2D : triangulaire P_1 , triangulaire P_2 et rectangulaire Q_1

Éléments finis tridimensionnels

On distingue les éléments finis tridimensionnels suivants, qui seront illustrés à la figure 12.2 :

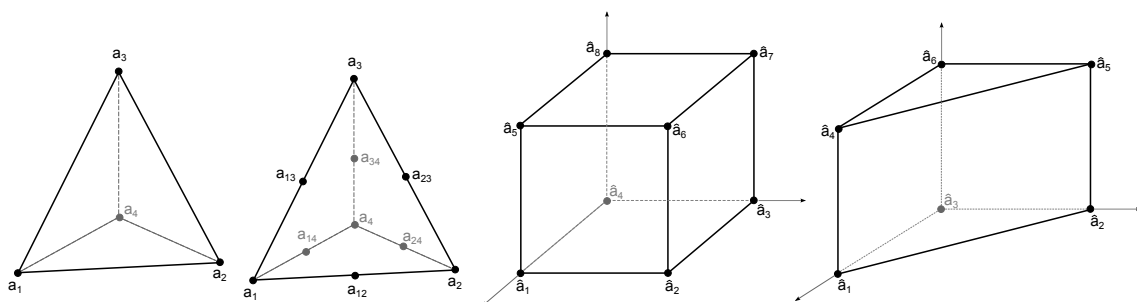


FIGURE 12.2: Éléments finis de Lagrange tridimensionnels : tétraédriques P_1 et P_2 , parallélépipédique Q_1 et prismatique

- les éléments tétraédriques, définis dans le tableau 12.4 ;
- les élément parallélépipédique, définis au tableau 12.5 ;
- et les élément prismatique, définis selon le tableau 12.6.

Élément	P_1	P_2
K	tétraèdre de sommets $\{a_1, a_2, a_3, a_4\}$	tétraèdre de sommets $\{a_1, a_2, a_3, a_4\}$
Σ	$\{a_1, a_2, a_3, a_4\}$	$\{a_i\}_{1 \leq i \leq 4} \cup \{a_{ij}\}_{1 \leq i < j \leq 4}$
P	P_1	P_2

Tableau 12.4: Éléments de Lagrange tridimensionnels tétrahédriques de degrés 1 et 2

Élément	Q_1
K	parallélépipède de sommets $\{a_1, \dots, a_8\}$ de côtés parallèles aux axes.
Σ	$\{a_i\}_{1 \leq i \leq 8}$
P	Q_1

Tableau 12.5: Élément de Lagrange tridimensionnel parallélépipédique de degré 1

Élément	Q_1
K	prisme droit de sommets $\{a_1, \dots, a_6\}$
Σ	$\{a_i\}_{1 \leq i \leq 6}$
P	$\{p(X, Y, Z) = (a + bX + cY) + Z(d + eX + fY), a, b, c, d, e, f \in \mathbb{R}\}$

Tableau 12.6: Élément de Lagrange tridimensionnel prismatique

12.1.3 Famille affine d'éléments finis et élément de référence

En fait, la notion de transformation affine entre un élément K d'un maillage \mathcal{T}_h et un élément de référence a déjà été utilisée en calcul de majoration d'erreur, mais la transformation elle-même n'a pas été détaillée.

Deux éléments finis $(\hat{K}, \hat{\Sigma}, \hat{P})$ et (K, Σ, P) sont affine-équivalents si et seulement si il existe une fonction affine F inversible telle que $K = F(\hat{K})$, $a_i = F(\hat{a}_i)$, $i=1, \dots, N$, et $P = \{\hat{p} \circ F^{-1}, \hat{p} \in \hat{P}\}$.

Définition 61 — Famille affine d'éléments finis. On appelle famille affine d'éléments finis une famille d'éléments finis tous affine-équivalents à un même élément fini appelé élément de référence.

D'un point de vue pratique, le fait de travailler avec une famille affine d'éléments finis permet de ramener tous les calculs d'intégrales à des calculs sur l'élément de référence. Voir figure 12.3 pour une illustration d'une transformation affine. Les éléments finis de références sont ceux obtenus,

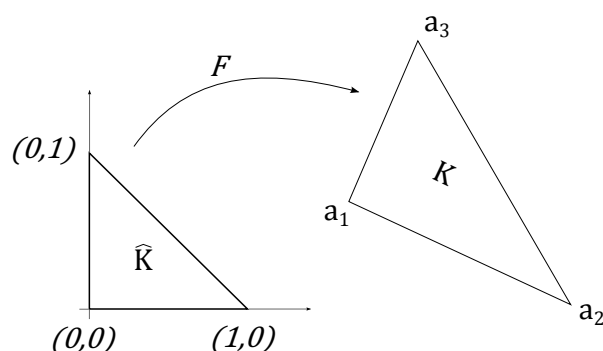


FIGURE 12.3: Transformation affine sur un triangle

dans les cas présentés ci-dessus avec :

- le segment $[0; 1]$ en une dimension ;
- le triangle unité de sommets $(0, 0)$, $(0, 1)$, $(1, 0)$,
- et le carré unité $[0; 1] \times [0; 1]$ en deux dimensions ;
- le tétraèdre unité de sommets $(0, 0, 0)$, $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$,

- le cube unité $[0; 1] \times [0; 1] \times [0; 1]$,
- le prisme unité de sommets $(0, 0, 0)$, $(0, 1, 0)$, $(1, 0, 0)$, $(0, 0, 1)$, $(0, 1, 1)$, $(1, 0, 1)$.

12.1.4 Construction de la base globale

Revenons à notre problème décrit par une équation aux dérivées partielles sous forme faible dans un domaine Ω sur lequel on réalise un maillage \mathcal{T}_h à partir d'une famille affine de N_e éléments finis $(K_i, \Sigma_i, P_i)_{i=1, \dots, N_e}$.

Par unisolvance, la solution approchée u_h sera entièrement définie sur chaque élément fini par ses valeurs sur les points de Σ_i , nommés nœuds du maillage. Notons (a_1, \dots, a_{N_h}) les nœuds du maillage ($N_h < N_e \cdot \text{card}_i$). Le problème approché revient à déterminer les valeurs de u_h aux points a_i : ce sont les degrés de liberté du problème approché. On va construire une base de V_h en associant à chaque degré de liberté a_i un vecteur de la base. On définit ainsi les fonctions de base globales φ_i ($i = 1, \dots, N_h$) par :

$$\varphi_i|_{K_j} \in P_j, j = 1, \dots, N_e \quad \text{et} \quad \varphi_i(a_j) = \delta_{ij}, 1 \leq i, j \leq N_h \quad (12.2)$$

et l'espace d'approximation interne est $V_h = \text{vect}\{\varphi_1, \dots, \varphi_{N_h}\}$. On remarquera qu'une telle fonction φ_i est nulle partout sauf sur les éléments dont a_i est un nœud.

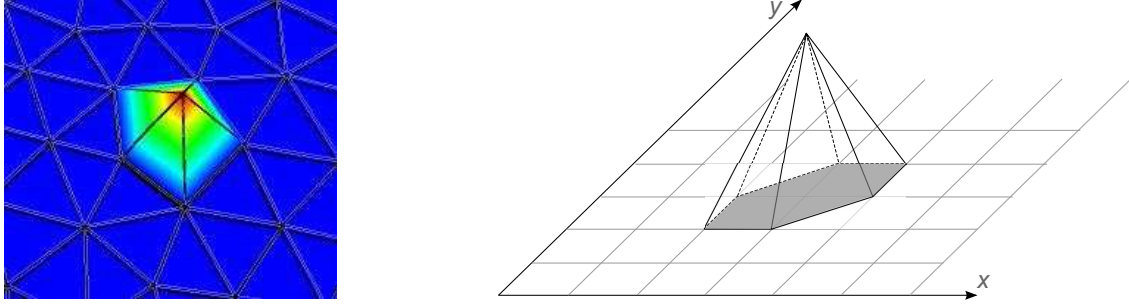


FIGURE 12.4: Base de V_h : exemple de fonction de base globale φ_i sur un maillage avec des éléments triangulaires P_1 .

De plus, sur un élément K dont a_i est un nœud, φ_i vaut 1 en a_i et 0 aux autres nœuds de K . Donc $\varphi_i|_K$ est une fonction de base locale de K . On voit donc que la fonction de base globale φ_i est construite comme réunion des fonctions de base locales sur les éléments du maillage dont a_i est un nœud.

Remarque. Ce qui précède est vrai dans le cas d'un maillage conforme, i.e si l'intersection entre deux éléments est soit vide, soit réduite à un sommet ou une arête en dimension 2, ou à un sommet, une arête ou une face en dimension 3.

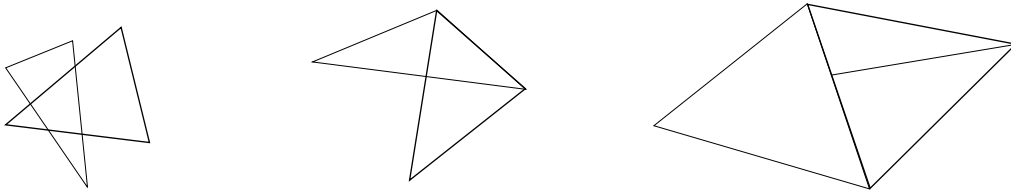


FIGURE 12.5: Maillage non conforme : situations interdites

12.1.5 Exemple : éléments finis \mathbb{P}_k en élasticité linéaire

La formulation faible du problème d'élasticité linéaire est donnée par la relation (8.42) que nous considérerons ici dans le cas de la condition de Dirichlet sur tout $\partial\Omega$, i.e. sous la forme :

trouver $u \in H_0^1(\Omega)^N$ tel que :

$$\int_{\Omega} \lambda \operatorname{div}(u) \operatorname{div}(v) + 2\mu \varepsilon(u) \cdot \varepsilon(v) = \int_{\Omega} f v, \quad \forall v \in H_0^1(\Omega)$$

Soit \mathcal{T}_h un maillage régulier de Ω . On introduit les espaces discrets :

$$V_h = \{u \in C(\Omega; \mathbb{R})^N / u|_K \in \mathbb{P}_k \text{ pour tout } K \in \mathcal{T}_h\}$$

et :

$$V_{0h} = \{u \in V_h / u = 0 \text{ sur } \partial\Omega\}$$

Soient $(\varphi_i)_{i=1, n_{ddl}}$ les fonctions de base associées aux degrés de liberté du treillis d'ordre k du maillage \mathcal{T}_h (n_{ddl} est le nombre de nœuds de degrés de liberté). L'approximation variationnelle du problème par la méthode des éléments finis \mathbb{P}_k consiste à déterminer $u \in V_{0h}$ tel que :

$$\int_{\Omega} \lambda \operatorname{div}(u) \operatorname{div}(v) + 2\mu \varepsilon(u) \cdot \varepsilon(v) = \int_{\Omega} f v, \quad \forall v \in V_{0h}$$

i.e. à résoudre le système :

$$\mathbf{K} \mathbf{q} = \mathbf{b}$$

avec :

$$\mathbf{K}_{ij} = \int_{\Omega} \lambda \operatorname{div}(\varphi_i) \operatorname{div}(\varphi_j) + 2\mu \varepsilon(\varphi_i) \cdot \varepsilon(\varphi_j)$$

et :

$$\mathbf{b}_i = \int_{\Omega} f \varphi_i$$

L'existence d'une solution à ce problème découle de l'application du théorème de Lax-Milgram.

Chaque élément de l'espace V_{0h} est défini de manière unique par ses valeurs dans \mathbb{R}^N aux nœuds de degrés de liberté du treillis d'ordre k , ce qui prouve que la dimension de l'espace V_{0h} est égale à $N.n_{ddl}$.

12.2 Éléments d'Hermite

12.2.1 Classe d'un élément fini

Une question naturelle est de savoir quelle est la régularité de la solution approchée u_h . En particulier, u_h est-elle continue ? dérivable ?

La solution approchée u_h est obtenue par combinaison linéaire des fonctions de base globales φ , la question revient donc à déterminer la régularité de celles-ci. Par construction, on voit que la régularité de φ_i sera donnée par sa régularité au niveau des interfaces entre les éléments adjacents formant son support.

Nous avons vu au paragraphe précédent que pour les éléments finis de Lagrange, φ_i est construite pour être continue d'un élément à l'autre, mais pas sa dérivée... c'est cette contrainte que nous allons introduire maintenant.

12.2.2 Éléments finis d'Hermite

Définition 62 — Éléments fini d'Hermite. Un élément fini d'Hermite ou élément fini général est un triplet (K, Σ, P) tel que :

- K est un élément géométrique de \mathbb{R}^n , compact, connexe, et d'intérieur non vide ;
- $\Sigma = \{\sigma_1, \dots, \sigma_N\}$ un ensemble de N formes linéaires sur l'espace des fonctions définies sur K , ou sur un sous-espace plus régulier contenant P ;
- P est un espace vectoriel de dimension finie de fonctions réelles définies sur K , et tel que Σ soit P -unisolvant.

Définition 63 — Opérateur de P-interpolation. Un opérateur de P-interpolation sur Σ est un opérateur π_K qui à toute fonction v définie sur K associe la fonction $\pi_K v$ de P définie par :

$$\pi_K v = \sum_{i=1}^N \sigma_i(v) p_i \quad (12.3)$$

Théorème 47 $\pi_K v$ est l'unique élément de P qui prend les mêmes valeurs que v sur les points de Σ .

On remarque immédiatement que si $\sigma_i(p) = p(ai)$, $i = 1, \dots, N$, on retrouve les éléments finis de Lagrange. Cette généralisation permet d'introduire des opérations de dérivation dans Σ , et donc d'améliorer la régularité des fonctions de V_h . Les fonctions de base globales φ_i , ($i = 1, \dots, N_h$) sont définies par :

$$\varphi_i|_{K_j} \in P_j, j = 1, \dots, N_e \quad \text{et} \quad \sigma_j(\varphi_i) = \delta_{ij}, 1 \leq i, j \leq N_h \quad (12.4)$$

Suivant les éléments utilisés, ces fonctions de base pourront être de classe C^1 ou même plus, et il en sera donc de même pour la solution approchée u_h .

12.2.3 Éléments uni- et bidimensionnels

Nous présentons quelques éléments finis d'Hermite utilisés classiquement, dont certains sont illustrés à la figure 12.6 :

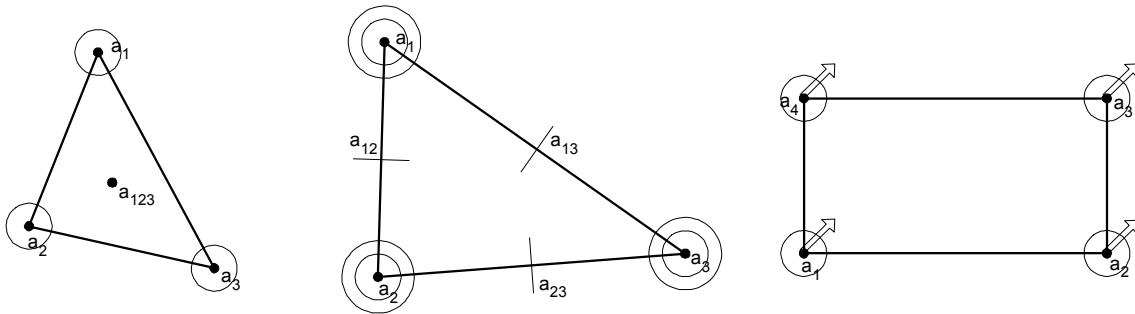


FIGURE 12.6: Éléments finis d'Hermite triangulaire cubique, élément d'Argyris et élément rectangulaire Q_3

- les éléments finis unidimensionnels cubiques et quintiques sont définis au tableau 12.7 ;
- les éléments finis bidimensionnels triangulaires : on distingue l'élément cubique d'Hermite, qui est C^0 , et l'élément d'Argyris, qui est C^1 . Ils sont présentés aux tableaux 12.8 et 12.9 ;
- l'élément fini bidimensionnel rectangulaire défini au tableau 12.10 ;

Élément	cubique	quintique
K	segment $[a; b]$	segment $[a; b]$
Σ	$\{p(a), p'(a), p(b), p'(b)\}$	$\{p(a), p'(a), p''(a), p(b), p'(b), p''(b)\}$
P	P_3	P_3
Régularité	C^1 et H^2	C^2 et H^3

Tableau 12.7: Éléments d'Hermite unidimensionnels de degrés 3 et 5

Élément	
K	triangle de sommets $\{a_1, a_2, a_3\}$
Σ	$\{p(a_i), \frac{\partial p}{\partial x}(a_i), i = 1, 2, 3\} \cup \{p(a_0)\}$
P	P_3
Régularité	C^0 , mais pas C^1

Tableau 12.8: Élément bidimensionnel triangulaire d'Hermite

Élément	
K	triangle de sommets $\{a_1, a_2, a_3\}$
Σ	$\{p(a_i), \frac{\partial p}{\partial x}(a_i), \frac{\partial p}{\partial y}(a_i), \frac{\partial^2 p}{\partial x^2}(a_i), \frac{\partial^2 p}{\partial y^2}(a_i), \frac{\partial^2 p}{\partial x \partial y}(a_i), i = 1, 2, 3\} \cup \{\frac{\partial p}{\partial n}(a_{ij}), 1 \leq i < j \leq 3\}$
P	P_5
Régularité	C^1

Tableau 12.9: Élément bidimensionnel triangulaire d'Argyris

Élément	Q_3
K	rectangle de sommets $\{a_1, a_2, a_3, a_4\}$ de côtés parallèles aux axes
Σ	$\{p(a_i), \frac{\partial p}{\partial x}(a_i), \frac{\partial p}{\partial y}(a_i), \frac{\partial^2 p}{\partial x \partial y}(a_i), i = 1, \dots, 4\}$
P	P_3
Régularité	C^1

Tableau 12.10: Élément bidimensionnel rectangulaire d'Hermite Q_3

12.2.4 Exemple : éléments finis d'Hermite pour l'équation des plaques

On se propose donc de résoudre la formulation variationnelle des plaques, donnée par l'équation (8.57) :

$$\int_{\Omega} \Delta u \Delta v = \int_{\Omega} f v, \quad \forall v \in H_0^2(\Omega)$$

en dimension 2 par une méthode d'éléments finis de type Hermite \mathbb{P}_5 .

Pour un maillage \mathcal{T}_h , on introduit l'espace discret :

$$V_h = \{v \in C^1(\overline{\Omega}) \text{ tel que } v|_{K_i} \in \mathbb{P}_5 \text{ pour tout } K_i \in \mathcal{T}_h\}$$

Nous allons vérifier l'unisolvance et le raccord au niveau des mailles pour les polynômes \mathbb{P}_5 .

L'unisolvance consiste donc à montrer que tout polynôme $p \in \mathbb{P}_5$ est caractérisé de manière unique sur un triangle K par les 21 valeurs réelles suivantes :

$$p(a_i), \quad \nabla p(a_i), \quad \nabla \nabla p(a_i), \quad \frac{\partial p(b_i)}{\partial n} \quad (12.5)$$

avec $i = 1, 2, 3$, (a_1, a_2, a_3) les sommets de K, (b_1, b_2, b_3) les milieux des côtés de K, et $\partial p(b_i)/\partial n$ la dérivée normale au côté de b_i .

Comme \mathbb{P}_5 est un espace de dimension 21, il suffit donc de montrer que l'application qui à un élément de \mathbb{P}_5 associe les 21 valeurs définies ci-dessus est injective (et donc bijective).

Quitte à effectuer un changement de variables, on se ramène au cas du triangle équilatéral défini par $a_1 = (-1, 0)$, $a_2 = (1, 0)$.

Soit $p \in \mathbb{P}_5$ un polynôme annulant les 21 valeurs, montrons qu'alors il est nul. Posons $q_1(x) = p(x, 0)$ et $q_2(x) = \partial p / \partial x_2(x, 0)$. Par hypothèse on a :

$$q_1(\pm 1) = q_1'(\pm 1) = q_1''(\pm 1) = 0$$

Comme q_1 est de degré au plus 5, $q_1 = 0$. Ainsi, p est divisible par x_2 , et donc il existe un polynôme $q(x_1, x_2)$ tel que $p(x_1, x_2) = x_2 q(x_1, x_2)$.

De même, comme $\partial p / \partial x_2 = \partial p / \partial n$ sur le segment $[a_1, a_2]$, on a donc par hypothèse que :

$$q_2(\pm 1) = q_2'(\pm 1) = q_2(0) = 0$$

et comme q_2 est de degré au plus 4, $q_2 = 0$.

Or $q_2(x) = q(x, 0)$, et donc q est divisible par x_2 . Et donc p est divisible par x_2^2 .

Par invariance par changement de côté, on en déduit de la même manière que p est divisible par $(1 + x_1 - x_2/\sqrt{3})^2$ et $(1 - x_1 - x_2/\sqrt{3})^2$. Le polynôme p , qui est de degré au plus 5, est donc divisible par un polynôme de degré 6, c'est donc qu'il est nul.

Le raccordement au niveau des mailles revient à prouver le lemme suivant :

Soient K et K' deux triangles ayant une arête commune $\Gamma = (a_1, a_2)$. Soient p_K et $p_{K'}$ deux éléments de \mathbb{P}_5 . Alors la fonction v définie sur $K \cup K'$ par :

$$v(x) = \begin{cases} p_K(x) & \text{si } x \in K \\ p_{K'}(x) & \text{si } x \in K' \end{cases}$$

est de classe C^1 si et seulement si, pour $i = 1, 2$:

$$p_K(a_i) = p_{K'}(a_i), \quad \nabla p_K(a_i) = \nabla p_{K'}(a_i), \quad \nabla \nabla p_K(a_i) = \nabla \nabla p_{K'}(a_i), \quad \frac{\partial p_K(b_i)}{\partial n} = \frac{\partial p_{K'}(b_i)}{\partial n} \quad (12.6)$$

avec les mêmes notations que pour l'équation (12.5).

L'application v est de classe C^1 si et seulement si les restrictions de p_K et $p_{K'}$ coïncident sur l'arête commune Γ , ainsi que les polynômes $\partial p_K / \partial n$ et $\partial p_{K'} / \partial n$.

Les polynômes p_K et $p_{K'}$, de degré au plus 5, coïncident sur Γ si et seulement si pour $i = 1, 2$:

$$p_K(a_i) = p_{K'}(a_i), \quad \frac{\partial p_K}{\partial \tau}(a_i) = \frac{\partial p_{K'}}{\partial \tau}(a_i), \quad \frac{\partial^2 p_K}{\partial \tau^2}(a_i) = \frac{\partial^2 p_{K'}}{\partial \tau^2}(a_i)$$

avec τ le vecteur unitaire tangent à l'arête.

Les polynômes $\partial p_K / \partial n$ et $\partial p_{K'} / \partial n$, de degré au plus 4, coïncident sur Γ si et seulement si pour $i = 1, 2$:

$$\frac{\partial p_K}{\partial n}(a_i) = \frac{\partial p_{K'}}{\partial n}(a_i), \quad \frac{\partial^2 p_K}{\partial n^2}(a_i) = \frac{\partial^2 p_{K'}}{\partial n^2}(a_i), \quad \frac{\partial p_K}{\partial n}(b) = \frac{\partial p_{K'}}{\partial n}(b)$$

On a donc prouvé que si les relations (12.6) sont vérifiées, alors v est de classe C^1 . Réciproquement, si v est de classe C^1 , les restrictions de $\partial p_K / \partial n$ et $\partial p_{K'} / \partial n$ à l'arête commune Γ coïncident, et pour $i = 1, 2$, on a $\partial^2 p_K = \partial n \partial \tau(a_i) = \partial^2 p_{K'} = \partial n \partial \tau(a_i)$. Les autres conditions étant satisfaites, le lemme est démontré.

Profitons-en pour exposer la méthode d'Argyris.

Notons tout d'abord que l'espace V_h est inclus dans $H^2(\Omega)$.

Ensuite, d'après le point précédent, un élément v de V_h est entièrement déterminé par les valeurs de v , ∇v et $\nabla \nabla v$ aux nœuds du maillage, ainsi que par le flux $\partial v / \partial n(b_k)$, b_k parcourant les milieux des arêtes k du maillage.

Notons n_s le nombre d'arêtes (side), et n_c le nombre de sommets (corner). On peut donc construire une base de V_h formée des éléments $(\varphi_{i,\alpha})_{(i,\alpha)}$ et (ψ_k) où $i \in \{1, \dots, n_s\}$, $\alpha \in \mathbb{N}^2$ le multi-indice tel que $|\alpha| \leq 2$ et $k \in \{1, \dots, n_c\}$ définis par :

$$\partial^\beta \varphi_{i,\alpha}(a_j) = \delta_j^i \delta_\alpha^\beta, \quad \frac{\partial \varphi_{i,\alpha}}{\partial n}(b_l) = 0, \quad \partial^\beta \psi_k(a_j) = 0, \quad \frac{\partial \psi_k}{\partial n}(b_l) = \delta_l^k$$

pour tout $j \in \{1, \dots, n_s\}$, $l \in \{1, \dots, n_c\}$, $\beta \in \mathbb{N}^2$ le multi-indice tel que $|\beta| \leq 2$.

Afin de résoudre l'équation des plaques, on introduit le sous-espace de V_h :

$$V_{0h} = V_h \cap H_0^2(\Omega)$$

qui est l'ensemble des fonctions de V_h qui s'annulent, ainsi que leurs dérivées partielles, sur le bord $\partial\Omega$. Cet espace est engendré par les éléments $(\varphi_{i,\alpha})$ et (ψ_k) où $i \in \{1, \dots, n_{s0}\}$ et $k \in \{1, \dots, n_{c0}\}$ parcourent respectivement les arêtes et sommets n'appartenant pas au bord de Ω . L'approximation variationnelle consiste à trouver $u_h \in V_{0h}$ tel que :

$$\int_{\Omega} \Delta u_h \Delta v_h = \int_{\Omega} f v_h, \quad \forall v \in V_{0h}$$

qui, d'après le théorème de Lax-Milgram, admet une solution unique. Il est équivalent au système matriciel :

$$\mathbf{K}_h \mathbf{q}_h = \mathbf{F}_h$$

où la matrice de rigidité, de taille $6n_{s0} + n_{c0}$, est définie par blocs de la manière suivante :

$$\mathbf{K}_h = \begin{bmatrix} \mathbf{D}_h & \mathbf{F}_h \\ \mathbf{F}_h^T & \mathbf{H}_h \end{bmatrix}$$

La matrice \mathbf{D}_h est également définie par blocs. Elle est constituée de 6×6 blocs de taille $n_{s0} \times n_{s0}$:

$$\mathbf{D}_h = \mathbf{E}_{\mathbf{h}}^{ij}{}_{(i,j) \in \{1, \dots, 6\}^2}, \text{ avec } \mathbf{E}_{\mathbf{h}}^{ij}{}_{kl} = \int_{\Omega} \Delta \varphi_{k,s_i} \Delta \varphi_{l,s_j}, \text{ où } (k,l) \in \{1, \dots, n_{s0}\}^2$$

avec s_i parcourant les multi-indices \mathbb{N}^2 de degré ≤ 2 .

La matrice \mathbf{F}_h est également définie par blocs. Elle est un « vecteur » colonne constitué de 6 blocs de taille $n_{s0} \times n_{c0}$:

$$\mathbf{F}_h = \mathbf{G}_{\mathbf{h}}^i{}_{i \in \{1, \dots, 6\}}, \text{ avec } \mathbf{G}_{\mathbf{h}}^i{}_{kl} = \int_{\Omega} \Delta \varphi_{k,s_i} \Delta \psi_l, \text{ où } (k,l) \in \{1, \dots, n_{s0}\} \times \{1, \dots, n_{c0}\}$$

La matrice \mathbf{H}_h , de taille $n_{c0} \times n_{c0}$ est définie par :

$$\mathbf{H}_{\mathbf{h}}{}_{kl} = \int_{\Omega} \Delta \psi_k \Delta \psi_l, \text{ où } (k,l) \in \{1, \dots, n_{c0}\}^2$$

Le vecteur \mathbf{F}_h comporte $6n_{s0} + n_{c0}$ composantes et est défini par :

$${}^T \mathbf{F}_h = {}^T \mathbf{c}_h^1, \dots, \mathbf{c}_h^6, \mathbf{d}_h$$

avec $c_h^i \in \mathbb{R}^{n_{s0}}$ et $d_h \in \mathbb{R}^{n_{c0}}$ les vecteurs :

$$(c_h^i)_k = \int_{\Omega} f_h \varphi_{k,s_i}, k \in \{1, \dots, n_{s0}\}, i \in \{1, \dots, 6\}$$

$$(d_h)_k = \int_{\Omega} f_h \psi_k, k \in \{1, \dots, n_{c0}\}$$

La solution u_h de l'approximation variationnelle est telle que :

$$u_h = \sum_{i=0}^5 \sum_{k=1}^{n_{s0}} q_h^{in_{s0}+k} \varphi_{k,s_{i+1}} + \sum_{k=1}^{n_{c0}} q_h^{6n_{s0}+k} \psi_k$$

avec \mathbf{q}_h la solution du système matriciel.

12.3 Traitement de plusieurs champs

Dans ce qui précède, le lecteur aura noté que l'on interpole un seul champ, en imposant ou non certaines régularités (dérivées partielles). Une question naturelle consiste à se demander comment traiter les problèmes ayant plusieurs champs inconnus (cela a déjà été abordé, et certaines stratégies on déjà été présentées : formulation mixte (Brezzi), utilisation de multiplicateurs de Lagrange).

On pourrait imaginer, si ces champs sont « dissociés » (sans lien les uns avec les autres), de modéliser chaque champ séparément, i.e. de traiter un problème par champ (avec pour chaque champ un choix de maillage et d'élément qui lui est propre).

En fait, les problèmes à plusieurs champs ne concernent généralement pas des champs indépendants. Lorsque l'on a plusieurs champs interdépendants, alors on dispose de relations entre ces champs, qui doivent elles-aussi être satisfaites.

Nous avons exposé au chapitre précédent pourquoi la mécanique reste un cas compliqué. Cela provient de ce que les champs inconnus sont les déplacements, déformations et contraintes. De plus, le chapitre précédent a permis une illustration du choix d'un modèle, même lorsque l'on ne recourt qu'à une théorie à un champ (déplacements).

Nous ne construirons pas ici un élément à plusieurs champs, mais nous allons donner une motivation pour le faire : celui du calcul des contraintes à l'interface entre deux matériaux différents que nous supposons parfaitement collés.

Considérons donc le problème décrit à la figure 12.7. Deux domaines Ω_1 et Ω_2 , constitués

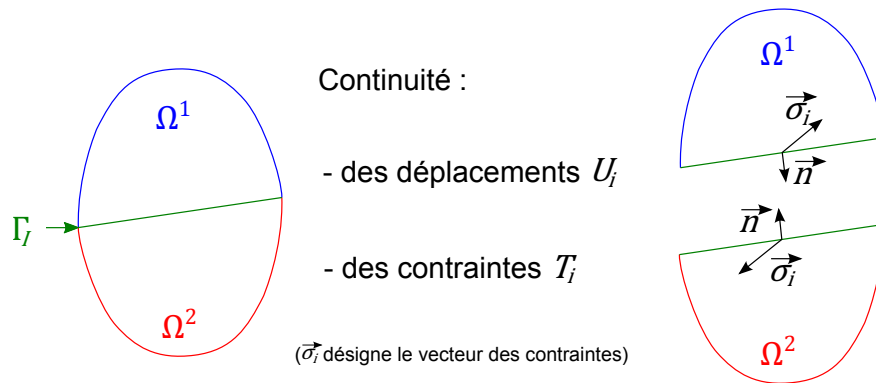


FIGURE 12.7: Interface et continuités

chacun de leur propre loi de comportement, ont une interface commune Γ_I . On supposera que le champ de déplacement est continu le long de Γ_I . La continuité de la composante normale du déplacement à l'interface traduit le fait qu'il n'y a pas décollement entre les deux domaines ; celle de la composante tangentielle qu'il n'y a pas glissement entre eux. On peut donc dire qu'en tout point de Γ_I , le déplacement est continu, ce que l'on note $u_i^1 = u_i^2$.

L'état d'équilibre des forces doit lui-aussi être vérifié le long de l'interface. Cela implique que les composantes normales des contraintes doivent être également continues le long de cette interface, i.e. que la trace du tenseur des contraintes doit être continue le long de l'interface. Par contre, les autres composantes peuvent (et doivent selon les cas) être discontinues.

Plusieurs stratégies sont envisageables :

Post-traitement : Dans cette méthode, on effectue le calcul en déplacements, de manière classique.

On obtient les contraintes de manière toute aussi classique, mais celles-ci n'ont évidemment pas les continuités et discontinuités souhaitées.

On utilise une méthode de post-traitement qui va modifier le calcul des contraintes sur les éléments situés de part et d'autre de l'interface, par exemple en imposant la vérification des équations d'équilibre (voir par exemple la méthode de Reissner locale, qui applique la fonctionnelle de Reissner uniquement le long de l'interface).

Élément mixte : La fonctionnelle d'Hellinger-Reissner possède les champs de déplacement et de contrainte comme inconnues. Elle conduit à un système de type mixte donné par l'équation (11.11), avec une matrice qui n'est plus définie-positive.

Il s'en suit que ces deux champs sont continus, et notamment que toutes les composantes des contraintes le sont, ce qui ne satisfait pas les conditions d'équilibre.

Si l'on souhaite utiliser ce type d'élément, il faudra par exemple, faire une condensation statique des composantes devant être discontinues, ou utiliser une méthode de post-traitement.

Élément hybride : La fonctionnelle de Pian et Tong (mixte hybride) présentée précédemment a l'avantage de ne pas faire intervenir de dérivée. Elle conduit elle-aussi à un système de type mixte, mais la matrice de rigidité n'est même plus symétrique (dans une discrétisation « brutale »). On sait la retraiter.

Toutes les composantes des contraintes sont là aussi continues, et il faudra appliquer les mêmes remèdes que ci-dessus.

Notons qu'un avantage de cette formulation c'est qu'elle nécessite la matrice de souplesse [S] au lieu de la matrice de Hooke [H], ce qui permet de traiter le cas des matériaux incompressibles.

Multiplicateurs de Lagrange : D'une manière évidente, la continuité du champ de déplacements à l'interface Γ_I entre deux éléments ¹ et ² peut être imposée par l'ajout à la fonctionnelle de la condition :

$$- \int_{\Gamma_I} \lambda_i (u_i^1 - u_i^2) d\Gamma \quad (12.7)$$

Cette condition est facile à écrire et facile à implémenter (mais on perd la définie-positivité de la matrice de rigidité).

De plus, l'interprétation physique de ces multiplicateurs de Lagrange montre que ceux-ci sont égaux aux composantes normales des contraintes.

Voilà quelques stratégies. D'autres peuvent exister, surtout si en plus on veut passer sur des modèles poutre ou plaque. Le but était de montrer que non seulement il est possible de développer de nombreux éléments finis, mais que même à partir d'éléments existants, il est toujours possible de construire une méthode numérique permettant d'obtenir les résultats souhaités. Les méthodes de post-traitement, que nous n'aborderons pas dans ce document, sont très riches et permettent de faire beaucoup de choses. Elles ont en outre l'avantage d'être parfois plus faciles à implémenter dans des codes industriels que de nouveaux éléments.

12.4 Validation pratique et indicateurs d'erreur

Les tests numériques des éléments et au delà des codes de calcul sont indispensables. Non seulement ils permettent de vérifier la satisfaction de critères de convergence et d'évaluer la précision des éléments finis développés, ce qui peut être réalisé par une analyse mathématique, mais ils permettent également de s'assurer de la bonne programmation. Plusieurs types de tests peuvent et doivent être faits, sur plusieurs problèmes types. Ces tests concernent aussi bien un seul élément, que plusieurs (patch-test), et même différents maillages. En mécanique, on distingue les cas listés dans le tableau 12.11.

12.4.1 Modes rigides et parasites

Le problème consiste à savoir si un élément fini, que l'on définira ici par sa matrice **K**, peut présenter l'état de déformation nulle ou d'énergie interne nulle (mode ou mouvement de corps rigide).

Une méthode consiste à déterminer le nombre de valeurs propres nulles de **K**. Ce nombre doit être égal au nombre de modes rigides : soit 3 en deux dimensions (2 translations, 1 rotation), 6 en trois dimensions et 1 en axisymétrique. On notera m_r le nombre de modes rigides. S'il y a plus de

Problème / solution de référence	Maillage	Vérification	Importance
Mouvement rigide et déformations constantes	1 élément	base polynomiale complète / modes parasites	convergence du modèle élément fini vers la solution théorique lorsque le nombre d'élément tend vers l'infini. Test particulièrement pour les non standard ou non conformes.
Mouvement rigide et déformations constantes	plusieurs éléments : patch-test cinématique ou mécanique	base complète, conformité, indépendance par rapport au maillage (vérification de l'assemblage), absence de modes parasites	
Champ de déformations générales	différents maillages	précision et influence de la distorsion sur les déplacements et la contrainte	qualité de l'élément par rapport à d'autres éléments existants.

Tableau 12.11: Test de validation

valeurs propres nulles que de modes rigides, alors c'est qu'il y a des modes parasites à énergie nulle. Ces modes parasites doivent disparaître après assemblage de plusieurs éléments afin d'éviter que la matrice de rigidité soit singulière (d'où le test sur plusieurs éléments). Le schéma d'intégration choisi pour la définition de l'élément (par exemple intégration complète, réduite ou sélective) peut influencer sur les modes parasites. C'est pourquoi, il n'est pas toujours aussi simple de déterminer mathématiquement ce facteur.

Dans ce cas, il est possible d'introduire un champ de déplacement représentant le mouvement rigide. Pour chaque mode rigide i , le vecteur \mathbf{u}_n^i associé à la matrice \mathbf{K} doit vérifier :

$$\mathbf{K}\mathbf{u}_n^i = \mathbf{0} \quad (12.8)$$

Un mode rigide quelconque étant une combinaison linéaire des modes \mathbf{u}_n^i , on construit un problème éléments finis dans lequel on impose m_r valeurs du vecteur \mathbf{u}_n^i et on détermine numériquement les $n - m_r$ composantes (non imposées) de \mathbf{u}_n^i . Celles-ci doivent être identiques aux valeurs théoriques associées au mode rigide i . Si ce n'est pas le cas, c'est qu'il existe des modes rigides.

12.4.2 Modes associés aux déformations constantes

Pour converger correctement l'élément doit également représenter exactement l'état de déformations ou contraintes constantes (ou plus généralement la représentation « constante » de tous les termes de la forme variationnelle).

Dans un premier temps, on calcule les efforts nodaux correspondant à une contrainte fixée. Dans un second temps, on introduit les efforts nodaux comme condition aux limites dans le modèle élément fini et on vérifie si l'on retrouve bien l'état de contrainte initialement choisi.

12.4.3 Patch-tests

Le domaine choisi doit posséder au moins un nœud à l'intérieur du domaine. On définit un champ de déplacement conduisant à un état de déformation désiré (constant ou non), que l'on introduit comme condition aux limites dans le modèle élément fini. On vérifie qu'au point intérieur, l'état de déformation obtenu est bien celui désiré.

12.4.4 Test de précision d'un élément

On confronte le modèle élément fini à un ou des cas tests représentatifs de ce pour quoi l'élément a été développé (élasticité tridimensionnelle, bidimensionnelle, problème de torsion, problème avec concentrations locales de déformations ou de contraintes, prise en compte des frontières courbes...).

Sur les composantes d'« intérêt », on regardera s'il y a bien convergence du modèle et sa vitesse de convergence en fonction du nombre d'éléments, de la distorsion de ceux-ci...

La « mesure » de l'écart peut demander une « jauge » particulière. C'est ce que l'on appelle un estimateur d'erreur. Nous n'avons abordé jusqu'à présent que la mesure de l'erreur due à la méthode des éléments finis elle-même, mais chaque problème particulier peut demander un estimateur particulier. Pour la mécanique, les estimateurs sont souvent liés à la qualité des contraintes ou à des estimateurs d'énergie.

Un indicateur local d'erreur pourra par exemple être basé sur l'écart entre les contraintes aux nœuds et aux frontières après extrapolation des points d'intégration de chaque élément. On rappelle que certaines contraintes doivent être nulles sur les frontières libres.

On trouve par exemple l'écart sur les contraintes équivalentes (selon un sens à définir en fonction du type de matériau : par exemple von Mises pour un matériau isotrope, Hill, Tsai-Hill, Tsai-Wu... pour les matériaux anisotropes, les mousses, les os, les composites...); l'écart sur la contrainte moyenne...

On peut également utiliser la densité d'énergie interne de déformation sur chaque élément, l'idéal étant que tous les éléments contribuent de manière identique à l'estimation de l'énergie interne totale.

Un indicateur global d'erreur pourra être par exemple la valeur de l'énergie potentielle totale. Plus sa valeur sera petite, meilleur sera le modèle.

On peut également s'assurer que certaines équations d'équilibre sont vérifiées à l'intérieur d'un élément ou le long d'une frontière entre éléments (par exemple à l'interface entre deux domaines, un problème évoqué un peu avant au paragraphe 12.3)...

12.5 Exemple : quelques variations sur le thème des éléments unidimensionnels

Dans ce paragraphe, nous appliquons ce que nous venons de voir sur quelques cas simples d'éléments unidimensionnels.

12.5.1 Élément de référence unidimensionnel linéaire à deux nœuds

Considérons un segment reliant deux points dans l'espace. Ce segment est défini par un point courant dont les coordonnées sont le vecteur \mathbf{x} . On peut paramétrer ce segment par le paramètre $\xi \in [-1, 1]$, et le vecteur position s'exprime par ${}^T\mathbf{x} = {}^T\mathbf{x}(\xi), y(\xi), z(\xi)$. Cet élément (segment) est défini par

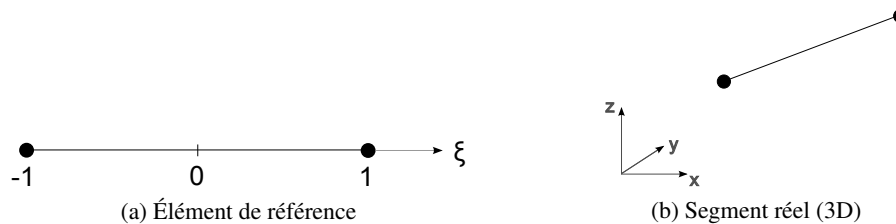


FIGURE 12.8: Correspondance entre élément de référence et géométrie réelle

ses seules extrémités. Nous avons donc deux nœuds de coordonnées (x_1, y_1, z_1) et (x_2, y_2, z_2) . Un point courant de cet élément sera obtenu par interpolation linéaire entre les deux nœuds, paramétrée par ξ . On cherche cette interpolation sous la forme $x(\xi) = {}^T\mathbf{N}(\xi)\mathbf{x}_n$, $y(\xi) = {}^T\mathbf{N}(\xi)\mathbf{y}_n$ et $z(\xi) = {}^T\mathbf{N}(\xi)\mathbf{z}_n$, avec ${}^T\mathbf{N}(\xi) = {}^T(N_1(\xi), N_2(\xi))$ et $N_1(\xi) = \frac{1}{2}(1 - \xi)$, $N_2(\xi) = \frac{1}{2}(1 + \xi)$. De manière plus « compacte », on peut écrire :

$$N_i(\xi) = \frac{1}{2}(1 + \xi_i \xi), \quad i = 1, 2, \quad \xi_i = \pm 1 \quad (\text{et } \xi \in [-1, 1]) \quad (12.9)$$

Dans cette interpolation (comme dans toute interpolation), on a une relation entre $d\mathbf{x}$ et $d\xi$. Notons $d\mathbf{x} = \mathbf{a}d\xi$. Alors :

$$\mathbf{t}_\mathbf{a} = {}^T(x_{,\xi}, y_{,\xi}, z_{,\xi}) = \frac{1}{2} {}^T(x_{21}, y_{21}, z_{21}) \quad (12.10)$$

avec ici $x_{21} = x_2 - x_1$, $y_{21} = y_2 - y_1$ et $z_{21} = z_2 - z_1$. On voit ainsi comment passer d'une intégrale sur le segment réel à une intégrale sur l'élément de référence.

De même, on peut chercher la relation entre $ds = d\mathbf{x} \cdot d\mathbf{x}$. Nous notons $ds = m d\xi$. On a alors :

$$m = |\mathbf{a}| = \frac{1}{2} \sqrt{x_{21}^2 + y_{21}^2 + z_{21}^2} = \frac{L_e}{2} \quad (12.11)$$

avec L_e la longueur de l'élément. On voit alors l'égalité des intégrations :

$$\int_{L_e} \dots ds = \int_{-1}^{+1} \dots m d\xi \quad (12.12)$$

Par ailleurs, on a la relation :

$$\frac{d}{d\xi} = \frac{ds}{d\xi} \frac{d}{ds} = m \frac{d}{ds} \quad (12.13)$$

12.5.2 Rappels sur la jacobienne et le jacobien d'une transformation

Nous considérons une transformation de \mathbb{R}^n dans \mathbb{R}^m donnée par la fonction $f : (x_1, \dots, x_n) \mapsto (f_{y_1}, \dots, f_{y_m})$. La matrice jacobienne (d'après Carl Jacobi) associée à f est définie par :

$$\mathbf{J}_f = \begin{bmatrix} {}^T\nabla \mathbf{f}_{y_1} \\ \vdots \\ {}^T\nabla \mathbf{f}_{y_m} \end{bmatrix} \quad (12.14)$$

i.e., sous forme complètement développée la jacobienne de f s'écrit :

$$\mathbf{J}_f = \begin{bmatrix} \frac{\partial f_{y_1}}{\partial x_1} & \dots & \frac{\partial f_{y_1}}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{y_m}}{\partial x_1} & \dots & \frac{\partial f_{y_m}}{\partial x_n} \end{bmatrix} \quad (12.15)$$

Au voisinage d'un point \mathbf{M} , l'approximation linéaire de la fonction f est donnée par :

$$f(\mathbf{X}) \approx f(\mathbf{M}) + \mathbf{J}_f(\mathbf{M}) \mathbf{MX} \quad (12.16)$$

La composée $f \circ g$ de fonctions différentiables est différentiable, et sa matrice jacobienne s'obtient par la formule :

$$\mathbf{J}_{f \circ g} = (\mathbf{J}_f \circ g) \cdot \mathbf{J}_g \quad (12.17)$$

Dans le cas où $m = n$, on appelle j_f le jacobien de f , défini comme le déterminant de sa matrice jacobienne :

$$j_f = \det(\mathbf{J}_f) \quad (12.18)$$

Dire que le jacobien est non nul revient donc à dire que la matrice jacobienne est inversible. Si le jacobien est positif au point \mathbf{M} , l'orientation de l'espace est conservée au voisinage de ce point. À l'inverse, l'orientation est inversée si le jacobien est négatif. Le jacobien d'une composée de

fonctions est le produit des jacobiens individuels. Le jacobien de la réciproque d'une fonction est l'inverse du jacobien de la fonction. Cette dernière propriété est liée au théorème d'inversion locale (qui peut être vu entre autre comme une extension du théorème des fonctions implicite en dimension supérieure à 1 dans le cas réel). Une fonction f de classe C^1 est inversible au voisinage de M avec une réciproque f^{-1} de classe C^1 si et seulement si son jacobien en M est non nul (théorème d'inversion locale). De plus, la matrice jacobienne de f^{-1} se déduit de l'inverse de la matrice jacobienne de f au moyen de la formule :

$$J_{f^{-1}} = (J_f \circ f^{-1})^{-1} \quad (12.19)$$

Théorème 48 — Théorème d'inversion locale. Soit f une application de U dans F , où U est un ouvert d'un espace de Banach réel et F un espace de Banach et soit x un point de U .

Si f est de classe C^p , avec p un entier strictement positif et si la différentielle de f au point x (définie au paragraphe 3.5) est un isomorphisme bicontinu, alors il existe un voisinage ouvert V de x et un voisinage ouvert W de $f(x)$ tels que f se restreigne en une bijection de V dans W dont la réciproque est de classe C^p .

Comme illustré au paragraphe précédent, le jacobien sert surtout pour effectuer des changement de variables dans le calculs des intégrales.

Théorème 49 — Théorème de changement de variables dans les intégrales multiples. Soient U un ouvert de \mathbb{R}^n , f une injection de classe C^1 de U dans \mathbb{R}^n et $V = f(U)$.

Si g est une fonction mesurable de V dans $[0; +\infty[$, on a égalité des intégrales pour la mesure de Lebesgue sur \mathbb{R}^n :

$$\int_V g(y_1, \dots, y_n) dy_1 \dots dy_n = \int_U g(f(x_1, \dots, x_n)) |\det J_f(x_1, \dots, x_n)| dx_1 \dots dx_n. \quad (12.20)$$

Si l'on considère un « petit » domaine, le volume de l'image de ce domaine par la fonction f sera celui du domaine de départ multiplié par la valeur absolue du jacobien.

12.5.3 Éléments de référence unidimensionnels linéaires à n nœuds

Il est possible de généraliser en considérant un segment ayant plusieurs nœuds intermédiaires comme indiqué sur la figure 12.9. Si l'on considère le cas général à n nœuds de la figure 12.10

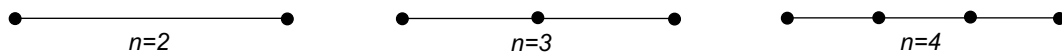


FIGURE 12.9: Nœuds intermédiaires

alors il vient :



FIGURE 12.10: Cas général

$$N_i(\xi) = \prod_{\substack{r=1 \\ r \neq i}}^n \frac{\xi_r - \xi}{\xi_r - \xi_i} \quad (12.21)$$

Si de plus, les n nœuds sont régulièrement espacés :

$$\xi_i = -1 + 2\frac{i-1}{n-1} \quad N_i(\xi) = \prod_{\substack{r=1 \\ r \neq i}}^n \frac{(2r-n-1) - \xi(n-1)}{2(r-i)} \quad (12.22)$$

12.5.4 Élément de référence unidimensionnel infini

Dans certains cas, il peut être nécessaire de prendre en compte des conditions aux limites situées à l'infini : problèmes de fondations, d'acoustique, de couplage fluide-structure.

Il est possible, comme aux paragraphes précédents, de proposer une transformation entre un élément de référence de longueur 2 et un élément réel infini, transformation illustrée sur la figure 12.11. Nous ne considérerons que le cas de l'élément unidimensionnel linéaire à 2 nœuds



FIGURE 12.11: Correspondance entre l'élément de référence et l'élément réel infini

dans ce paragraphe, mais on pourrait étudier le cas de l'élément linéaire unidimensionnel à n nœuds, ou des éléments bidimensionnels par exemple. Nous allons toujours avoir une interpolation de la fonction solution sous la forme :

$$u = \frac{1-\xi}{2}u_1 + \frac{1+\xi}{2}u_2 \quad (12.23)$$

car concrètement u_2 est connu (et fini) : c'est le « lieu » où l'on situe l'infini dans le modèle éléments finis. Par contre, l'interpolation du point courant sera donnée par :

$$x = x_1 + \frac{1+\xi}{1-\xi}\alpha \quad (12.24)$$

où α est une constante. On aura également :

$$u_{,x} = \frac{u_2 - u_1}{4\alpha}(1-\xi)^2 \quad (12.25)$$

12.5.5 Élément fini de barre unidimensionnel

On considère la barre de la figure 12.12, de section A , de longueur L , constituée d'un matériau homogène isotrope de module d'Young E et densité ρ et soumise uniquement à son poids propre. On souhaite discrétiser ce problème à l'aide d'un élément unidimensionnel linéaire tel que présenté

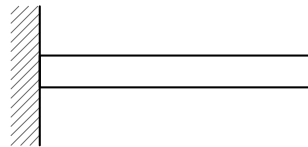


FIGURE 12.12: Barre encastree-libre

au paragraphe 12.5.1. La forme variationnelle de ce problème d'élasticité linéaire unidimensionnel (théorie des poutres) est :

$$W = \int_0^L EA v_{,x} u_{,x} dx - \int_0^L \rho g A v dx = 0, \quad \forall v \quad (12.26)$$

avec les conditions aux limites $u(0) = 0$ et $v(0) = 0$. Cette forme variationnelle s'écrit pour l'élément (ou pour chaque élément si on en avait plusieurs) :

$$W = {}^T \mathbf{v} (\mathbf{K} \mathbf{u} - \mathbf{f}) \quad (12.27)$$

Nous avons vu que la géométrie, représentée par un élément, est approximée par :

$$x = \frac{1-\xi}{2} x_1 + \frac{1+\xi}{2} x_2, \quad \xi \in [-1; 1] \quad (12.28)$$

La représentation de la fonction solution, est :

$$u = \frac{1-\xi}{2} u_1 + \frac{1+\xi}{2} u_2, \quad \xi \in [-1; 1] \quad (12.29)$$

Nous obtenons la matrice de rigidité élémentaire :

$$\mathbf{K} = \frac{2EA}{L} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad (12.30)$$

et le vecteur des forces nodales élémentaires :

$$\mathbf{f} = \rho g A \frac{L}{4} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (12.31)$$

En prenant en compte les conditions aux limites $u_1 = 0$ et $v_1 = 0$, l'expression $W = 0$ donne :

$$u_2 = \frac{\rho g L^2}{8E} \quad (12.32)$$

La déformation est donnée par $\varepsilon = u_{,x} = {}^T \mathbf{B} \mathbf{u}_n$ avec ${}^T \mathbf{B} = \frac{1}{L} (-1, 1)$. On obtient :

$$\varepsilon = \frac{\rho g L}{8E} \quad (12.33)$$

Elle est constante sur l'élément. Quant à la contrainte, elle est obtenue par $\sigma = E\varepsilon$, soit :

$$\sigma = \frac{\rho g L}{8} \quad (12.34)$$

Elle est également constante sur l'élément.

12.5.6 Assemblage de trois éléments unidimensionnels linéaires à deux nœuds

Poursuivons le cas de la poutre du paragraphe précédent, mais considérons une discrétisation de la barre à l'aide de trois éléments, comme montré sur la figure 12.13. Chaque élément e (de

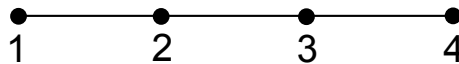


FIGURE 12.13: Discrétisation de la barre en trois éléments

longueur L_e) possède une matrice de rigidité élémentaire \mathbf{k}_e et un vecteur des forces nodales élémentaires \mathbf{f}_e définis par :

$$\mathbf{K}_e = \frac{2EA}{L_e} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad \text{et} \quad \mathbf{f}_e = \rho g A \frac{L_e}{4} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (12.35)$$

Les variables nodales sont ${}^T\mathbf{q} = {}^T\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4$. Les matrices élémentaires, issues de l'équation (12.35), s'écrivent en fonction de ces variables nodales :

$$\mathbf{K}_1 = \frac{2EA}{L_1} \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{K}_2 = \frac{2EA}{L_2} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{K}_3 = \frac{2EA}{L_3} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

et les forces nodales, quant à elles, s'expriment sous la forme :

$$\mathbf{f}_1 = \rho g A \frac{L_1}{4} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \mathbf{f}_2 = \rho g A \frac{L_2}{4} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \quad \mathbf{f}_3 = \rho g A \frac{L_3}{4} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

L'assemblage permettant de constituer le système complet est simplement obtenu par :

$$\mathbf{K} = \mathbf{K}_1 + \mathbf{K}_2 + \mathbf{K}_3 \quad \text{et} \quad \mathbf{F} = \mathbf{f}_1 + \mathbf{f}_2 + \mathbf{f}_3 \quad (12.36)$$

On résout alors $\mathbf{K}\mathbf{q} = \mathbf{F}$ avec la condition aux limites $u_1 = 0$.

Remarque. La matrice \mathbf{K} obtenue sur cet exemple a un caractère bande (la matrice est vide en dehors d'une zone centrée sur la diagonale). Ceci est dû à la numérotation des nœuds. Une autre numérotation pourrait faire perdre ce caractère important (stockage et résolution). C'est pourquoi les programmes éléments finis peuvent comporter une étape de renumérotation automatique des nœuds.

12.5.7 Élément de barre unidimensionnel de type Hermite (élément subparamétrique)

En restant encore sur notre élément linéaire unidimensionnel, nous allons voir comment ajouter des contraintes sur les dérivées aux nœuds, comme illustré à la figure 12.14. La géométrie est encore

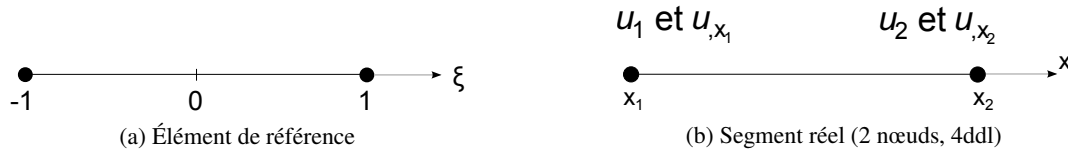


FIGURE 12.14: Élément de barre unidimensionnel de type Hermite

une fois approximée par :

$$x = \frac{1-\xi}{2}x_1 + \frac{1+\xi}{2}x_2, \quad \xi \in [-1; 1] \quad (12.37)$$

Mais cette fois, l'approximation de la fonction solution est voulue sous la forme :

$$u = {}^T(N_1, N_2, N_3, N_4) \mathbf{u}_n \quad \text{avec les ddl} \quad {}^T\mathbf{u}_n = {}^T(u_1, u_{,x_1}, u_2, u_{,x_2}) \quad (12.38)$$

Les fonctions N_i choisies sont des fonctions cubiques de type Hermite assurant une continuité de u et $u_{,x}$ aux nœuds 1 et 2. On a :

$$N_1 = \frac{1}{4}(1-\xi)^2(2+\xi); \quad N_2 = \frac{L}{8}(\xi^2-1)(1-\xi); \quad N_3 = \frac{1}{4}(1+\xi)^2(2-\xi); \quad N_4 = \frac{L}{8}(\xi^2-1)(1+\xi) \quad (12.39)$$

Le champ $u_{,x}$, qui est la déformation ε , est donné par $\varepsilon = {}^T\mathbf{B}\{u_n\}$ avec :

$${}^T\mathbf{B} = {}^T \left(\frac{3}{2L}(\xi^2-1), \quad \frac{1}{4}(3\xi^2-2\xi-1), \quad \frac{3}{2L}(1-\xi^2), \quad \frac{1}{4}(3\xi^2+2\xi-1) \right) \quad (12.40)$$

La contrainte, dans le cas où le coefficient d'élasticité H est constant est donnée par :

$$\sigma_x = H {}^T\mathbf{B} \mathbf{u}_n \quad (12.41)$$

12.5.8 Élément mixte unidimensionnel

Continuons avec notre élément de référence linéaire unidimensionnel. Cette fois-ci nous souhaitons le mettre en relation avec un segment réel dont les inconnues nodales sont les déplacements u_1 et u_2 , mais qui possède en plus une approximation constante de la contrainte par élément, comme montré sur la figure 12.15. Nous cherchons donc une approximation C^0 du déplacement et C^{-1} de

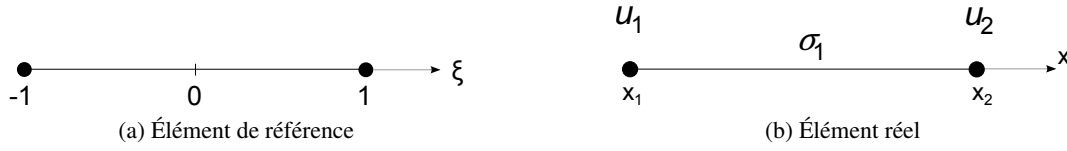


FIGURE 12.15: Élément mixte unidimensionnel

la contrainte. Pour la géométrie, nous avons une fois encore :

$$x = \frac{1-\xi}{2}x_1 + \frac{1+\xi}{2}x_2, \quad \xi \in [-1; 1] \quad (12.42)$$

Pour la contrainte, nous avons :

$$\sigma_x(x) = \sigma_1 \text{ (constante)} \quad (12.43)$$

La fonctionnelle d'Hellinger-Reissner sous sa deuxième forme (ou fonctionnelle mixte), donnée dans sa formulation générale par l'équation (8.52), s'écrit, dans le cas de l'élasticité linéaire unidimensionnelle :

$$W = \int_0^L A \left(-\sigma_x^* \frac{1}{H} \sigma_x + \sigma_x^* u_{,x} + v_{,x} \sigma_x \right) dx - \int_0^L A v f_x dx - (v f_S)_{S_f} = 0 \quad \forall v, \sigma^* \quad (12.44)$$

avec les conditions aux limites $u = \bar{u}$ et $u^* = 0$ sur S_u . De manière discrétisée, il vient :

$$W = {}^T(v_1, v_2, \sigma_1^*) \left(\begin{bmatrix} \mathbb{O} & \mathbf{b} \\ {}^T\mathbf{b} & -a \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \sigma_1 \end{pmatrix} - \begin{pmatrix} \mathbf{f}_n \\ \sigma_1 \end{pmatrix} \right) \quad (12.45)$$

avec :

$$\mathbf{b} = \int_0^L \frac{A}{L} \begin{pmatrix} -1 \\ 1 \end{pmatrix}; \quad a = \int_0^L \frac{A}{H} dx; \quad \mathbf{f}_n = \int_{-1}^{+1} A \begin{pmatrix} N_1 \\ N_2 \end{pmatrix} \frac{L}{2} f_x d\xi = A f_x \frac{L}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (12.46)$$

On pourrait s'arrêter là avec le système précédent à résoudre. Toutefois, σ_1 est une variable locale sans couplage avec les autres éléments. On peut donc l'exprimer en fonction des autres degrés de liberté de l'élément u_1 et u_2 . La dernière ligne du système donne : $\sigma_1^* = ({}^T\mathbf{b}\mathbf{u}_n - a\sigma_1) = 0, \forall \sigma_1^*$, d'où :

$$\sigma_1 = \frac{1}{a} {}^T\mathbf{b}\mathbf{u}_n \quad (12.47)$$

Le système devient alors :

$$W = {}^T\mathbf{v}_n (\mathbf{b}\sigma_1 - \mathbf{f}_n) = {}^T\mathbf{v}_n (\mathbf{K}\mathbf{u}_n - \mathbf{f}_n) \quad (12.48)$$

avec :

$$\mathbf{K} = \mathbf{b} \frac{1}{a} {}^T\mathbf{b} \quad (12.49)$$

Dans le cas où H est constant, la matrice de rigidité obtenue est identique à celle du modèle en déplacements du paragraphe 12.5.5.

12.6 Sur les déplacements imposés

12.6.1 Problème considéré

Nous repartons de l'élément de barre 1D défini au paragraphe 12.5.5, et nous rappelons que la matrice de rigidité élémentaire et le vecteur des forces nodales élémentaires sont :

$$\mathbf{K} = \frac{2EA}{L} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad \text{et} \quad \mathbf{f} = \rho g A \frac{L}{4} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (12.50)$$

L'assemblage de trois éléments, décrit au paragraphe 12.5.6, conduit à :

$$\mathbf{K} = \mathbf{K}_1 + \mathbf{K}_2 + \mathbf{K}_3 \quad \text{et} \quad \mathbf{F} = \mathbf{f}_1 + \mathbf{f}_2 + \mathbf{f}_3 \quad (12.51)$$

On traitera le cas où les trois éléments ont la même longueur ($L_1 = L_2 = L_3 = L$) et où seul le nœud 4 est soumis à une force. On obtient alors un système de type $\mathbf{K}\mathbf{q} = \mathbf{F}$ à résoudre, avec la condition aux limites $u_1 = 0$, qui s'écrit explicitement dans ce cas :

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ F' \end{pmatrix} \quad (12.52)$$

(avec $F' = \rho g L^2 / (8E)$).

12.6.2 Retour sur la résolution de systèmes linéaires

Soit à résoudre un système linéaire de type :

$$\mathbf{Ax} = \mathbf{b} \quad (12.53)$$

Ce système peut être réécrit par blocs sous la forme :

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad (12.54)$$

Si A_{11} est inversible, alors on peut écrire la première ligne :

$$x_1 = A_{11}^{-1} (b_1 - A_{12}x_2) \quad (12.55)$$

ce qui donne, dans la deuxième ligne :

$$A_{22} - A_{21}A_{11}^{-1}A_{12}x_2 = b_2 - A_{21}A_{11}^{-1}b_1 \quad \text{que l'on note : } Sx_2 = c \quad (12.56)$$

La matrice S est appelée complément de Schur (présenté au paragraphe 18.5.1), ou matrice condensée sur les degrés de liberté x_2 . Ce calcul est également identique à celui présenté à propos de la condensation statique au paragraphe 13.5.1. Dans ce cas, la matrice S est la matrice de rigidité du super-élément considéré.

Le système initial (12.53) est équivalent à :

$$\begin{bmatrix} A_{11} & A_{12} \\ \mathbb{O} & S \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ c \end{pmatrix} \quad (12.57)$$

qui peut être réécrit :

$$\begin{bmatrix} \mathbf{I} & \mathbb{O} \\ A_{21}A_{11}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ \mathbb{O} & S \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} \mathbf{I} & \mathbb{O} \\ A_{21}A_{11}^{-1} & \mathbf{I} \end{bmatrix} \begin{pmatrix} b_1 \\ c \end{pmatrix} \quad (12.58)$$

12.6.3 Complément de Schur et déplacements imposés

Nous cherchons toujours à résoudre notre système :

$$\mathbf{K}\mathbf{q} = \mathbf{F} \quad (12.59)$$

Considérons la matrice \mathbf{D} , matrice booléenne, permettant d'extraire de l'ensemble des inconnues nodales \mathbf{q} le sous-ensemble $\mathbf{q}_2 = \mathbf{D}\mathbf{q}$ sur lequel portent les conditions de déplacements imposés \mathbf{q}_d (ici on traite le cas général, i.e. les déplacements imposés peuvent être nuls ou non). La matrice \mathbf{D} n'est pas rectangulaire : si on souhaite imposer d déplacements parmi les n degrés de liberté du système, alors \mathbf{D} est de dimension $d \times n$, et donc \mathbf{q}_2 est bien un vecteur à d composantes, comme \mathbf{u}_d .

Le système $\mathbf{K}\mathbf{q} = \mathbf{F}$ peut alors s'écrire, par blocs, en utilisant le complément de Schur (12.57) sous une forme triangulaire d'où l'on tire le sous-système correspondant aux déplacements non imposés :

$$\mathbf{K}_{11}\mathbf{q}_1 = \mathbf{f}_1 - \mathbf{K}_{12}\mathbf{u}_d \quad (12.60)$$

Ce système est réduit, i.e. possède moins d'inconnues que le problème initial, mais il est nécessaire de modifier le chargement extérieur.

Son inconvénient est qu'il est nécessaire d'effectuer un tri explicite au sein des degrés de liberté, ce qui a pour conséquence de « remplir » le terme \mathbf{K}_{12} , et par suite peut générer un surcoût de calcul lorsque le nombre de degrés de liberté bloqués est grand.

C'est ainsi que procède le code ABAQUS pour imposer les déplacements.

12.6.4 Multiplicateurs de Lagrange et déplacements imposés

Une autre technique, pour imposer des déplacements, consiste à utiliser les multiplicateurs de Lagrange : les contraintes sont relaxées, puis réintroduites via des multiplicateurs de Lagrange. Dans ce cas, il n'est plus nécessaire de séparer explicitement les degrés de liberté, mais la taille du problème augmente puisqu'il faut lui adjoindre les multiplicateurs. Le système à résoudre s'écrit très simplement :

$$\begin{bmatrix} \mathbf{K} & -\mathbf{D}^T \\ -\mathbf{D} & \mathbf{0} \end{bmatrix} \begin{pmatrix} \mathbf{q} \\ \lambda \end{pmatrix} = [\mathbf{A}] \begin{pmatrix} \mathbf{q} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{F} \\ -\mathbf{u}_d \end{pmatrix} \quad (12.61)$$

Il est intéressant de remarquer que la première ligne montre que les multiplicateurs de Lagrange sont bien les réactions aux appuis. Par contre, la matrice du système (12.61), \mathbf{A} , si elle reste bien symétrique, n'est plus définie positive. Elle possède des valeurs propres négatives. Toutefois, on dispose du résultat suivant :

Théorème 50 si \mathbf{K} est symétrique définie positive, alors :

$$\mathbf{A} \text{ inversible} \iff \mathbf{D} \text{ injective} \iff \ker \mathbf{D} = \emptyset \quad (12.62)$$

ce qui correspond aux cas où toutes les liaisons sont indépendantes.

Afin de ne pas détruire la structure bande de \mathbf{K} (ce que ferait une factorisation de Gauß), et d'éviter des pivotages (que ferait une factorisation de Crout, i.e. $\mathbf{LD}^T\mathbf{L}$), on peut développer une autre technique, dite technique de double multiplicateur, qui est celle employée dans CAST3M :

$$\begin{cases} \mathbf{K}\mathbf{q} = \mathbf{F} - \mathbf{D}^T(\lambda_1 - \lambda_2) \\ \lambda_1 = \lambda_2 \\ \mathbf{D}\mathbf{q} = \mathbf{u}_d \end{cases} \quad (12.63)$$

L'inconvénient reste l'augmentation de la taille du système lorsque l'on a de nombreux blocages.

12.6.5 Actions extérieures et déplacements imposés

L'interprétation physique des multiplicateurs de Lagrange conduit naturellement à une autre méthode : il est possible de considérer un déplacement imposé comme une action extérieure, par exemple comme la réaction d'un ressort ayant une raideur k très grande et un déplacement imposé à sa base. On se retrouve alors avec un système du type :

$$(\mathbf{K} + {}^T\mathbf{D}k\mathbf{D})\mathbf{q} = \mathbf{F} + {}^T\mathbf{D}k\mathbf{D}\mathbf{u}_d \quad (12.64)$$

dans lequel on n'a fait qu'ajouter des termes à la matrice de rigidité sur les degrés de liberté correspondant à ces déplacements et au vecteur des forces généralisées sur les efforts duaux.

Le problème réside dans le choix de la valeur de k : trop petite, la condition est mal imposée ; trop grande, le système devient très mal conditionné, voire numériquement singulier.

12.6.6 Retour sur notre exemple

Nous devons toujours résoudre :

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ F' \end{pmatrix} \quad (12.65)$$

avec la condition aux limites $u_1 = 0$. Tout d'abord, on pourra essayer de résoudre sans introduire de conditions aux limites afin de voir que le système est alors singulier. On obtient $u_1 = u_2 = u_3 = u_4$ et $F' = 0$. La méthode la plus rudimentaire consiste à prendre en compte directement et explicitement la condition aux limites. On supprime donc du système matriciel la ligne et la colonne correspondant à u_1 , et on résout :

$$u_1 = 0 \text{ et } \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{pmatrix} u_2 \\ u_3 \\ u_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ F' \end{pmatrix} \quad \text{d'où : } \begin{cases} u_1 = 0 \\ u_2 = F' \\ u_3 = 2F' \\ u_4 = 3F' \end{cases} \quad (12.66)$$

Séparons maintenant les déplacements imposés des autres degrés de liberté, par une matrice booléenne \mathbf{D} : \mathbf{D} est une matrice 1×4 , et seul le terme $D_{11} = 1$ est non nul : $\mathbf{q}_2 = \mathbf{D}\mathbf{q} = \mathbf{u}_1$, et le déplacement imposé est $\mathbf{u}_d = \mathbf{0}$. Le système s'écrit :

$$\begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix} \begin{pmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{pmatrix} \quad (12.67)$$

avec :

$$\begin{bmatrix} \mathbf{K}_{11} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} & \mathbf{K}_{12} = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \\ \mathbf{K}_{21} = \begin{bmatrix} -1 & 0 & 0 \end{bmatrix} & \mathbf{K}_{22} = \begin{bmatrix} 1 \end{bmatrix} \end{bmatrix} \begin{pmatrix} \mathbf{q}_1 = \begin{pmatrix} u_2 \\ u_3 \\ u_4 \end{pmatrix} \\ \mathbf{q}_2 = (u_1) \end{pmatrix} = \begin{pmatrix} \mathbf{f}_1 = \begin{pmatrix} 0 \\ 0 \\ F' \end{pmatrix} \\ \mathbf{f}_2 = (u_d = 0) \end{pmatrix}$$

Par la méthode du complément de Schur, on est encore ramené à la résolution du système (12.66). On pourra s'amuser à calculer \mathbf{K}_{11}^{-1} , \mathbf{S} ... Par la méthode des multiplicateurs de Lagrange, on obtient le système :

$$\begin{bmatrix} 1 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ F' \\ 0 \end{pmatrix} \quad \text{d'où : } \begin{cases} u_1 = 0 \\ u_2 = F' \\ u_3 = 2F' \\ u_4 = 3F' \\ \lambda = -F' \end{cases} \quad (12.68)$$

Par la technique du double multiplicateur, on obtient le système :

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 1 & -1 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ F' \\ 0 \\ u_d = 0 \end{pmatrix} \quad \text{d'où : } \begin{cases} u_1 = 0 \\ u_2 = F' \\ u_3 = 2F' \\ u_4 = 3F' \\ \lambda_1 = -F' \\ \lambda_2 = -F' \end{cases} \quad (12.69)$$

En considérant une action extérieure, on obtient le système :

$$\begin{bmatrix} 1+k & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix} = \begin{pmatrix} 0+k \cdot u_d = 0 \\ 0 \\ 0 \\ F' \end{pmatrix} \quad \text{d'où : } \begin{cases} u_1 = 0 \\ u_2 = F' \\ u_3 = 2F' \\ u_4 = 3F' \end{cases} \quad (12.70)$$

Dans tous les exemples présentés, nous avons résolu les systèmes à la main, ce qui masque d'éventuels problèmes numériques (notamment dans le dernier cas).

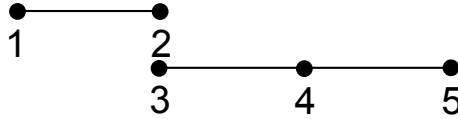
12.6.7 Relations linéaires entre degrés de liberté

Cette situation se présente par exemple lors de la prise en compte de conditions de symétrie, de conditions aux limites périodiques... Il s'agit à chaque fois de conditions de type :

$$\mathbf{D}\mathbf{q} = \mathbf{u}_d \quad (12.71)$$

où \mathbf{D} n'est plus forcément booléenne.

Nous allons l'appliquer ici, à titre d'exemple simple, au cas où la poutre considérée est maintenant constituée de deux poutres, que l'on va « coller » par une telle relation. La poutre est constituée



de la poutre 1 – 2 et de la poutre 3 – 4 – 5, les points 2 et 3 étant astreints à rester collés ensemble. En ajoutant la relation $u_2 = u_3$ par l'intermédiaire d'un multiplicateur de Lagrange, c'est-à-dire en ajoutant le terme $\lambda_1(u_2 - u_3)$, on obtient le système :

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 & 0 & -1 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ \lambda_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ F' \\ 0 \\ u_d = 0 \end{pmatrix} \quad (12.72)$$

auquel il faut également ajouter la condition aux limites $u_1=0$. Si cette condition aux limites est imposée par un multiplicateur de Lagrange λ_2 , alors finalement, il faut résoudre :

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & -1 \\ -1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ F' \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{d'où : } \begin{cases} u_1 = 0 \\ u_2 = F' \\ u_3 = F' \\ u_4 = 2F' \\ u_5 = 3F' \\ \lambda_1 = -F' \\ \lambda_2 = -F' \end{cases} \quad (12.73)$$

Chapitre 13

Calcul efficient : qualité des résultats et efforts de calcul

13.1 Amélioration d'un modèle : méthodes r , h et p

Les indicateurs d'erreur ont pour but de nous indiquer si « nous sommes loin de la solution », afin de pouvoir modifier le modèle si besoin. Ainsi, par itérations successives, on pourra tendre vers une solution de plus en plus proche de la solution exacte.

Les indicateurs locaux nous permettent notamment de faire des modifications uniquement locales du modèle, i.e. uniquement là où il y en a besoin, par exemple en raffinant le maillage.

Globalement, il existe trois stratégies pour améliorer la précision de la solution obtenue :

Méthode r : Pour un maillage et un type d'élément donné, il s'agit de déplacer les nœuds, en fonction des indicateurs d'erreur, et donc sans impacter le nombre de degrés de liberté du système. Ainsi, la taille des éléments peut augmenter (maillage plus grossier) dans les zones les moins sollicitées et diminuer (maillage plus fin) dans les zones les plus sollicitées. On voit que la restriction de la méthode est qu'elle ne joue pas sur le nombre de nœuds, et que par conséquent, sans violer les contraintes de distorsion, elle est limitée.

Méthode h : En conservant le même type d'élément, on les subdivise dans les zones les plus sollicitées selon le ou les indicateurs d'erreur choisis. Dans cette méthode, on augmente le nombre de nœuds afin de contrôler les erreurs et la précision du modèle. Il est nécessaire de se fixer une limite dans la précision recherchée afin de ne pas trop raffiner le maillage.

Méthode p : À nombre d'éléments constant, dans les zones les plus sollicitées, on va modifier les éléments en introduisant des fonctions de formes polynomiales d'ordre plus élevé, dites hiérarchiques. La complexité du système est cette fois encore accrue.

Évidemment, ces méthodes peuvent être combinées. La méthode $-hp$ - propose de modifier à la fois le maillage et les fonctions d'interpolation. Elle semble aujourd'hui être la méthode optimale en terme d'efficacité et de vitesse de convergence.

Notons que les méthodes $-r$ - et $-h$ - dépendent des capacités du mailleur automatique implémenté. De nombreux travaux existent sur le maillage automatique, nous n'entrerons pas dans ce détail.

13.2 Post-traitement

Une manière d'améliorer les résultats est de recourir à des méthodes dites de post-traitement, i.e. des méthodes qui, à partir des données issues de la résolution du système matriciel correspondant au problème, fournissent des données complémentaires ou améliorent toute ou partie des données déjà disponibles.

En fonction des problèmes (donc des données disponibles et des données souhaitées) de nombreuses méthodes existent. Elles sont généralement dédiées à un problème donné.

Un exemple déjà abordé dans ce document est celui où, à partir des déplacements nodaux obtenus par un calcul éléments finis d'une structure mécanique composée de deux matériaux différents dans une discrétisation classique « en déplacements », on souhaite remonter aux contraintes à l'interface entre lesdits matériaux.

Une méthode déjà mentionnée est la méthode dite de « Reissner local » qui consiste à intégrer les équations d'équilibre sous la forme mixte de Reissner, mais uniquement sur des paires d'éléments ayant des propriétés matérielles différentes et possédant une face commune.

On obtient alors un champ de contrainte complètement continu dont on ne retient que les composantes correspondant à la trace des contraintes, les autres étant calculées comme dans la méthode classique. On montre que l'on améliore grandement la qualité de l'approximation des contraintes aux interfaces, mêmes avec un maillage grossier.

Un autre exemple serait, disposant d'un point de pression constant par élément, de calculer la pression en un point quelconque comme interpolation (linéaire ou non) à partir des points disponibles.

Un troisième exemple serait à partir des données nodales dans un modèle unidimensionnel (poutre ou barre selon le cas), de remonter à la répartition des contraintes en un point quelconque de la structure (donc via les hypothèses faites dans le modèle sur la répartition des contraintes dans une section + via une interpolation lorsque l'on se trouve dans une section ne passant pas par un nœud)...

13.3 Exemple d'implémentation d'un post-traitement dans ANSYS

Dans ce chapitre, nous présentons une manière d'implémenter la méthode « Reissner local » mentionnée aux paragraphes 13.2 et 12.3. Il s'agit de calculer les contraintes à l'interface entre deux matériaux en utilisant la fonctionnelle mixte d'Hellinger-Reissner donnée à l'équation (8.52).

Le but est de montrer que même sur un maillage grossier, il est possible de correctement estimer les contraintes aux interfaces, pour peu que l'on dispose d'une méthode appropriée. Nous espérons qu'au passage, nous démontrerons également qu'il est assez simple d'implémenter des fonctions dans un code existant.

13.3.1 Macro dans ANSYS

Si un code ne dispose pas de la méthode que l'on souhaite utiliser, il est toujours possible de retraiter les résultats... toutefois, selon les codes, il est plus ou moins aisé d'implémenter des fonctions dans ledit code.

ANSYS est souvent qualifié de « code industriel », ce qui pourrait laisser supposer qu'il n'a pas la même capacité à résoudre les problèmes que des codes dits « de recherche » ou « spécialisés ». Or il n'en est rien (même si en toute rigueur, il y a une quinzaine d'années il était plus faible que d'autres sur les non linéarités, ce qui n'est plus vrai depuis longtemps).

C'est vrai que disposant d'une interface utilisateur remarquable (comparée à celle de nombreux autres codes), il est possible de ne réaliser les calculs que via cette interface, i.e. sans passer par un fichier batch... ce que ne font de toutes façons pas les gens sérieux (donc cet argument ne tient pas).

Le qualificatif d'industriel peut également laisser penser que le logiciel est plus fermé que d'autres, mais ce n'est que partiellement le cas, en tous les cas on peut y remédier facilement, et c'est pourquoi nous allons présenter l'implémentation d'une macro dans ce code.

Pour implémenter une nouvelle fonction dans ANSYS, plusieurs méthodes s'offrent à nous. Nous citerons :

- écrire le bout de programme et recompiler le noyau : c'est faisable en théorie, mais personnellement je n'ai jamais réussi...

- partir des fichiers de résultats binaires pour les retravailler : c'est également théoriquement faisable, mais il faut décortiquer les formats d'écriture desdits fichiers, les relire, réécrire dans le même format... et c'est donc beaucoup de travail (purements informatique) ;
- programmer la fonction directement sous ANSYS (ANSYS possède un langage assez riche et puissant) ;
- ou alors, et c'est la voie que nous allons montrer, utiliser ANSYS en « coopération » avec un petit programme extérieur.

La structure de la macro ANSYS que nous proposons (et qui peut être également incluse dans les menus d'ANSYS, mais nous n'entrons pas dans ce niveau de détail d'interfaçage) est très simple :

- on récupère les données dont nous avons besoin et on les stocke dans des fichiers externes (simples) ;
- on lance un programme externe (dont la structure sera exposée plus bas) ;
- on réintègre les résultats dans ANSYS.

On peut alors se servir de toutes les fonctions de visualisation disponibles dans ANSYS avec les résultats modifiés...

```

1 /nop
2 nall
3 !
4 ! Parametres
5 *get,Nnoeuds,node,,count
6 *get,Nmater,mat,,count
7 *get,Nelem,elem,,count
8 *cfopen,temp,par
9 *vwrite,Nnoeuds
10 (E13.7,' ')
11 *vwrite,Nelem
12 (E13.7,' ')
13 *vwrite,Nmater
14 (E13.7,' ')
15 *cfclose
16 !
17 ! Matériaux
18 *cfopen,temp,mat
19 *dim,Mater,,Nmater,3
20 *do,I,1,Nmater
21   *get,Mater(I,1),Ex,I
22   *get,Mater(I,2),Ey,I
23   *get,Mater(I,3),Nuxy,I
24 *enddo
25 *vwrite,Mater(1,1),Mater(1,2),Mater
    (1,3)
26 (3(E13.7,' '))
27 *cfclose
28 !
29 ! Noeuds
30 *dim,CoordX,,Nnoeuds
31 *dim,CoordY,,Nnoeuds
32 *dim,CoordZ,,Nnoeuds
33 *vget,CoordX(1),node,,loc,x
34 *vget,CoordY(1),node,,loc,y
35 *cfopen,temp,coo
36 *vwrite,CoordX(1),CoordY(1)
37 (2(E13.7,' '))
38 *cfclose
39 !
40 ! Elements + numero du materiau
41 *dim,E1,,Nelem,9
42 *vget,E1(1,1),elem,1,node,1
43 *vget,E1(1,2),elem,1,node,2
44 *vget,E1(1,3),elem,1,node,3
45 *vget,E1(1,4),elem,1,node,4
46 *vget,E1(1,5),elem,1,node,5
47 *vget,E1(1,6),elem,1,node,6
48 *vget,E1(1,7),elem,1,node,7
49 *vget,E1(1,8),elem,1,node,8
50 *vget,E1(1,9),elem,1,attr,mat
51 *cfopen,temp,elt
52 *vwrite,E1(1,1),E1(1,2),E1(1,3),E1
    (1,4),E1(1,5),E1(1,6),E1(1,7),E1
    (1,8),E1(1,9)
53 (9(E13.7,' '))
54 *cfclose
55 !
56 ! Resultats
57 *vget,CoordX(1),node,1,U,x
58 *vget,CoordY(1),node,1,U,y
59 *cfopen,temp,res
60 *vwrite,CoordX(1),CoordY(1)
61 (2(E13.7,' '))
62 *cfclose
63 !
64 ! Contraintes
65 *vget,CoordX(1),node,1,S,x
66 *vget,CoordY(1),node,1,S,y
67 *vget,CoordZ(1),node,1,S,xy
68 *cfopen,temp,str
69 *vwrite,CoordX(1),CoordY(1),CoordZ(1)
70 (3(E13.7,' '))
71 *cfclose
72 !
73 ! run locreiss.exe
74 /sys,'locreiss.exe'
75 !
76 ! Effacer les fichiers
77 ! generes par ANSYS:
78 !   - temp.par: parametres
79 !   - temp.coo: coordonnees nodales
80 !   - temp.elt: incidences
81 !   - temp.mat: proprietes materielles
82 !   - temp.res: déplacements nodaux
83 !   - temp.str: contraintes nodales
84 ! generes par locreiss.exe:
85 !   - temp.dum
86 !   - temp.tmp
87 ! Il reste les fichiers
88 !   - temp.sij: contraintes nodales
    modifiees
89 !
90 !/sys,'rm ./temp.par'
91 !/sys,'rm ./temp.coo'
92 !/sys,'rm temp.elt'
93 !/sys,'rm temp.mat'
94 !/sys,'rm temp.res'
95 !/sys,'rm temp.str'

```

```

96 !/sys , 'rm temp.dum'
97 !/sys , 'rm temp.tmp'
98 !
99 ! lire les valeurs dans temp.sij
100 ! ces fichiers contiennent: sxx, syx,
    sxy
101 ! ce sont toutes les composantes
    continues.
102 ! Elles n'ont pas toutes un sens:
103 ! Sxx discontinu, mais dans le repere
104 ! local, donc on la laisse pour qu'
    ANSYS
105 ! puisse faire la ! rotation de repere
106 ! si necessaire.
107 *vread,CoordX(1),temp,sxx,,
108 (E13.7)
109 *vread,CoordY(1),temp,syx,,
110 (E13.7)
111 *vread,CoordZ(1),temp,sxy,,
112 (E13.7)
113 *vput,CoordX(1),node,1,S,x
114 *vput,CoordY(1),node,1,S,y
115 *vput,CoordZ(1),node,1,S,xy
116 !
117 ! Cleanup
118 CoordX(1)=
119 CoordY(1)=
120 CoordZ(1)=
121 Mater(1,3)=
122 E1(1,9)=
123 Nnoeuds=
124 Nmater=
125 Nelem=
126 !
127 /gop

```

Le programme extérieur locreiss.exe, est structuré comme suit :

```

1  PROGRAM VM2
2  C
3  IMPLICIT DOUBLE PRECISION (A-H,O-Z)
4  IMPLICIT INTEGER (I-N)
5  C
6  C
7  C But:
8  C 1. on relit les fichiers de ANSYS generes avec la macro INTERF
9  C 2. on determine les noeuds de post-traitement
10 C 3. on effectue un Reissner local aux interfaces
11 C
12 C WARNING:
13 C On ne traite que le cas de PLANE 82 avec une geometrie definie
14 C dans le plan (X,Y)
15 C
16 C MAIN of: VM2
17 C
18 C
19 C PARAMETER(NODMAX=1000,NELTMAX=300,MATMAX=4)
20 C
21 C COMMON /VMCTS/ COMPI,ZERO,ONE,IINP,IOUT,ICON,MONI
22 C COMMON /NPB/ NNOD,NELT,NMAT,NCAL
23 C
24 C
25 C DIMENSION COORD(NODMAX,3),DISP(NODMAX,2)
26 C DIMENSION XMAT(MATMAX,3),NLT(NELTMAX,9)
27 C DIMENSION NCALC(NODMAX,5),NADJ(NODMAX,8),NDT(NODMAX,8)
28 C DIMENSION D1(3,3),D2(3,3)
29 C
30 C Presetings
31 C MONI=0 ! monitoring level (debug)
32 C ICON=6 ! standard output = screen
33 C IINP=10 ! input file
34 C IOU=20 ! output file
35 C ZERO=0.0D0 ! zero
36 C ONE=1.0D0 ! one
37 C COMPI=DACOS(-ONE) ! pi
38 C
39 C
40 C Read input files
41 C CALL READANS(NODMAX,NELTMAX,MATMAX,COORD,XMAT,NLT,DISP)
42 C
43 C Find Adjacent elements
44 C CALL CHELT(NODMAX,NELTMAX,NLT,NCALC,NADJ,NDT)
45 C
46 C Material properties
47 C CALL MATER(MATMAX,XMAT,D1,D2,D3,D4,NODMAX,NADJ,NELTMAX,NLT)
48 C
49 C Computation: Local Reissner
50 C CALL LOCREISS(NODMAX,NELTMAX,D1,D2,D3,D4,
51 C x COORD,NLT,NADJ,DISP)
52 C

```



```

53 C—— Average multi-computed nodes
54   CALL AVERAGE(NODMAX,MULTI)
55 C
56 C—— Last line of VM2
57   END

```

La routine READANS permet de lire les données. Celles-ci sont stockées dans les matrices COORD(NODMAX,3), DISP(NODMAX,2), XMAT(MATMAX,3) et NLT(NELTMAX,9).

la routine CHELT détecte les interfaces, i.e. les faces dont les nœuds appartiennent à des éléments dont les propriétés matérielles sont différentes. Le vecteur NADJ(NODMAX,8) contient les numéros des deux éléments adjacents (positions 1 et 2) ainsi que les numéros des nœuds de l'interface (positions 3, 4 et 5).

La routine MATER construit, pour chacune des faces de l'interface, la matrice de rigidité (aussi bien pour le cas isotrope que pour le cas orthotrope).

La routine LOCREISS calcule les contraintes nodales le long de chaque face.

Les contraintes ayant été calculées pour les nœuds de chacune des faces de l'interface, on moyenne les résultats pour un nœud appartenant à plusieurs faces. C'est ce que fait la routine AVERAGE. À ce stade, nous disposons donc des contraintes nodales, calculées par la fonctionnelle mixte d'Hellinger-Reissner, le long de toutes les interfaces présentes dans le modèle. Il ne reste plus qu'à écrire ces résultats pour qu'ils soient relus par la macros ANSYS.

13.3.2 Poutre en U

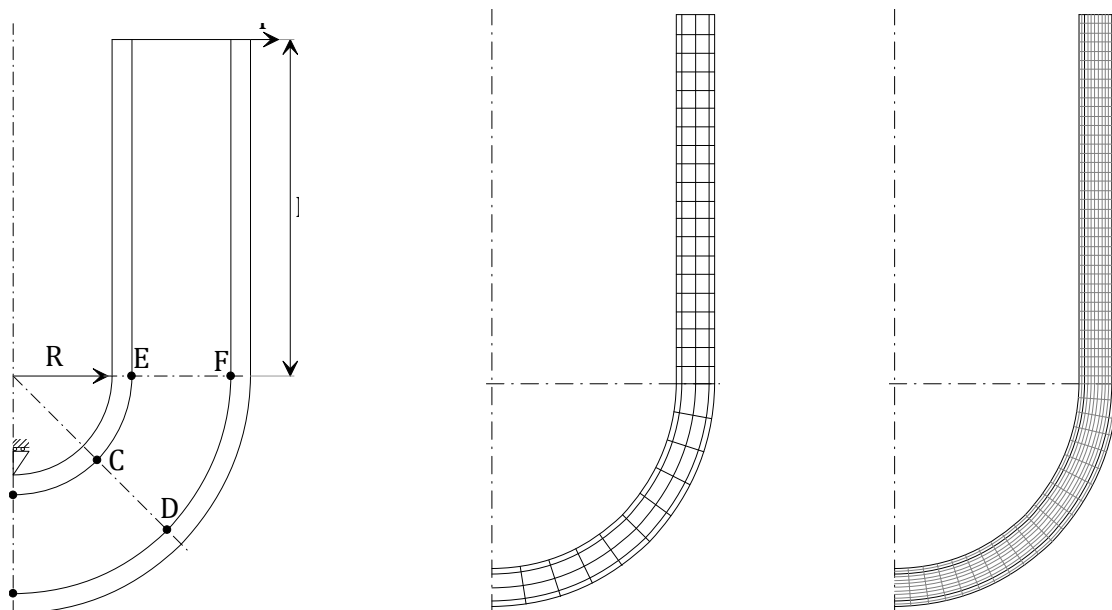


FIGURE 13.1: a) Géométrie de la poutre en U, b) le maillage utilisé pour le tester la méthode Reissner local, c) maillage permettant d'obtenir la « solution de référence » puisqu'il n'y a pas de solution analytique

Disposant de la macro précédente, nous allons l'appliquer au calcul de la poutre en U présentée à la figure 13.1a. Il s'agit d'une poutre sandwich dont les peaux sont en aluminium et l'âme en résine époxy. Par symétrie, seule une demi-poutre est modélisée. Elle est soumise à une force unitaire sur chaque montant. Les points pour lesquels nous nous intéresserons aux contraintes sont les points indiqués sur la figure 13.1a, à savoir les points A, B, C, D, E et F.

Le listing ANSYS est le suivant :

```

1  !* geom: partie droite
2  long=20
3  H1=0.2
4  H2=1.8
5  H3=2
6  !*
7  !* geom coude
8  Rint=10
9  Rh1=Rint+H1
10 Rh2=Rint+H2
11 Rh3=Rint+H3
12 !*
13 !* maillage
14 NCUTS=40
15 NRCUTS=20
16 NCORE=16
17 NSKIN=2
18 Forc=1
19 !*
20 /PREP7
21 !*
22 ET,1,PLANE82
23 !*
24 UIMP,1,EX,, ,70000,
25 UIMP,1,NUXY,, ,0.3,
26 UIMP,1,EMIS,, ,1,
27 !*
28 !*
29 UIMP,2,EX,, ,3400,
30 UIMP,2,NUXY,, ,0.34,
31 UIMP,2,EMIS,, ,1,
32 !*
33 k,1,0,-Rh3
34 k,2,Rh3,0
35 k,3,Rh2,0
36 k,4,0,-Rh2
37 k,5,0,-Rh1
38 k,6,Rh1,0
39 k,7,Rint,0
40 k,8,0,-Rint
41 k,9,Rh3,long
42 k,10,Rh2,long
43 k,11,Rh1,long
44 k,12,Rint,long
45 k,13,0,0
46 !*
47 larc,1,2,13,Rh3,NRCUTS
48 1,2,3,NSKIN
49 larc,4,3,13,Rh2,NRCUTS
50 1,1,4,NSKIN
51 1,4,5,NCORE
52 1,3,6,NCORE
53 larc,5,6,13,Rh1,NRCUTS
54 1,5,8,NSKIN
55 larc,8,7,13,Rint,NRCUTS
56 1,6,7,NSKIN
57 !*
58 1,2,9,NCUTS
59 1,9,10,NSKIN
60 1,10,3,NCUTS
61 1,10,11,NCORE
62 1,11,6,NCUTS
63 1,11,12,NSKIN
64 1,12,7,NCUTS
65 !*
66 mat,1
67 a,1,2,3,4
68 a,5,6,7,8
69 a,2,9,10,3
70 a,6,11,12,7
71 mat,2
72 a,4,3,6,5
73 a,3,10,11,6
74 mat,1
75 amesh,1
76 amesh,2
77 amesh,3
78 amesh,4
79 mat,2
80 amesh,5
81 amesh,6
82 !*
83 DL,4,,symm
84 DL,5,,symm
85 DL,8,,symm
86 !*
87 dk,8,uy,0
88 !*
89 fk,9,fx,Forc
90 !*
91 FINISH
92 /SOLU
93 /STAT,SOLU
94 SOLVE
95 FINISH
96 /POST1
97 pldisp,1
98 *USE,INTERF
99 ...

```

On utilise la macro en tapant `*USE,INTERF`, et les contraintes sont directement modifiées dans ANSYS. On peut alors utiliser les mêmes fonctions qu’habituellement pour visualiser les différentes composantes des contraintes.

Le listing Cast3M du même problème est le suivant :

```

1  * geom: partie droite
2  long=20.0;
3  H1=0.2;
4  H2=1.8;
5  H3=2.0;
6  *
7  * geom coude
8  Rint=10.0;
9  Rh1=Rint+H1;
10 Rh2=Rint+H2;
11 Rh3=Rint+H3;
12 MRint = -1. * Rint;
13 MRh1 = -1. * Rh1;
14 MRh2 = -1. * Rh2;
15 MRh3 = -1. * Rh3;
16 *
17 * maillage
18 NCUTS=40;
19 NRCUTS=20;
20 NCORE=16;
21 NSKIN=2;
22 Forc=1.0;
23 *
24 OPTION DIMENSION 2 ELEMENT QUA4;
25 *
26 k1 = 0. MRh3;

```

```

27 k2 = Rh3 0.;
28 k3 = Rh2 0.;
29 k4 = 0. MRh2;
30 k5 = 0. MRh1;
31 k6 = Rh1 0.;
32 k7 = Rint 0.;
33 k8 = 0. MRint;
34 k9 = Rh3 long;
35 k10 = Rh2 long;
36 k11 = Rh1 long;
37 k12 = Rint long;
38 k13 = 0. 0.;
39 *
40 L1 = CERCLE NRCUTS k1 k13 k2 ;
41 L2 = DROITE k2 k3 NSKIN;
42 L3 = CERCLE NRCUTS k4 k13 k3 ;
43 L4 = DROITE k1 k4 NSKIN;
44 L5 = DROITE k4 k5 NCORE;
45 L6 = DROITE k3 k6 NCORE;
46 L7 = CERCLE NRCUTS k5 k13 k6 ;
47 L8 = DROITE k5 k8 NSKIN;
48 L9 = CERCLE NRCUTS k8 k13 k7 ;
49 L10 = DROITE k6 k7 NSKIN;
50 *
51 L11 = DROITE k2 k9 NCUTS;
52 L12 = DROITE k9 k10 NSKIN;
53 L13 = DROITE k10 k3 NCUTS;
54 L14 = DROITE k10 k11 NCORE;
55 L15 = DROITE k11 k6 NCUTS;
56 L16 = DROITE k11 k12 NSKIN;
57 L17 = DROITE k12 k7 NCUTS;
58 *
59 SURF1 = DALLER L1 L2 L3 L4;
60 SURF2 = DALLER L2 L11 L12 L13;
61 PEAUEXT = SURF1 ET SURF2;
62 * ELIMINE PEAUEXT;
63 SURF1 = DALLER L3 L6 L7 L5;
64 SURF2 = DALLER L6 L13 L14 L15;
65 AME = SURF1 ET SURF2;
66 * ELIMINE AME;
67 SURF1 = DALLER L8 L7 L10 L9;
68 SURF2 = DALLER L10 L15 L16 L17;

69 PEAUINT = SURF1 ET SURF2;
70 * ELIMINE PEAUINT;
71 PoutU = PEAUEXT ET AME ET PEAUINT;
72 *
73 Model1 = MODL peauext MECANIQUE
           ELASTIQUE ISOTROPE QUA4;
74 Model2 = MODL ame MECANIQUE ELASTIQUE
           ISOTROPE QUA4;
75 Model3 = MODL peauint MECANIQUE
           ELASTIQUE ISOTROPE QUA4;
76 ModelTot = Model1 ET Model2 ET Model3;
77 *
78 Mater1=MATERIAU Model1 YOUNG 70000.0
           NU 0.3 RHO 2700.0;
79 Mater2=MATERIAU Model2 YOUNG 3400.0 NU
           0.34 RHO 1000.0;
80 Mater3=MATERIAU Model3 YOUNG 70000.0
           NU 0.3 RHO 2700.0;
81 *
82 MR1 = RIGIDITE Model1 Mater1;
83 MR2 = RIGIDITE Model2 Mater2;
84 MR3 = RIGIDITE Model3 Mater3;
85 Mrigid = MR1 ET MR2 ET MR3;
86 *
87 CondL2 = BLOQUER UX (L4 ET L5 ET L8);
88 CondL1 = BLOQUER UY k8;
89 CondLtot = CondL1 ET CondL2;
90 FOR1 = FORCE(Forc 0.) k9;
91 Mrigid = Mrigid ET CondLtot;
92 *
93 Dep11 = RESOUD Mrigid For1 ;
94 *
95 UX1 = EXCO 'UX' dep11;
96 UY1 = EXCO 'UY' dep11;
97 TRACE UX1 PoutU;
98 TRACE UY1 PoutU;
99 *
100 def0 = DEFORMEE poutU Dep11 0.0 BLEU;
101 def1 = DEFORMEE poutU Dep11 ROUG;
102 TRACE (def0 ET def1);
103 *

```

Il reste maintenant à avoir les résultats dans les repères locaux élémentaires, et voir ce qui se passe aux interfaces...

13.3.3 Résultats

Dans les listings ci-dessus, nous n'avons pas exposé comme atteindre les contraintes, car cela sera fait en TP. Les résultats directement obtenus sont donnés à la figure 13.2 pour la déformée et les déplacements selon x et y , et à la figure 13.3 pour les déformations.

Néanmoins nous présentons quelques résultats issus de l'analyse des contraintes aux interfaces avec ou sans utilisation de la méthode de Reissner local. Nous nous concentrerons sur les résultats numériques et graphiques permettant d'apprécier la précision de la méthode de Reissner local par rapport aux résultats d'ANSYS aux points A, B, E et F ainsi que d'étudier l'influence de la valeur du rayon R.

Une première remarque s'impose : plus la valeur du rayon est faible, plus les résultats sont mauvais, quelle que soit la méthode employée (et vous devez savoir pourquoi à ce niveau du document).

Au point A, l'influence des conditions aux limites est encore sensible : en plus de la condition de symétrie de la structure, le blocage du déplacement vertical induit une légère détérioration des résultats numériques. L'influence de cette condition aux limites n'est plus visible au point B.

En A et B, la composante discontinue est σ_{xx} , les composantes σ_{yy} et σ_{xy} sont continues. Par contre, aux points E et F, c'est σ_{yy} la composante discontinue et σ_{xx} et σ_{xy} les composantes

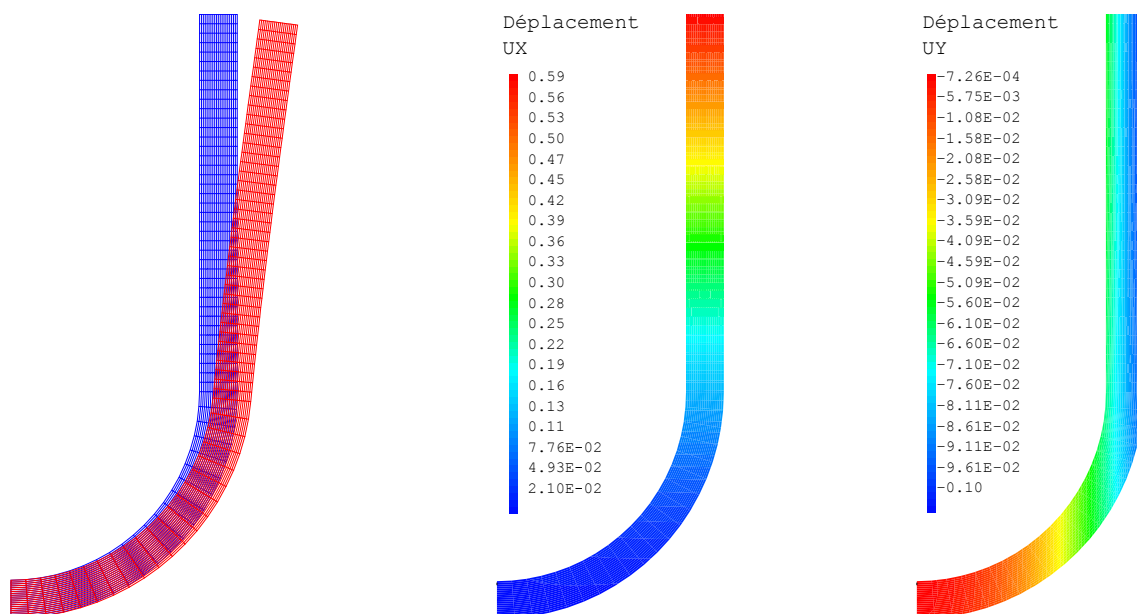


FIGURE 13.2: Déformée et déplacements

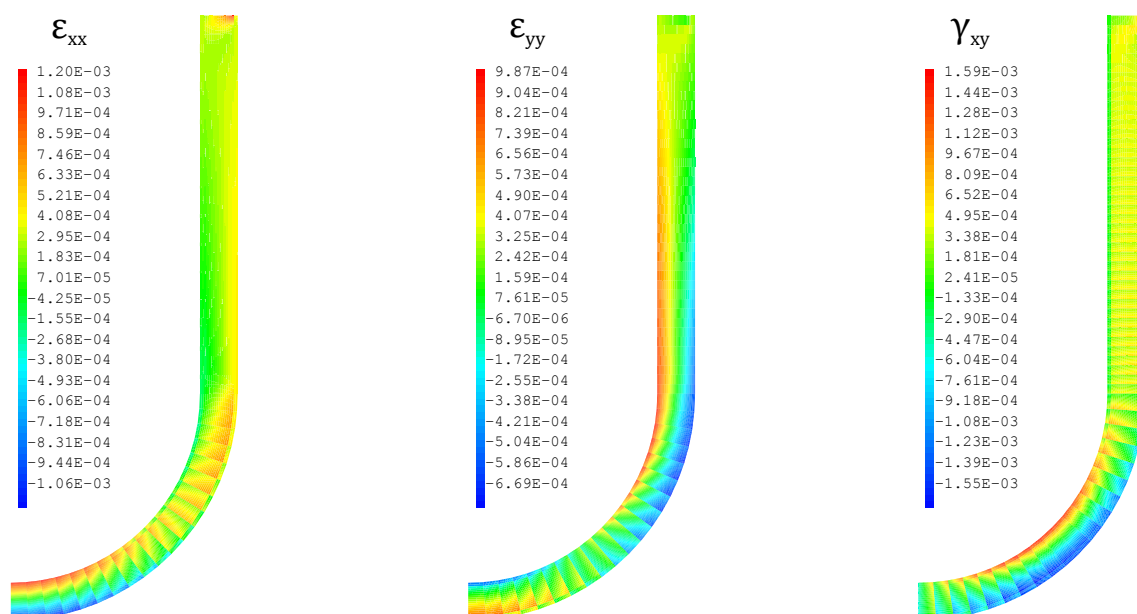


FIGURE 13.3: Déformations

continues.

D'après les résultats présentés ci-après, il est visible que la méthode de Reissner local permet d'améliorer les résultats par rapport à ANSYS.

Méthode	R (mm)	σ_{xx} peau (MPa)	σ_{xx} âme (MPa)	σ_{yy} (MPa)	σ_{xy} (MPa)
Ref	10	72,085	4,0121	1,5883	-0,0023339
Reiss	10	71,957	3,9670	1,5987	-0,0155720
ANSYS	10	71,957	3,9670	1,5116	0,0086231
Ref	8	66,500	3,8400	1,8556	-0,0022273
Reiss	8	66,665	3,7863	1,8888	-0,0118667
ANSYS	8	66,665	3,7863	1,7943	0,0111910
Ref	5	57,289	3,6598	2,6317	-0,0018770
Reiss	5	57,727	3,5771	2,6979	-0,0071024
ANSYS	5	57,727	3,5771	2,6066	0,0128210
Ref	3	49,551	3,6827	3,9304	-0,0014368
Reiss	3	49,454	3,5552	4,0110	-0,0046055
ANSYS	3	49,454	3,5552	3,9621	0,0106410
Ref	2	42,699	3,8289	5,4227	-0,0009727
Reiss	2	42,362	3,6593	5,4785	-0,0037443
ANSYS	2	42,362	3,6593	5,5264	0,0059470
Ref	1	27,171	4,3142	9,2850	-0,0004173
Reiss	1	25,876	4,0707	9,1122	-0,0021565
ANSYS	1	25,876	4,0707	9,6345	0,0005231

Tableau 13.1: Résultats au point A

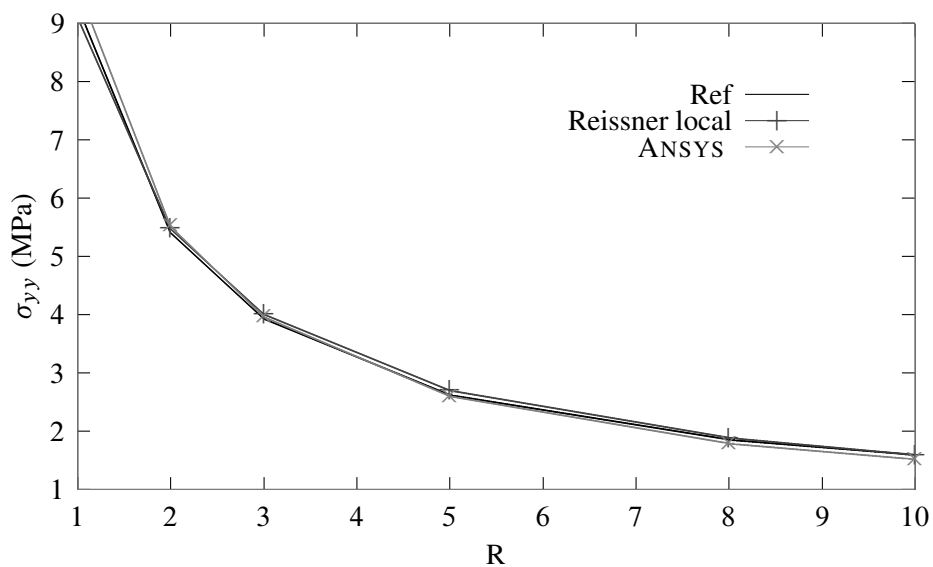


FIGURE 13.4: σ_{yy} en A

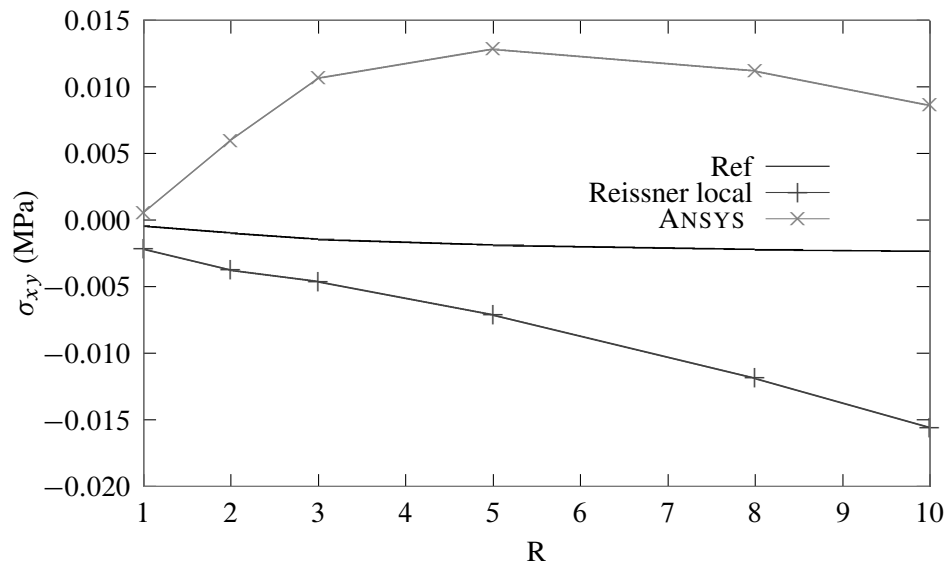


FIGURE 13.5: σ_{xy} en A

Méthode	R (mm)	σ_{xx} peau (MPa)	σ_{xx} âme (MPa)	σ_{yy} (MPa)	σ_{xy} (MPa)
Ref	10	-72,248	-3,0668	1,4080	-0,0023570
Reiss	10	-72,542	-3,1193	1,2573	-0,0138800
ANSYS	10	-72,542	-3,1193	1,5464	-0,0198700
Ref	8	-67,723	-2,7905	1,5799	-0,0026011
Reiss	8	-68,001	-2,8457	1,4570	-0,0107700
ANSYS	8	-68,001	-2,8457	1,7111	-0,0197300
Ref	5	-61,079	-2,3199	2,0320	-0,0024099
Reiss	5	-61,335	-2,3813	1,9460	-0,0067001
ANSYS	5	-61,335	-2,3813	2,1535	-0,0186210
Ref	3	-56,823	-1,9080	2,6611	-0,0022154
Reiss	3	-57,072	-1,9756	2,5888	-0,0045050
ANSYS	3	-57,072	-1,9756	2,7738	-0,0176540
Ref	2	-54,772	-1,6200	3,2396	-0,0021234
Reiss	2	-55,026	-1,6902	3,1612	-0,0036124
ANSYS	2	-55,026	-1,6902	3,3438	-0,0172760
Ref	1	-52,491	-1,1687	4,2871	-0,0020797
Reiss	1	-52,777	-1,2391	4,1625	-0,0031088
ANSYS	1	-52,777	-1,2391	4,3710	-0,0174200

Tableau 13.2: Résultats au point B

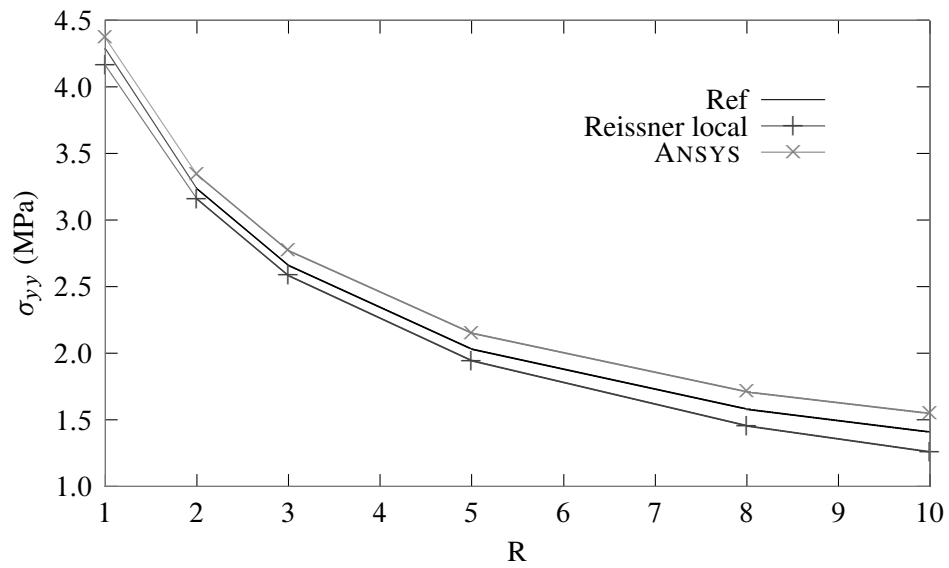


FIGURE 13.6: σ_{yy} en B

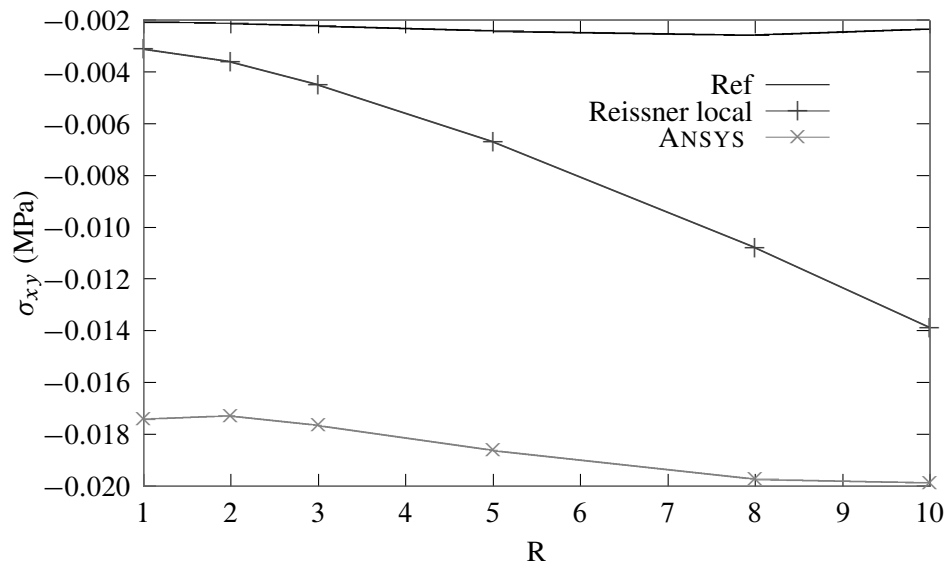


FIGURE 13.7: σ_{xy} en B

Méthode	R (mm)	σ_{yy} peau (MPa)	σ_{yy} âme (MPa)	σ_{xx} (MPa)	σ_{xy} (MPa)
ref	10	46,0340	2,410200	0,57548	0,77518
Reiss	10	46,0265	2,397450	0,48986	0,71342
ANSYS	10	46,0265	2,397450	0,62529	0,80572
ref	8	45,8625	2,443000	0,68465	0,83563
Reiss	8	45,8505	2,421450	0,62992	0,76781
ANSYS	8	45,8505	2,421450	0,71910	0,90760
ref	5	45,4490	2,539450	1,00620	1,00390
Reiss	5	45,5025	2,499550	1,00810	0,91559
ANSYS	5	45,5025	2,499550	0,94099	1,19800
ref	3	44,5760	2,694400	1,56820	1,25790
Reiss	3	45,0015	2,644550	1,62260	1,10690
ANSYS	3	45,0015	2,644550	1,30760	1,64980
ref	2	43,1510	2,856600	2,23920	1,50420
Reiss	2	44,1375	2,805000	2,32600	1,25870
ANSYS	2	44,1375	2,805000	1,77080	2,12250
ref	1	37,6330	3,184100	4,02110	1,90380
Reiss	1	40,1735	3,143850	4,04390	1,41220
ANSYS	1	40,1735	3,143850	3,04360	3,15350

Tableau 13.3: Résultats au point E

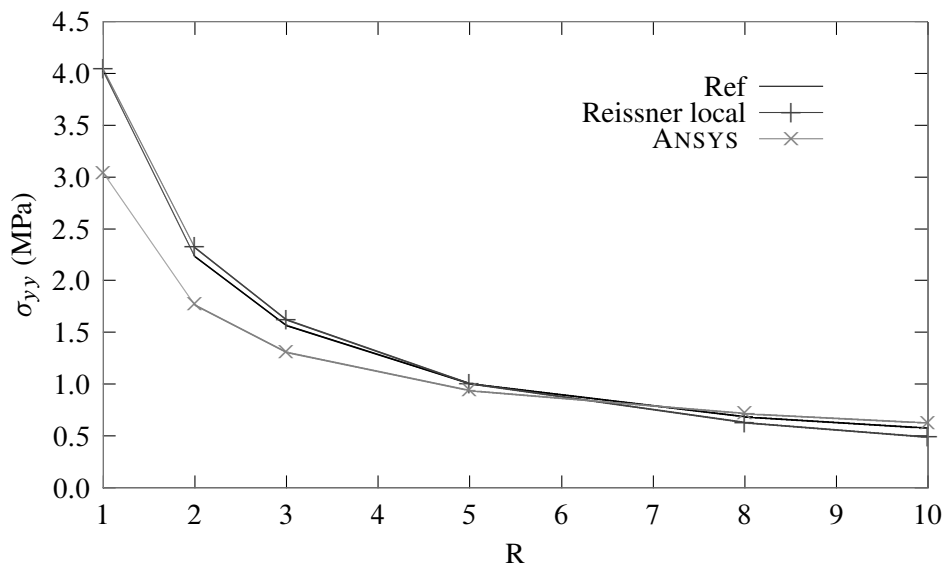


FIGURE 13.8: σ_{xx} en E

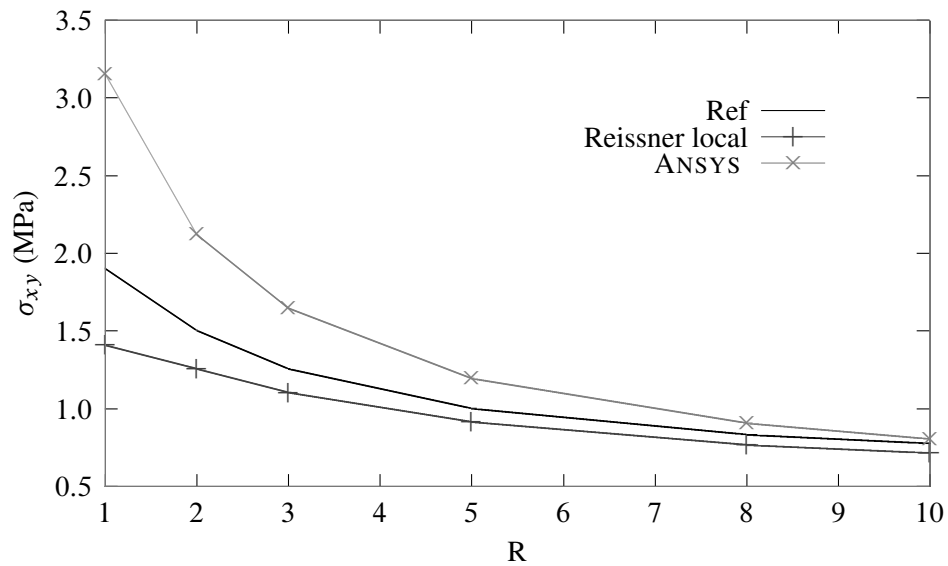


FIGURE 13.9: σ_{xy} en E

Méthode	R (mm)	σ_{yy} peau (MPa)	σ_{yy} âme (MPa)	σ_{xx} (MPa)	σ_{xy} (MPa)
ref	10	-47,438	-2,1425	0,54431	0,30073
Reiss	10	-47,510	-2,1533	0,39416	0,38936
ANSYS	10	-47,510	-2,1533	0,65814	0,30000
ref	8	-47,4555	-2,11415	0,62480	0,24743
Reiss	8	-47,5605	-2,12825	0,50447	0,33138
ANSYS	8	-47,5605	-2,12825	0,72921	0,21663
ref	5	-47,415	-2,03645	0,83167	0,10558
Reiss	5	-47,5335	-2,05940	0,76054	0,18360
ANSYS	5	-47,5335	-2,05940	0,87666	0,00251
ref	3	-47,223	-1,92305	1,11940	-0,10042
Reiss	3	-47,1885	-1,95170	1,07760	-0,01314
ANSYS	3	-47,1885	-1,95170	1,05090	-0,28747
ref	2	-46,940	-1,81445	1,38480	-0,29901
Reiss	2	-46,6760	-1,84130	1,35230	-0,18934
ANSYS	2	-46,6760	-1,84130	1,21420	-0,55238
ref	1	-46,1935	-1,61030	1,85860	-0,67358
Reiss	1	-45,4620	-1,62185	1,81270	-0,50497
ANSYS	1	-45,4620	-1,62185	1,52280	-1,03400

Tableau 13.4: Résultats au point F

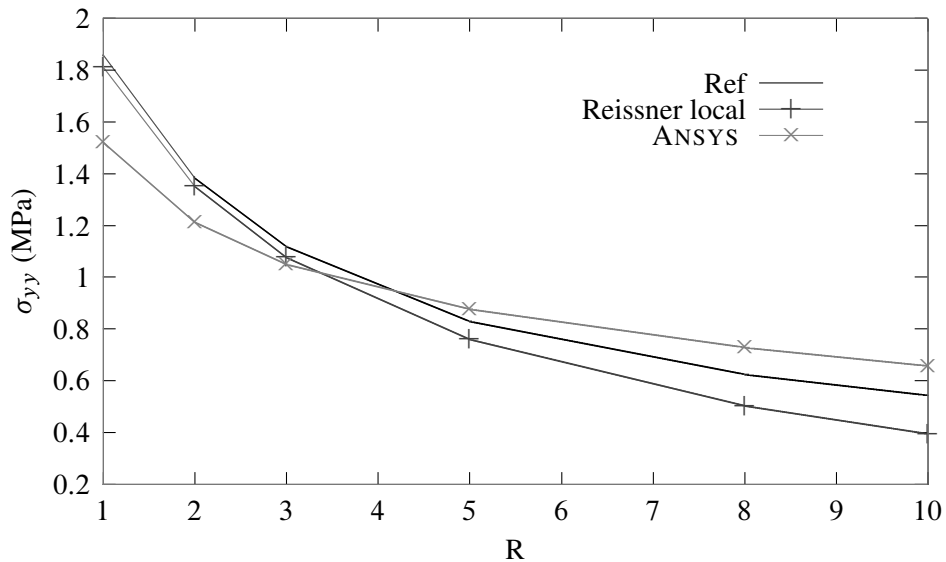


FIGURE 13.10: σ_{xx} en F

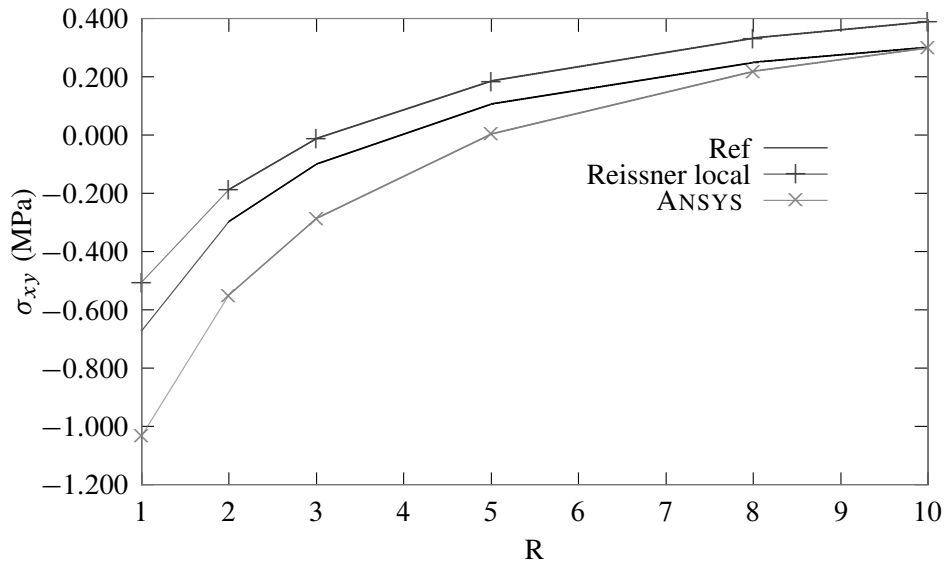


FIGURE 13.11: σ_{xy} en F

13.4 Sous-structuration et simulation multi-échelles

Le fonctionnement des produits industriels met en jeu des phénomènes physiques à des échelles très différentes. La prise en compte de tous ces phénomènes dans les simulations numériques demanderait d'utiliser des modèles extrêmement détaillés, entraînant des coûts d'identification et/ou de calcul prohibitifs. La simulation multi-échelles est une réponse à cette problématique. Elle consiste à simuler chaque phénomène à l'échelle la plus pertinente, i.e. en utilisant plusieurs modèles de tailles et de finesses différentes ; cela permet, grâce à des solveurs adaptés, de réaliser des simulations qui seraient inaccessibles par des approches plus directes.

Imaginons que nous souhaitions calculer un avion, mais que, pour des raisons évidente de sécurité, nous ayons besoin d'obtenir une précision « boulon par boulon ». On se doute bien que l'on ne peut mailler tout l'avion avec un tel niveau de détail... On recourt alors à une cascade

de modèles de dimensions et de finesses différentes, allant de l'avion tout entier (mais modélisé relativement grossièrement) à des détails structuraux modélisés très finement (mais limités à de petites zones), comme montré sur la figure 13.12.

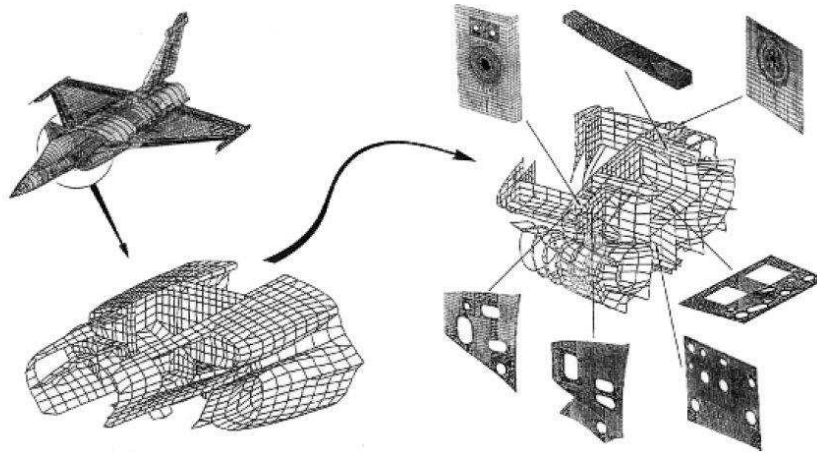


FIGURE 13.12: Sous-structuration en aéronautique

La simulation est donc décomposée en différents niveaux, chacun représentant une échelle différente. Pour coordonner ces niveaux entre eux, on utilise généralement une approche dite descendante : on commence par simuler le comportement global de l'avion, puis les résultats sont utilisés pour déterminer les conditions aux limites appliquées sur les niveaux inférieurs. Toutefois, il est parfois également nécessaire de combiner ces approches descendantes à des approches ascendantes : les résultats des simulations fines sont utilisés pour construire des modèles de comportements plus grossiers.

De très nombreuses méthodologies multi-échelles existent et reposent toutes sur le fait de simuler chaque phénomène à l'échelle la plus pertinente. Pour cela, il est nécessaire :

1. de distinguer différentes échelles dans la modélisation et dans la simulation ;
2. de modéliser les relations existant entre ces différentes échelles.

Considérons un problème comportant deux échelles. Le point 1 évoqué ci-dessus se traduit par le fait de disposer de deux modèles distincts :

- Un modèle macroscopique

Il représente le produit et son environnement extérieur et est constitué d'une géométrie et d'un modèle de comportement relativement grossiers (puisque n'ayant pas vocation à représenter les phénomènes microscopiques) ;

- Un modèle microscopique

Il possède une description géométrique et un maillage suffisamment fins ainsi que soit modélisé le comportement détaillé. Il ne comprend qu'une ou quelques « cellules types » de petites dimensions.

Il est souvent possible, pour les matériaux notamment, de disposer d'hypothèses simplificatrices (telles que l'élasticité linéaire) permettant de se restreindre à une seule cellule type. Pour le cas des matériaux anisotropes constitués de plusieurs autres matériaux tels que les matériaux composites et fibreux, on se reportera au chapitre 15.

Toutefois, il peut arriver que l'on ne dispose pas de telles hypothèses et il soit alors nécessaire de multiplier le nombre de cellules (jusqu'à parfois couvrir tout le modèle macro !))

Le point 2 consiste donc à relier les deux échelles de modélisation.

En effet, les modèles macroscopique et microscopique ne sont pas indépendants puisqu'ils modélisent la même physique à des échelles différentes. Il est donc nécessaire qu'ils soient cohérents l'un vis-à-vis de l'autre, en tout point et à chaque instant de la simulation. Pour cette raison, les modélisations multi-échelles comportent des couplages, i.e. des modèles d'interactions, entre les

échelles.

Dans le cadre de la mécanique des solides déformables, on procède généralement comme suit :

- Le modèle de comportement macroscopique : qui modélise des phénomènes se produisant « au cœur du matériau », doit correspondre à la relation contraintes/déformations observée sur la cellule microscopique ;
- Le modèle de comportement microscopique : qui traduisent la façon dont la cellule est sollicitée par son environnement extérieur, doivent correspondre à l'état de contraintes ou de déformations macroscopique.

Le modèle macroscopique ayant une résolution beaucoup plus grossière, la cohérence des deux modèles ne peut donc pas se traduire par une correspondance exacte, point par point, des conditions aux limites ou du comportement. Pour cette raison, la plupart des approches multi-échelles postulent que les quantités macroscopiques doivent correspondre à des « moyennes » des quantités microscopiques correspondantes, la définition mathématique exacte de cette moyenne variant fortement d'une approche à l'autre.

On peut donc dire que :

- *Le modèle de comportement macroscopique d'un élément de volume est choisi égal au comportement moyen de la cellule microscopique correspondante, calculé en utilisant une technique d'homogénéisation ;*
- *Les conditions aux limites microscopiques sont appliquées en moyenne, car les champs de contraintes ou de déplacements macroscopiques sont beaucoup trop grossiers.*

Le problème est donc d'échanger des données pertinentes entre les différentes échelles, compte tenu des différentes résolutions des modèles.

Dans le sens microscopique vers macroscopique, on utilisera les techniques d'homogénéisation dont il sera l'objet au chapitre 15.

Dans le sens macroscopique vers microscopique, il s'agit donc de spécifier des conditions aux limites à appliquer sur le bord des cellules microscopiques, à partir d'une solution macroscopique.

Les méthodes les plus simples se contentent d'imposer directement le champ de déplacements (ou de contraintes) macroscopique comme condition aux limites. Le modèle macroscopique ayant par définition une résolution beaucoup plus grossière que le modèle microscopique, il est incapable de capturer l'allure microscopique des déplacements ou des contraintes. Ces conditions aux limites sont donc souvent trop imprécises par rapport aux finalités du modèle microscopique. Cela peut dégrader fortement la qualité des résultats et donc restreindre le domaine de validité de ces approches.

Pour palier ce problème, on écrit les conditions aux limites microscopiques de manière plus subtile en les décomposant en la somme d'un terme moyen et d'un terme de moyenne nulle. Ainsi, le champ microscopique u^m sera écrit $u^m = u^M + v^m$ où u^M est le champ macroscopique (i.e. la moyenne du champ microscopique), et v^m est le reste (à moyenne nulle). Notre problème devient donc de déterminer le reste v^m , et c'est là que les méthodes divergent le plus. On distingue néanmoins deux grosses familles :

- Condition de périodicité
On postule que l'on connaît l'allure du reste. On peut alors enchaîner un calcul macroscopique avec un solveur spécifique puis un calcul microscopique avec un solveur classique ;
- Couplage des cellules microscopiques
Lorsqu'il n'est pas si simple de « séparer » les échelles, alors il faut résoudre en même temps les deux problèmes microscopique et macroscopique avec échange de données.

Ces deux approches vont être un peu plus détaillées maintenant.

13.4.1 Condition de périodicité – méthodes multi-niveaux

Souvent, les méthodes multi-niveaux sont basées sur des conditions de périodicité inspirées de l'homogénéisation périodique. Ces conditions supposent que la partie microscopique du champ

de déplacement, v^m , est périodique, i.e. est égale sur chaque paire de faces opposées de la cellule type considérée. La même hypothèse est formulée sur les contraintes. On obtient ainsi un ensemble de conditions aux limites permettant de déterminer entièrement la solution microscopique à partir d'une déformation (ou d'une contrainte) macroscopique imposée.

Le résultat microscopique ainsi obtenu est pertinent à deux conditions : 1) il faut que la microstructure soit effectivement périodique, et 2) que le principe de Saint-Venant s'applique, i.e. que l'on se trouve suffisamment loin de la surface de la pièce (y compris des détails géométriques tels que des trous, des fissures...). Dans le cas contraire, des effets de bord peuvent affecter l'allure de la solution, qui n'est alors plus périodique : il faut donc recourir à d'autres modélisations.

La simulation multi-niveaux fait appel à deux types de modèles, chacun équipé de son solveur :

- un modèle macroscopique ne possédant pas de relation de comportement du matériau prédéfinie, équipé d'un solveur éléments finis modifié ;
- un ensemble de modèles microscopiques ne possédant pas de conditions aux limites prédéfinies, équipés de solveurs éléments finis classiques.

En fait, le solveur éléments finis modifié est un solveur éléments finis classique avec une petite différence : à chaque fois que le solveur macroscopique a besoin du comportement d'un élément de volume quelconque, il envoie l'état de déformation macroscopique de cet élément au solveur microscopique. Ce dernier dispose alors toutes les données nécessaires pour simuler numériquement son comportement à l'échelle microscopique. Il est donc capable de renvoyer l'état de contraintes macroscopiques, selon la décomposition mentionnée ci-dessus. Les cellules microscopiques étant pour ainsi dire « indépendantes », il est possible de recourir à des calculateurs parallèles.

Le domaine de validité de ces méthodes est bon dès que l'on sait faire des hypothèses réalistes sur l'allure de la solution microscopique. Ce n'est pas toujours le cas : par exemple, la fissuration, lorsqu'elle sort d'un cadre microscopique pour atteindre un cadre macroscopique, se prête très mal à cet exercice. Le cas de la fissuration sera abordé au chapitre 21.

En résumé :

Les méthodes multi-niveaux abordent la simulation à l'échelle macroscopique et se « nourrissent » du comportement simulé à l'échelle microscopique.

Dans les méthodes multi-niveaux, l'emploi d'hypothèses de périodicité se traduit par des « sauts » de contraintes et de déplacements d'une cellule à l'autre ; si les échelles sont mal séparées, ces sauts sont non négligeables. Ne correspondant a priori pas à la physique, ils traduisent un écart avec la réalité.

13.4.2 Couplage des cellules microscopiques – méthodes de décomposition de domaine

Les méthodes multi-niveaux présentées au paragraphe précédent sont mises à mal lorsque le comportement microscopique « déborde » un peu sur le comportement macroscopique, i.e. lorsque les échelles ne sont pas suffisamment bien séparées. On recourt alors aux méthodes de décomposition de domaines, dont la validité est plus large, mais qui sont plus complexes.

Puisque nous ne pouvons plus supposer une allure de la solution microscopique, nous n'allons utiliser que les seules connaissances disponibles a priori : la continuité du champ de déplacements ainsi que la vérification par les contraintes du principe d'action-réaction. Il « suffit » donc d'écrire qu'à l'interface entre deux cellules microscopiques, il y a égalité des déplacements et nullité de la somme des traces des contraintes (on retrouve ce que nous avons déjà plusieurs fois évoqué avec l'état des contraintes à l'interface entre deux matériaux différents).

La présence de couplages entre les cellules microscopiques (qui en quelque sorte correspond à une généralisation de la condition de périodicité du paragraphe précédent) change complètement le déroulement de la simulation par rapport aux méthodes multi-niveaux vues au paragraphe précédent. En effet, la prise en compte de ces couplages implique d'échanger directement des données entre les différents solveurs microscopiques, qui ne sont plus « indépendants ». En contrepartie, cela permet

de propager une information fine sur l'ensemble de la pièce et, ainsi, de se passer de l'hypothèse de séparation des échelles : il n'est plus nécessaire de modéliser séparément les phénomènes microscopiques et macroscopiques.

Concrètement, les méthodes de décomposition de domaine sont des solveurs qui ont généralement un fonctionnement multi-échelles. Elles partent d'un modèle microscopique du produit décomposé en sous-structures, et consistent à coupler les sous-structures en échangeant des contraintes et des déplacements sur les interfaces. Des versions plus ou moins simples existent. Notons bien qu'il n'y a pas de modèle macroscopique dans une telle approche.

En résumé :

Dans la décomposition de domaine, la simulation est abordée à l'échelle la plus fine : la sous-structuration et le problème grossier ne sont utilisés que pour améliorer l'efficacité de la résolution.

Une simulation par décomposition de domaine conduit toujours à un champ de déplacement microscopique continu sur toute la structure, et un champ de contraintes microscopique équilibré (au sens des éléments finis) sur toute la structure.

Actuellement, la simulation multi-échelles est encore relativement peu répandue dans le monde de l'ingénierie. Elle représente en effet un changement considérable par rapport aux pratiques usuelles de simulation ; de plus, la plupart des logiciels de calcul multi-échelles sont des outils développés par des chercheurs, qui n'ont pas encore l'ergonomie et la robustesse des solveurs utilisés dans l'industrie. Les industriels attendent donc l'apparition d'outils mieux adaptés à leurs problématiques avant d'envisager une utilisation plus fréquente de ces méthodes.

Cependant, à plus long terme, la simulation multi-échelles suscite un intérêt considérable dans l'industrie : en rendant accessibles à la simulation des phénomènes qui ne peuvent actuellement être étudiés qu'expérimentalement, elle constitue un pas important en direction du « virtual testing ». Pour cette raison, elle fait toujours l'objet de nombreux projets de recherche. Ceux-ci concernent aussi bien la modélisation, avec notamment la mise au point de « matériaux virtuels » (qui ne sont rien d'autre que des modèles multi-échelles, notamment pour les matériaux composites), que les solveurs qui sont en constante évolution.

13.5 Super-éléments

Histoire

Originellement introduit dans l'aéronautique dans les années 60 (d'où le choix de la figure 13.12), le concept de sous-structuration répondait à trois motivations :

- faciliter la division du travail : des sous-structures avec des fonctions différentes (fuselage, ailes, train d'atterrissage...) pouvaient être traitées par des groupes d'experts différents. Chaque groupe pouvait à loisir améliorer, raffiner... sa partie tant que l'interface avec les autres parties restait inchangée.
- profiter de la répétition : en remarquant qu'une même structure peut contenir plusieurs sous-structures identiques, il est possible de diminuer le temps d'étude (par exemple symétrie des ailes...)
- contourner les limitations des ordinateurs : les ordinateurs de l'époque atteignaient vite leurs limites (par exemple en terme de taille mémoire). Diviser une structure complexe, que l'on était incapable de calculer en une seule fois, permettait de sauvegarder des résultats sous-structure par sous-structure puis d'effectuer « l'assemblage » des résultats.

Si les deux premiers points sont toujours d'actualité, le troisième l'est moins, surtout depuis le recourt aux algorithmes parallèles.

C'est d'ailleurs le développement de procédures pour le calcul parallèle qui a conduit les mathématiciens appliqués au concept de sous-domaines, alors qu'ils devaient grouper des éléments pour des raisons de calcul.

Dans ce paragraphe, nous allons parler du concept de super-élément, qui n'est qu'une application de ce qui a été présenté au paragraphe précédent.

Un super-élément est un groupement d'éléments qui, après assemblage, peuvent être vus comme un élément individuel du point de vue du calcul. Cet assemblage peut être requis pour des raisons de modélisation ou de calcul.

Pour constituer un super-éléments, les éléments groupés ne peuvent être pris au hasard. Ils doivent au moins constituer une « structure » en eux-même, mais d'autres conditions sont nécessaires qui seront détaillées plus loin.

Comme nous l'avons dit au paragraphe précédent, il y a deux voies duales pour considérer ce processus. L'approche descendante consiste à considérer un super-élément comme constitué d'un ensemble d'éléments. On parle alors de macro-élément. L'approche ascendante consiste à considérer un super-élément comme un sous-ensemble d'une structure complète. On parle alors de sous-structure.

Finalement, quand parle-t-on de sous-structure ou de macro-élément ? En fait, il n'y a pas de règle, et le terme générique de super-élément couvre tout le spectre depuis l'élément individuel jusqu'à la structure complète.

13.5.1 Condensation statique

En tant qu'assemblage de plusieurs éléments, un super-élément possède :

- des degrés de liberté internes : qui ne sont pas connectés à des éléments n'appartenant pas au super-élément considéré. Les nœuds ayant des degrés de liberté internes sont dits nœuds internes.
- des degrés de liberté aux frontières : qui sont connectés à au moins une entité (élément, super-élément) n'appartenant pas au super-élément considéré.

L'opération consistant à éliminer tous les degrés de liberté internes est appelée condensation statique ou simplement condensation. Regardons ce qui se passe d'un point de vue matriciel. Pour cela, considérons que nous ayons à résoudre le système :

$$\mathbf{K}\mathbf{q} = \mathbf{F} \quad (13.1)$$

Le vecteur \mathbf{q} se compose des composantes internes \mathbf{q}_i et des composantes de frontière \mathbf{q}_b , de sorte que le système, une fois réordonné s'écrit :

$$\begin{bmatrix} \mathbf{K}_{bb} & \mathbf{K}_{bi} \\ \mathbf{K}_{ib} & \mathbf{K}_{ii} \end{bmatrix} \begin{pmatrix} \mathbf{q}_b \\ \mathbf{q}_i \end{pmatrix} = \begin{pmatrix} \mathbf{f}_b \\ \mathbf{f}_i \end{pmatrix} \quad (13.2)$$

Si la matrice \mathbf{K}_{ii} n'est pas singulière, alors la seconde équation peut être résolue en terme de variables internes en :

$$\mathbf{q}_i = \mathbf{K}_{ii}^{-1} (\mathbf{f}_i - \mathbf{K}_{ib}\mathbf{q}_b) \quad (13.3)$$

En reportant cela dans la première équation, on obtient le système avec matrice de rigidité condensée :

$$\tilde{\mathbf{K}}_{bb}\mathbf{q}_b = \tilde{\mathbf{f}}_b \quad (13.4)$$

avec :

$$\tilde{\mathbf{K}}_{bb} = \mathbf{K}_{bb} - \mathbf{K}_{bi}\mathbf{K}_{ii}^{-1}\mathbf{K}_{ib} \quad \text{et} \quad \tilde{\mathbf{f}}_b = \mathbf{f}_b - \mathbf{K}_{bi}\mathbf{K}_{ii}^{-1}\mathbf{f}_i \quad (13.5)$$

Après condensation, on peut donc bien considérer le super-élément, d'un point de vue calculatoire, comme un élément individuel.

Notons que la matrice \mathbf{K}_{ii} n'est pas singulière si elle possède la condition de rang suffisant, i.e. si elle ne contient que des modes à énergie nulle correspondant aux modes rigides (cette condition a déjà évoquée à propos de la validation des éléments). Si cela n'est pas le cas, le super-élément est dit flottant, et peut quand même être traité (en utilisant les projecteurs et inverses généralisés, mais c'est un peu plus compliqué).

La condensation statique est une opération matricielle appelée inversion partielle ou élimination partielle ou pseudo-inversion.

13.5.2 Remonter aux degrés de liberté internes

Une fois le système condensé résolu, on obtient les valeurs aux nœuds internes en réutilisant la formule :

$$\mathbf{q}_i = \mathbf{K}_{ii}^{-1} (\mathbf{f}_i - \mathbf{K}_{ib}\mathbf{q}_b) \quad (13.6)$$

13.6 Pseudo-inversion et réanalyse

On peut être amené, notamment lors de phases de conceptions, à devoir considérer plusieurs problèmes « relativement proches » les uns des autres (i.e. tester plusieurs configurations). On est alors tenté d'utiliser tout ou partie de la première modélisation afin de réaliser les suivantes.

L'idée de la méthode de réanalyse est d'analyser le comportement d'une structure élastique par éléments finis sans particulariser le système d'équations final par la prise en compte de conditions cinématiques.

On obtient alors une solution générale de ce système faisant intervenir une matrice de rigidité régularisée. La nature de cette matrice (somme d'une matrice bande et d'une matrice pleine) ne permet pas d'utiliser les méthodes les plus classiques et optimales de résolution, mais il est possible d'en développer d'autres permettant d'accéder à la quasi-inverse d'une matrice singulière semi-définie positive (tout en profitant de son caractère bande).

Il est alors possible, tout en modifiant les conditions cinématiques et les chargements appliqués, de procéder à des réanalyses qui consistent alors simplement à résoudre des systèmes dits secondaires de tailles très inférieures au système global (mais un surcoût a été « payé » initialement pour calculer la pseudo-inverse).

Des problèmes de contact avec ou sans frottement entre solides élastiques peuvent bénéficier de cette méthode, ainsi que la modélisation du comportement élastique incompressible.

13.6.1 Modification du chargement uniquement

Restons sur le problème structurel correspondant au système matriciel :

$$\mathbf{K}\mathbf{q} = \mathbf{F} \quad (13.7)$$

où \mathbf{K} est de dimension $n \times n$.

Si l'on souhaite considérer plusieurs cas de chargement $\mathbf{F}_1, \dots, \mathbf{F}_k$, on peut soit résoudre k fois le système précédent, soit résoudre le système :

$$\mathbf{K}\bar{\mathbf{q}} = \bar{\mathbf{F}} \quad (13.8)$$

où $\bar{\mathbf{F}}$ est la matrice $n \times k$ des k vecteurs de chargement, et $\bar{\mathbf{q}}$ est la matrice $n \times k$ des k vecteurs solutions correspondants.

Cette méthode, la plus simple des méthodes de réanalyse, permet malgré tout d'économiser des opérations.

13.6.2 Modification de la matrice

Considérons maintenant le cas où ce n'est plus le chargement \mathbf{F} qui peut varier d'une analyse à l'autre, mais la matrice \mathbf{K} .

On est amené à chercher une solution du système :

$$(\mathbf{K} + \Delta\mathbf{K})\mathbf{q} = \mathbf{F} \quad (13.9)$$

en fonction de la solution du système non perturbé, i.e. à calculer $(\mathbf{K} + \Delta\mathbf{K})^{-1}$ en fonction des autres matrices. On rappelle que $\Delta\mathbf{K}$ est appelé perturbation de la matrice \mathbf{K} .

Certaines formules existent pour des modification mineures de la matrice \mathbf{K} , mais nous ne les présenterons pas.

13.6.3 Modification des conditions cinématiques

Nous ne considérons ici que des modifications des conditions cinématiques. Généralement, les conditions cinématiques sont prises en compte en supprimant ou en modifiant les équations de l'équilibre avant résolution, ce qui rend le système régulier. Nous avons vu que le système à résoudre (incluant les conditions cinématiques) revient à chercher le minimum de la forme quadratique :

$$\mathcal{L} = \frac{1}{2} {}^T \mathbf{q} \mathbf{K} \mathbf{q} - {}^T \mathbf{q} \mathbf{F} \quad (13.10)$$

Notons que l'on peut écrire les p conditions cinématiques sous la forme : $\mathbf{C} \mathbf{q} = \delta$ avec \mathbf{C} une matrice $p \times n$, et \mathbf{q} et δ des vecteurs.

Nous avons également déjà vu que ces conditions peuvent être prises en compte par l'intermédiaire de p multiplicateurs de Lagrange λ dans la fonctionnelle précédente qui devient alors :

$$\mathcal{L}^* = \frac{1}{2} {}^T \mathbf{q} \mathbf{K} \mathbf{q} - {}^T \mathbf{q} \mathbf{F} - {}^T \lambda (\mathbf{C} \mathbf{q} - \delta) \quad (13.11)$$

Le système à résoudre est symétrique, régulier, mais non défini-positif.

Pour pallier la lenteur des algorithmes disponibles pour la résolution numérique d'un tel système (méthode de Gauß ou de décomposition avec pivotage), il est préférable de régulariser le système.

Pour cela, on s'appuie sur la connaissance de la matrice \mathbf{K} : celle-ci est singulière d'ordre r , où r correspond aux mouvements de corps rigide, ou modes rigides. Cela veut également dire qu'elle possède $(n - r)$ valeurs propres strictement positives (et r valeurs propres nulles).

Il est évident qu'il faut au moins disposer de $p \geq n$ conditions aux limites cinématiques pour que le système puisse admettre une solution. C'est évidemment l'hypothèse que nous ferons (sinon le problème est mal posé).

Nous considérons la matrice \mathbf{R} de dimension $n \times r$ des r vecteurs propres de \mathbf{K} correspondant à la valeur propre nulle. Quitte à construire ces vecteurs (qui sont orthogonaux), nous les choisirons normés. On a alors :

$$\begin{cases} \mathbf{K} \mathbf{R} = \mathbf{0} \\ {}^T \mathbf{R} \mathbf{R} = \mathbf{I}_r \end{cases} \quad (13.12)$$

On introduit la matrice :

$$\mathbf{K}_\alpha = \mathbf{K} + \alpha \mathbf{R}^T \mathbf{R} \quad (13.13)$$

qui, pour $\alpha > 0$ est régulière car symétrique et définie-positive. Les colonnes de \mathbf{R} sont vecteurs propres de \mathbf{K}_α pour la valeur propre α .

Dans le système $\mathbf{K} \mathbf{q} = \mathbf{F}$, on remplace \mathbf{K} par $\mathbf{K}_\alpha - \mathbf{R}^T \mathbf{R}$, et en introduisant le fait que $\mathbf{K}_\alpha \mathbf{R}^T \mathbf{R} = \alpha \mathbf{R}^T \mathbf{R}$, on obtient finalement :

$$\mathbf{K}_\alpha (\mathbf{I}_n - \mathbf{R}^T \mathbf{R}) \mathbf{q} = \mathbf{F} \quad (13.14)$$

Si l'on change de variable en posant $\mathbf{v} = (\mathbf{I}_n - \mathbf{R}^T \mathbf{R}) \mathbf{q}$ (\mathbf{v} est la projection de \mathbf{q} sur l'orthogonal du noyau de \mathbf{K}), \mathbf{v} devient l'inconnue du système régularisé :

$$\mathbf{K}_\alpha \mathbf{v} = \mathbf{F} \quad (13.15)$$

et l'on retrouvera la solution du système initial \mathbf{q} par :

$$\mathbf{q} = \mathbf{v} + \mathbf{R}^T \mathbf{R} \mathbf{q} = \mathbf{K}_\alpha^{-1} \mathbf{F} + \mathbf{R}^T \mathbf{R} \mathbf{q} \quad (13.16)$$

On peut remarquer que \mathbf{K}_α^{-1} , que l'on note \mathbf{S}_α , possède les mêmes valeurs propres que \mathbf{K}_α (et donc les mêmes que \mathbf{K} et $\mathbf{R}^T \mathbf{R}$). On est donc naturellement amené à décomposer \mathbf{S}_α comme :

$$\mathbf{S}_\alpha = \mathbf{S} + \frac{1}{\alpha} \mathbf{R}^T \mathbf{R} \quad (13.17)$$

avec $\mathbf{S} \mathbf{R} = 0$ et $\mathbf{R}^T \mathbf{S} = 0$.

C'est cette matrice \mathbf{S} que l'on appelle quasi-inverse de \mathbf{K} , car :

$$\mathbf{S} \mathbf{K} = \mathbf{K} \mathbf{S} = \mathbf{I} - \mathbf{R}^T \mathbf{R} \quad (13.18)$$

Elle est de dimension $n \times n$, symétrique, semi-définie positive (ses valeurs propres sont de même signe), admet les mêmes valeurs propres que \mathbf{K} et en particulier ceux de la valeur propre nulle dont \mathbf{R} est une base, mais elle ne possède pas de caractère bande.

Voici brossé, en quelques lignes, les idées principales de la méthode. Nous n'irons pas plus loin dans sa présentation.

On rappelle que l'intérêt de la méthode est qu'une fois une première étape consistant à intégrer les données relatives à la structure (géométrie, discrétisation, matériaux) réalisée, on peut alors effectuer autant de fois que nécessaire la seconde étape qui porte sur la prise en compte des chargements et conditions aux limites.

Notons qu'il est possible de modifier un peu la forme du système secondaire afin de pouvoir prendre en compte, lors de la seconde étape, des conditions plus complexes, comme des conditions mixtes par exemple...

13.6.4 Deux mots de statistiques

La méthode de réanalyse permet donc, pour un coût de calcul maîtrisé, d'obtenir les réponses à plusieurs problèmes proches. Il se peut donc tout à fait que l'on désire faire varier un paramètre afin d'en étudier l'influence. Notons X ce paramètre variable, et $Y = \mathcal{F}(X)$ la réponse du système correspondante.

Dans ce cas, si l'on connaît la répartition statistique de X , i.e. sa densité de probabilité (encore appelée fonction de masse), alors on peut réaliser un échantillonnage correct de la plage dans laquelle ce paramètre varie afin de pouvoir obtenir des statistiques sur la réponse Y . Comme on raisonne sur un nombre de points finis, on utilisera donc une formulation discrète. Par exemple, une approximation de la moyenne de la réponse sera donnée par une formule de type :

$$E[Y] = \frac{\sum p_i \mathcal{F}(X_i)}{\sum p_i} \quad (13.19)$$

où p_i est le poids, ou la pondération, correspondant à la fréquence à laquelle la valeur X_i du paramètre X apparaît. On pourrait tout aussi bien, suite à l'analyse des réponses Y , approcher celle-ci par une fonction quelconque, i.e. donner une forme analytique à la fonction \mathcal{F} , et raisonner ensuite de manière continue.

Si c'est vraiment la caractérisation d'un système sous sollicitations aléatoires qui nous intéresse, alors on effectuera plutôt une analyse stochastique, et on se reportera au chapitre 22.

13.7 Dérivées d'ordre supérieur

Au paragraphe précédent, nous avons vu comment, sur une structure donnée, il était possible de prendre en compte plusieurs chargements et conditions aux limites cinématiques sans avoir à refaire tout le calcul.

Dans le même état d'esprit, nous allons voir maintenant comment optimiser la forme d'une structure, sans refaire tous les calculs.

Considérons le cas d'une conception ayant pour but de déterminer la forme la plus adaptée selon certains critères (rigidité, déformée, contraintes, énergie transmise...). Chaque évaluation d'une fonction coût conduit à une analyse par éléments finis. L'utilisation de dérivées par rapport à la géométrie, ou plus généralement par rapport à la fonction coût, permet de réduire ce nombre d'analyse.

En fait, peut-être contrairement à l'intuition, le calcul de ces dérivées est relativement peu coûteux ; il n'est pas difficile et peut être fait automatiquement. Les dérivées d'ordre supérieur d'une fonction coût peuvent en fait être calculées avec autant de précision que la fonction elle-même. Ainsi, en un seul calcul, il est possible d'obtenir un développement de Taylor de la fonction coût, et donc d'éviter de nombreuses analyses.

13.7.1 Dérivées par rapport à la géométrie

En plus de considérer un domaine borné Ω de \mathbb{R}^n , nous allons considérer une perturbation V de Ω , et nous noterons :

$$\Omega + V = (I + V)(\Omega); \quad V \in W^{1,\infty}(\Omega; \mathbb{R}^n) \quad (13.20)$$

Nous considérons la fonction coût $J(\Omega, u_\Omega)$ choisie pour décrire le problème d'optimisation. Cette fonction dépend donc naturellement du domaine Ω et de la solution du problème sur ce domaine u_Ω . Nous allons donner un sens à la dérivée de la fonction coût par rapport aux variations du domaine.

Si l'on considère le cas simple $J(\Omega, u) = \int_\Omega u$, alors il vient :

$$\frac{dJ}{d\Omega}(\Omega, u_v, V) = \int_\Gamma u_\Omega V \cdot n + \int_\Omega u'_{\Omega;v} \quad (13.21)$$

avec une fois encore n la normale extérieure de $\Gamma = \partial\Omega$, et :

$$u'_{\Omega;v} = \lim_{t \rightarrow 0} \frac{u_{\Omega+tV}(x) - u_\Omega(x)}{t}, \quad \forall x \in \Omega \quad (13.22)$$

La méthode de dérivation par transport conduit à :

$$\frac{dJ}{d\Omega}(\Omega, u_v, V) = \int_\Gamma u_\Omega \operatorname{div} V + \int_\Omega \dot{u}_{\Omega;v} \quad (13.23)$$

avec :

$$\dot{u}_{\Omega;v} = \lim_{t \rightarrow 0} \frac{u_{\Omega+tV} \circ (I + tV)(x) - u_\Omega(x)}{t}, \quad \forall x \in \Omega \quad (13.24)$$

On définit alors la dérivée de la fonction coût par rapport aux variations du domaine par :

$$\frac{\partial J}{\partial \Omega}(\Omega, u, V) = \int_\Gamma u V \cdot n \quad (13.25)$$

La difficulté tient à ce que l'ensemble des domaines Ω ne constitue pas un espace vectoriel. En effet, les perturbations ne s'ajoutent pas, ou au moins ne sont pas associatives : $(\Omega + V) + W \neq (\Omega + W) + V$.

Toutefois, si l'on utilise un paramétrage du domaine, il est alors possible d'utiliser les outils classiques du calcul différentiel dans les espaces normés. On introduit alors le paramètre F par :

$$J(F, u) = J(F(\Omega), u \circ F^{-1}), \quad F \in W^{1,\infty}(\Omega; \mathbb{R}^n) \quad (13.26)$$

et l'on a alors $(F + V) + W = (F + W) + V$.

La dérivée partielle de J par rapport au paramètre F est définie sans ambiguïté. Dans notre exemple, cette dérivée, prise au point $F = I$ est :

$$\frac{\partial J}{\partial F}(I, u) \cdot V = \int_{\Omega} u \operatorname{div} V \quad (13.27)$$

On pourra alors nous faire remarquer que :

$$\frac{\partial J}{\partial \Omega}(\Omega, u, V) \neq \frac{\partial J}{\partial F}(I, u) \cdot V \quad (13.28)$$

ce à quoi nous répondrons que c'est le prix à payer pour se ramener à un espace vectoriel. Cette démarche peut alors être généralisée aux ordres supérieurs.

Nous parlerons plus en détail de dérivées par rapport à la géométrie dans le chapitre 16 sur l'optimisation, où sera abordée l'optimisation de forme.

13.7.2 Calcul des dérivées

Nous n'entrons pas dans le détail, mais en cours de calcul se pose la question intéressante : obtient-on le même résultat si l'on discrétise d'abord et dérive ensuite ? En fait, la réponse est oui, sous certaines conditions de régularité que nous ne mentionnerons pas dans le cadre de ce document.

Encore une fois, nous n'avons fait qu'effleurer le problème, juste pour « faire connaître » la méthode. Il nous semblait intéressant de présenter la notion de dérivée par rapport à la géométrie.

De telles méthodes sont d'ores et déjà implémentées dans certains codes de calculs. Leur intérêt devrait apparaître clairement au lecteur (nous l'espérons).

Chapitre 14

Le maillage

Résumé — À ce niveau du document, on peut considérer que la méthode des éléments finis a été présentée, au moins en ce qui concerne les aspects les plus classiques (et même un peu plus). Nous avons décidé, avant d’entrer dans le détail de « subtilités » liées au comportement des matériaux et à la non-stationnarité, d’insérer ici un petit chapitre sur le maillage, dont les techniques de construction n’ont rien de commun avec celles relatives aux éléments eux-mêmes. De plus, nous nous restreindrons aux maillages de type Delaunay.-Voronoi

Le maillage n’a pas seulement un intérêt en calcul scientifique. Le maillage est un « support » de représentation tridimensionnel utilisé par exemple dans les jeux vidéo ainsi que dans les animations 3D. Dans ce dernier cas, on s’intéresse à la qualité du maillage en lien avec la qualité de rendu de l’image générée lorsque l’on applique une texture sur le maillage. On s’intéresse également à comment « faire bouger » le maillage pour qu’un personnage n’apparaissent pas distordu pendant une animation...

L’opération de maillage peut se faire à partir de plusieurs données :

- soit à partir de données à utiliser pour la discrétisation : sommets, arêtes...
- soit à partir de données de type CAO décrivant uniquement les entités géométriques.

Nous nous contenterons de présenter quelques outils de maillage relatifs au premier cas, i.e. lorsque nous disposons de points, arêtes... mais le second cas n’est pas vraiment plus compliqué.

Les méthodes de construction de maillage sont essentiellement :

- Maillage par triangles/tétraèdres :
 - Méthodes utilisant le critère de Delaunay : les bases de la méthode seront exposées au paragraphe 14.1 ;
 - Méthodes par avancement de fronts : le principe en sera détaillé au paragraphe 14.2 ;
 - Méthodes par décomposition spatiale ;
- Maillage par quadrangles/hexaèdres : quelques remarques seront faites au paragraphe 14.4
 - Maillage par avancement de fronts ;
 - Maillage par décomposition en domaines.

Parfois, il ne s’agit pas de mailler un domaine, mais de le remailler. Les méthodes d’adaptation de maillage les plus connues sont :

- Raffinement ;
- Transformation des éléments ;
- Déplacements de nœuds ;
- Simplification de maillage.

L’utilisation de différentes méthodes est illustré à la figure 14.1 issue de [6].

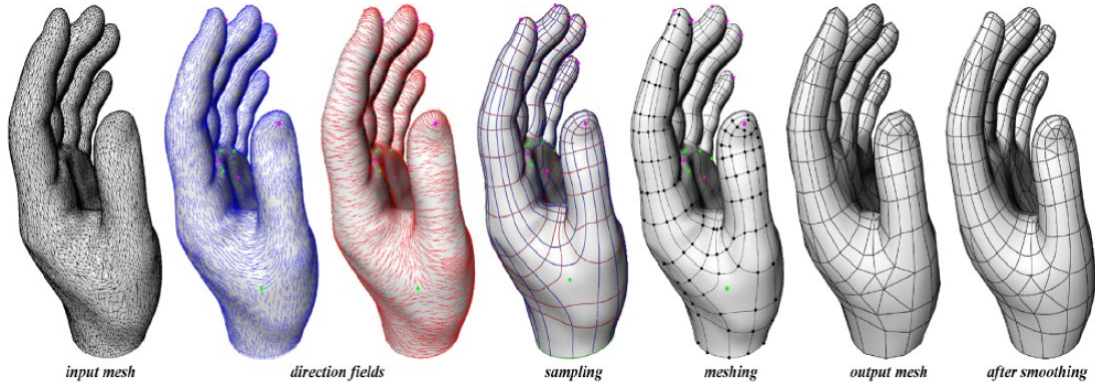


FIGURE 14.1: Manipulations sur des maillages

14.1 Maillage de Delaunay

14.1.1 Maillage simplicial

L'opération de maillage consiste à discrétiser un domaine (i.e. un milieu continu ou plutôt sa modélisation géométrique) par des éléments (éléments finis nous concernant), si possible bien proportionnés : au paragraphe 10.3, nous avons déjà présenté les dimensions géométriques représentatives d'un maillage que sont le diamètre maximum des éléments h et le facteur de forme du maillage σ , ainsi que le diamètre h_K d'un élément K et sa rondeur ρ_K . Ces deux derniers paramètres sont représentés sur la figure 10.2.

Dans ce chapitre, nous traiterons du cas bidimensionnel. Nous expliquerons les bases théoriques, mais ne rentrerons pas dans les détails pratiques de la programmation d'algorithmes de maillage (il y a de très bons cours disponibles sur le sujet).

Dans la suite nous aurons besoin des notions suivantes :

- un segment fermé (resp. ouvert) d'extrémités a et b de \mathbb{R}^d est noté $[a, b]$ (resp. $]a, b[$) ;
- un convexe E est un ensemble tel que : $\forall (a, b) \in E^2, [a, b] \subset E$;
- le convexifié d'un ensemble E de points de \mathbb{R}^d , noté $\mathcal{C}(E)$ est le plus petit convexe contenant E .
- un domaine Ω (i.e. un ouvert de \mathbb{R}^d) est dit polygonal si son bord $\Gamma = \partial\Omega$ est formé d'un nombre fini de segments ;
- un n -simplex (x_0, \dots, x_n) est le convexifié des $n + 1$ points de \mathbb{R}^d affine indépendant. Cela implique que $n \leq d$. Les sommets sont des 0-simplex, un segment est un 1-simplex, un triangle un 2-simplex, un tétraèdre est un 3-simplex.

Définition 64 — Maillage simplicial. Un maillage simplicial $\mathcal{T}_{d,h}$ d'un ouvert polygonal Ω_h de \mathbb{R}^d est un ensemble de d -simplex K^k de \mathbb{R}^d pour $k = 1 \dots N_t$, tel que l'intersection de deux d -simplex distincts \bar{K}^i et \bar{K}^j de $\mathcal{T}_{d,h}$ soit l'ensemble vide ou le p -simplex commun à \bar{K}^i et \bar{K}^j avec $p \leq d$.

En termes plus simples : le maillage $\mathcal{T}_{d,h}$ est constitué de N_t éléments K^k ($k = 1 \dots N_t$) appelés d -simplex (qui sont des triangles pour $d = 2$ et des tétraèdres pour $d = 3$) tels que l'intersection (de l'adhérence) de deux éléments \bar{K}^i et \bar{K}^j soit soit nulle (éléments parfaitement séparés), soit un point, une arête ou une face (i.e. un p -simplex avec $p \leq d$) commun aux deux éléments.

On notera $\mathcal{T}_{0,h}$ l'ensemble des sommets de $\mathcal{T}_{d,h}$, $\mathcal{T}_{1,h}$ l'ensemble de ses arêtes et $\mathcal{T}_{d-1,h}$ l'ensemble de ses faces. Le bord $\partial\mathcal{T}_{d,h}$ est l'ensemble des faces n'appartenant qu'à un seul d -simplex de $\mathcal{T}_{d,h}$.

Théorème 51 Pour tout ouvert polygonal Ω_h de \mathbb{R}^2 , il existe un maillage de cet ouvert sans sommet interne (voir figure 14.2).

Les sommets de ce maillage sont les points anguleux du bord $\partial\Omega_h$.

Malheureusement ce théorème n'est plus vrai en dimension plus grande que 2, car il existe des configurations d'ouvert polyédrique non-convexe qu'il est impossible de mailler sans point interne.

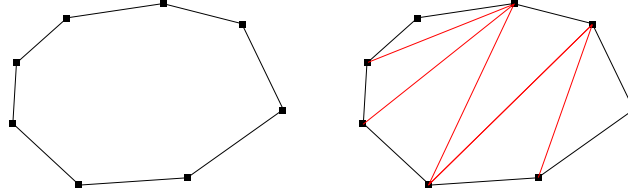


FIGURE 14.2: a) Points et arêtes et b) maillage de Delaunay sans point interne

14.1.2 Maillage de Delaunay-Voronoi

Dans le cas général, la construction d'un maillage nécessite de connaître :

- un ensemble de points \mathcal{S} ;
- un ensemble d'arêtes \mathcal{A} définissant le maillage de la frontière des sous-domaines.
- un ensemble, qui peut être vide, de sous-domaines \mathcal{D} à mailler.

Bien que les diagrammes de Voronoï existent en dimension quelconque, nous ne les présentons qu'en dimension 2. Aussi, les éléments précédents deviennent-ils, en dimension 2 :

$$\mathcal{S} = \{x^i \in \mathbb{R}^2, i = 1, \dots, N_p\} \quad (14.1)$$

$$\mathcal{A} = \{(s_1^i, s_2^i) \in \{1 \dots N_p\}^2, i = 1, \dots, N_a\} \quad (14.2)$$

$$\mathcal{D} = \{(a^i, \text{sens}^i) \in \{1 \dots N_a\} \times \{-1, 1\}, i = 1, \dots, N_d\} \quad (14.3)$$

i.e. une arête a^i est définie par ses deux sommets (s_1^i, s_2^i) qui sont des points de \mathcal{S} , et un sous-domaine est défini par une arête frontière a^i et un sens de parcourt (positif ou négatif).

Définition 65 Les diagrammes de Voronoï sont les polygones convexes $V^i, i = 1, \dots, N_p$ formés par l'ensemble des points de \mathbb{R}^2 plus proches de x^i que des autres points x^j , soit :

$$V^i = \{x \in \mathbb{R}^2 / \|x - x^i\| \leq \|x - x^j\|, \forall j \in \{1, \dots, N_p\}\} \quad (14.4)$$

Chaque point x^i (point générateur) étant considéré comme une île d'où partent des bateaux, la région de Voronoï du point x^i est la région où un bateau issu de l'île x^i arrive avant tout bateau issu d'une autre île. Notons qu'il est possible de jouer avec la métrique pour définir des diagrammes de Voronoï sur d'autres géométries que la géométrie euclidienne, et qu'il est possible également d'étendre la définition au cas où « les bateaux ne vont pas tous à la même vitesse ». Dans ce dernier cas, on parle de diagramme de Voronoï pondéré, que nous n'utiliserons pas dans ce document.

Les diagrammes de Voronoï sont des polygones obtenus comme intersections finies de demi-espaces et sont donc convexes. De plus, les sommets v^k de ces polygones sont à égale distance des points $\{x^{i^k}, j = 1, \dots, n_k\}$ de \mathcal{S} , où le nombre n_k est généralement égal ou supérieur à 3. À chacun de ces sommets v^k , nous pouvons associer le polygone convexe construit avec les points $\{x^{i^k}, j = 1, \dots, n_k\}$ en tournant dans le sens trigonométrique. Ce maillage est généralement formé de triangles, sauf si il y a des points cocycliques.

L'usage informel des diagrammes de Voronoï remonte à Descartes en 1644. Dirichlet a utilisé des diagrammes de Voronoï en dimension 2 ou 3 dans son étude des formes quadratiques en 1850. Le médecin britannique John Snow a utilisé un diagramme de Voronoï en 1854 pour montrer que la majorité des personnes mortes dans l'épidémie de choléra de Soho (à Londres) vivait plus près de la pompe infectée de Broad Street que de n'importe quelle autre pompe.

Les diagrammes de Voronoï portent le nom du mathématicien russe Georgy Fedoseevich Voronoï qui a défini et étudié le cas général en dimension n en 1908. Les diagrammes de Voronoï qui sont utilisés en géophysique et en météorologie pour analyser des données de distributions spatiales (comme les mesures de chutes de pluie) sont appelés polygones de Thiessen du nom du météorologiste américain Alfred H. Thiessen. Ils sont également très utiles en géométrie algorithmique, en particulier pour des problèmes de représentation ou de quantification, et sont utilisés dans le champ de la robotique pour créer un protocole pour éviter les obstacles détectés. Pour la modélisation de phénomènes naturels, ils servent pour les études de la compétition végétale (écologie et sylviculture), pour les territoires d'animaux (zoologie) et des clans et tribus néolithiques (anthropologie et archéologie), ainsi que pour les modèles de zones urbaines (géographie). Il explique aussi la répartition (et la forme) des tâches du pelage des girafes et des écailles de tortues.



Descartes Snow Voronoï



Delaunay

La triangulation de Delaunay a été inventée par le mathématicien russe Boris Delaunay en 1934. D'après la définition de Delaunay, le cercle circonscrit d'un triangle constitué de trois points de l'ensemble de départ est vide s'il ne contient pas d'autres sommets que les siens. Ainsi, les autres points sont autorisés sur le périmètre en lui-même mais pas à l'intérieur strict du cercle circonscrit. La condition de Delaunay affirme qu'un réseau de triangles est une triangulation de Delaunay si tous les cercles circonscrits des triangles du réseau sont vides. Ceci constitue la définition originale en deux dimensions. En remplaçant les cercles par des sphères circonscrites, il est possible d'étendre la définition à la dimension trois... mais en fait, on peut l'étendre en dimension quelconque. Les triangulations de Delaunay maximisent le plus petit angle de l'ensemble des angles des triangles, évitant ainsi les triangles allongés.

On parle souvent de la triangulation de Delaunay comme du dual du diagramme de Voronoï qui lui est associé. En fait les deux sont liés de la façon suivante :

- Les sommets du diagramme de Voronoï sont les centres des cercles circonscrits des triangles de la triangulation de Delaunay. Les arêtes du diagramme de Voronoï sont sur les médiatrices des arêtes de la triangulation de Delaunay ;
- Chaque germe (ou point générateur) du diagramme de Voronoï constitue un sommet dans la triangulation de Delaunay. Ces sommets sont reliés entre eux par une arête si et seulement si les cellules sont adjacentes.

Définition 66 — Maillage de Delaunay. On appelle maillage de Delaunay strict, le maillage dual des diagrammes de Voronoï, construit en reliant deux points x^i et x^j , si les diagrammes V^i et V^j ont un segment en commun.

Pour rendre le maillage triangulaire, il suffit de découper les polygones qui ne sont pas des triangles en triangles. Nous appelons ces maillages des maillages de Delaunay de l'ensemble \mathcal{S} .

Le domaine d'un maillage de Delaunay d'un ensemble de points \mathcal{S} est l'intérieur du convexe $\mathcal{C}(\mathcal{S})$ de l'ensemble de points \mathcal{S} .

Il existe une propriété qui permet de savoir si un maillage est un maillage de Delaunay. C'est la propriété de la boule ouverte.

Théorème 52 — Propriété de la boule ouverte. Un maillage $\mathcal{T}_{d,h}$ est un maillage de Delaunay s'il est tel que pour tout triangle T du maillage, le disque ouvert $D(T)$ correspondant au cercle

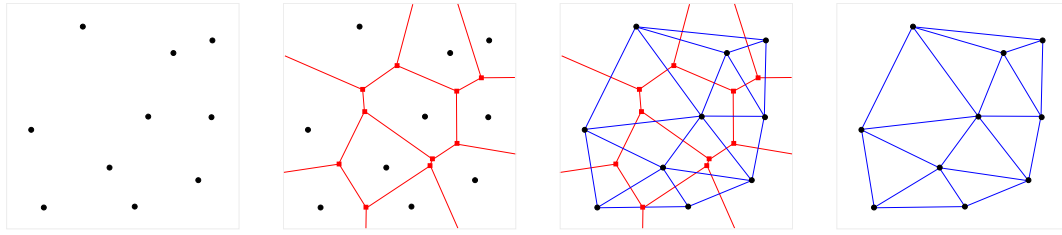


FIGURE 14.3: a) points, b) ajout du diagramme de Voronoï, c) ajout de la triangulation de Delaunay et d) maillage seul

circonsrit à T ne contient aucun sommet :

$$D(T) \cap \mathcal{T}_{0,h} = \emptyset \quad (14.5)$$

Réciproquement, si le maillage $\mathcal{T}_{d,h}$ d'un domaine convexe vérifie la propriété de la boule ouverte, alors il est de Delaunay.

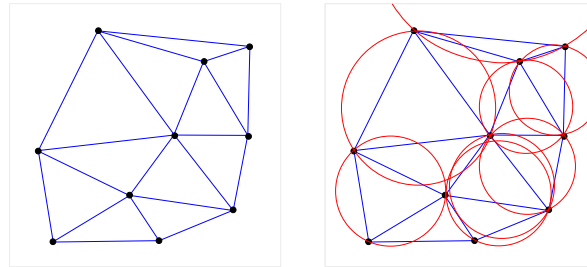


FIGURE 14.4: a) Triangulation de Delaunay et b) cercles circonsrits : propriété de la boule ouverte

La propriété de la boule ouverte peut être appliquée également au cas d'un quadrilatère convexe, puisque celui-ci peut être découpé en deux triangles adjacents. Il suffit alors de vérifier la propriété pour toutes les paires de triangles adjacents formant un quadrilatère convexe.

Si le critère de la boule ouverte n'est pas vérifié pour un quadrilatère convexe, alors on fera un échange de diagonal pour résoudre le problème.

14.1.3 Remarques

À ce niveau, nous sommes en mesure de générer un maillage... toutefois, celui-ci n'est pas forcément exempt de problèmes :

- si les seuls points disponibles sont sur la frontière, il va falloir générer des points internes au maillage. En effet, le maillage généré existe, mais peut présenter des distorsions inadmissibles en termes de calcul. Le critère le plus naturel pour distribuer les points internes est d'imposer en tout point x de \mathbb{R}^2 , le pas de maillage $h(x)$. En pratique, on ne dispose pas toujours de cette information et il faut la construire à partir des points disponibles, i.e. des points de la frontière. D'autre part, dans de nombreuses applications, on préfère donner le nombre de subdivisions souhaitées entre deux sommets : un peu de géométrie différentielle (longueurs d'arcs) permet de se ramener à donner une valeur à $h(x)$;
- il existe des cas où le maillage généré peut ne pas respecter la frontière (par exemple pour un domaine en forme de « U » avec peu de points. Il existe une solution de forçage de la frontière par permutation d'un certain nombre de diagonales. Voir figure 14.5 ;
- il faut traiter le cas des trous... mais ce n'est pas réellement difficile.

Comme il ne s'agit pas ici de donner un cours sur les algorithmes de maillage, nous en resterons là. Les maillages triangulaires sont plutôt bien maîtrisés. Il y a de nombreux cas à traiter pour couvrir toutes les configurations existantes, mais les bases théoriques et les réalisations pratiques existent dans tous les cas.

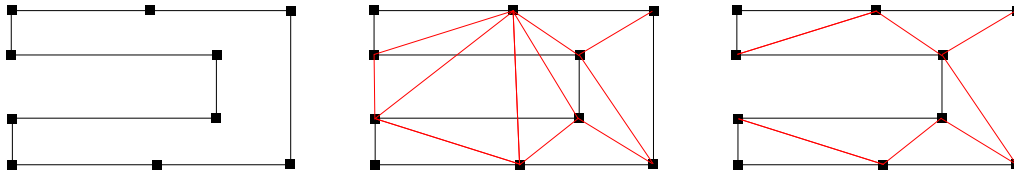


FIGURE 14.5: a) Points et arêtes, b) maillage de Delaunay ne respectant pas la frontière, c) maillage respectant la frontière obtenu en échangeant des diagonales

14.2 Maillage par avancement de fronts

Le front initial est constitué de la frontière (ou des frontières, s'il y a des trous), i.e. des nœuds et des arêtes. L'algorithme est très simple : pour chaque arête représentée par le segment $[x^i, x^j]$, on crée un nouveau point x^{N_p+1} tel que le triangle formé des trois points x^i, x^j et x^{N_p+1} soit équilatéral.

Évidemment il faut quelques règles supplémentaires pour que cela fonctionne bien :

- on met à jour en permanence le front dès qu'un point est créé, et on ne s'arrête que lorsque toutes les arêtes du front ont été balayées ;
- en cours de calcul, de nouveaux fronts peuvent apparaître (cela correspond à plusieurs domaines non encore maillés à l'intérieur du maillage en cours) ;
- on vérifie si des nœuds du front courant ne seraient pas candidats pour être le nouveau point x^{N_p+1} du triangle : par exemple, en s'assurant qu'un nœud existant n'appartient pas au cercle de centre le point x^{N_p+1} théorique (triangle équilatéral) et de rayon un certain paramètre fixé ;
- lorsque plusieurs nœuds existants sont candidats, on sélectionne celui qui fournit le triangle le plus équilatéral possible ;
- on supprime toutes les possibilités qui intersectent un front existant (pas de recouvrement d'éléments) ;
- on rejette les triangles inversés.

14.3 Maillage par transformation

Le principe est on ne peut plus simple. Il s'agit de mettre en bijection deux domaines : d'une part le domaine à mailler compliqué Ω_h , et d'autre part un domaine de référence plus simple : rectangle, sphère...

Pour faire simple, on peut imaginer mailler une ellipse à partir du maillage d'un disque, un rectangle à partir du maillage d'un carré... on peut dire que l'on fait du morphing sur un maillage. Dans la pratique, on part de la surface d'un domaine tridimensionnel complexe. On transforme cette surface en une surface plane par une certaine transformation. On maille cette nouvelle surface dans le plan. Puis on effectue la transformation inverse afin de disposer d'un maillage de la surface dans \mathbb{R}^3 .

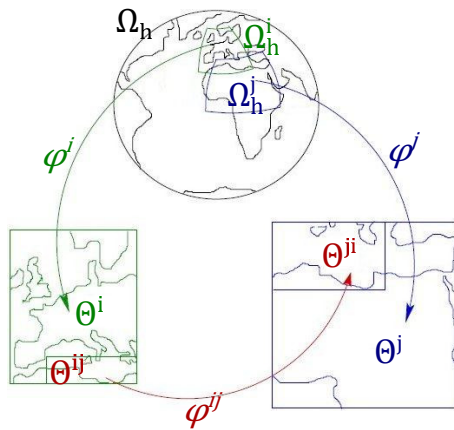


FIGURE 14.6: Domaines et transformations

Considérons que nous souhaitons mailler notre domaine Ω_h qui est une surface dans \mathbb{R}^2 ou \mathbb{R}^3 . On appelle transformation φ , un homéomorphisme transformant Ω_h en un autre domaine (plus simple par exemple, ou au moins plan si la surface de Ω_h est dans \mathbb{R}^3) noté Θ . Comme nous utilisons des homéomorphismes, les transformations inverses existent et sont continues. Nous sommes donc en mesure de repasser de Θ à Ω_h .

Dans le cas général, illustré à la figure 14.6, ce domaine Ω_h peut lui-même être déjà décomposé en un certain nombre de sous-domaines Ω_h^i avec recouvrement partiel des sous-domaines (les Ω_h^i sont des ouverts de la variété Ω_h). On appelle transformations φ^i , les homéomorphismes transformant Ω_h^i en les sous-domaines Θ^i (avec recouvrements partiels) constituant Θ . On appelle fonctions de transition φ^{ij} les transformations permettant de passer du sous-domaine Θ^{ij} de Θ^i correspondant à la zone de recouvrement dans Θ^i avec Θ^j au sous-

domaine Θ^{ji} de Θ^j correspondant à la zone de recouvrement dans Θ^j avec Θ^i . Cela permet de s'assurer de la bonne transition entre les cartes locales Θ^i (les Θ^i forment un atlas de cartes).

La transformation considérée peut conserver une ou plusieurs des propriétés suivantes :

- les angles : transformation conforme ;
- les aires : transformation authalique ;
- les longueurs : transformation isométrique.

Un cas « simple » est celui où Ω_h est une surface de genre zéro sans bord (une telle surface correspond à la surface d'un volume sans trou, c'est donc quelque chose de très général). Une surface de genre zéro sans bord est homéomorphe à la sphère unité S^2 et l'on peut donc trouver un homéomorphisme φ , qui peut être compliqué, entre Ω_h et la sphère, i.e. on est dans le cas précédent où l'on passe directement de Ω_h à S^2 par un unique homéomorphisme φ . Il ne reste plus ensuite qu'à mailler la sphère, puis à envoyer tous les sommets du maillage sur Ω_h par φ^{-1} , tout en conservant la même connectivité. Il faut toutefois veiller à mailler S^2 assez finement pour s'assurer que le maillage de Ω_h ne s'auto-intersecte pas. Ainsi, la densité locale des sommets du maillage de S^2 doit dépendre de Ω_h , et donc de φ .

Le maillage de la sphère n'est pas si simple qu'on le pense. On peut là encore appliquer la même technique : définir un homéomorphisme entre la demi-sphère et le disque, mailler le disque. Le maillage de la sphère, obtenu par symétrie de la demi-sphère doit alors « se recoller », du moment que le maillage du bord du cercle est conforme et laissé conforme par l'homéomorphisme. Si l'on procède au maillage du cercle par la méthode de Delaunay telle que décrite en partant du contour du cercle, on est alors assuré que le maillage du contour sera conforme.

Dans les deux cas, passage du cercle à la demi-sphère et passage de la sphère à Ω_h , on pourra remarquer que la densité du maillage n'est généralement pas la même. Il s'agit du phénomène d'étirement. Si l'on souhaite que la surface finale Ω_h ait une densité de maillage fixée, il faut savoir comment φ^{-1} modifie cette densité pour pouvoir mailler S^2 en conséquence... idem pour le cercle. Cela n'est pas simple, surtout si Ω_h est compliqué.

14.4 Remarques sur le maillage quadrangulaire et hexaédrique

Même en 2D, il est plus difficile de générer un maillage quadrangulaire que triangulaire. Une raison fondamentale est que tout polygone ne peut être décomposé en un ensemble de quadrangles : il peut être nécessaire d'ajouter des sommets sur le bord. Et en pratique, il est nécessaire et suffisant d'avoir un nombre pair d'arêtes sur le bord géométrique.

En 3D, on ne peut pas mailler tout volume dont le bord est maillé par une surface quadrangulaire. Disposer d'un nombre pair de quadrangles sur le bord n'est qu'une condition nécessaire. Une condition suffisante est d'avoir un nombre pair de quadrangles maillant une surface topologiquement équivalente à une sphère. Dans ce cas, le volume délimité par la surface en question admet un maillage hexaédrique.

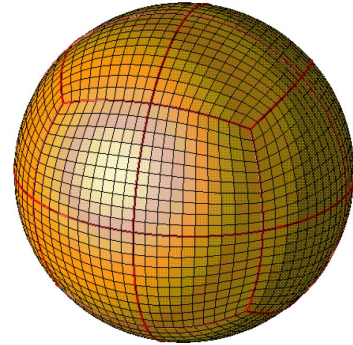


FIGURE 14.7: Maillage en quadrangles d'aires égales

Le maillage d'une sphère en quadrangles de même aire est un problème difficile qui ressort du problème de Tammes, ou problème des dictateurs. Une solution a été donnée par Lemaire et Weill en 2000 [54]. Un maillage obtenu est donné à la figure 14.7. Un tel maillage comporte $24n^2$ éléments par construction : la sphère est « découpée » par son cube inscrit et chacun des six morceaux obtenus est divisé en quatre par symétrie. Toutefois, il faut noter que les quadrangles considérés ne sont pas plans, ce sont des morceaux de sphères. Pour l'instant il n'y a pas de solution « parfaitement satisfaisante » avec des quadrangles plans, dans un sens qui reste à définir... puisque l'invariant d'Euler-Poincaré implique par exemple qu'il n'y a pas de polygonalisation de la sphère (qui est de caractéristique d'Euler-Poincaré de 2) avec uniquement des rectangles (ou des hexagones d'ailleurs). Au passage, rappelons que les triangles permettent eux une polygonalisation d'une surface, quelque soit sa caractéristique, d'où leur succès... mais attention à réaliser un maillage « pas trop trivial » afin d'éviter le cas de triangles trop distordus venant se connecter sur un même sommet et conduisant à des singularités en terme de calcul numérique (et pour cela, utilisons donc la triangulation de Delaunay-Voronoi).

Aucune méthode n'est apte à traiter tous les cas possibles concernant le maillage avec des quadrangles ou des hexaèdres. Néanmoins de nombreux algorithmes sont disponibles :

- Les maillages structurés (obtention de la répétition régulière de motifs identiques) fonctionnent bien mais ne sont pas adaptés à toutes les géométries.
- On peut subdiviser le domaine en plusieurs parties pour appliquer sur chacune d'elles un maillage structuré ou un maillage obtenu par déformation d'un maillage structuré.
- On peut définir deux faces topologiquement équivalentes du modèle comme les surfaces source et cible et découper les surfaces de liaisons entre la source et la cible pour créer un maillage structuré dans une direction.
- On peut raisonner comme en CAO en générant des maillages par extrusion ou par révolution.
- Le paving est une variante de l'avancée de front. On essaye de faire un maillage de type structuré en reportant un front « parallèle » au front existant, avec le même nombre de nœuds. Il sera nécessaire d'avoir des distorsions à certains endroits et autour des trous, il faudra faire des coutures.
- Le Q-morph se propose de transformer un maillage triangulaire en maillage quadrangulaire par avancée de front. Cela ne fonctionne pas en 3D.
- Les méthodes basées sur une grille considèrent la surface ou le volume découpé(e) en carrés ou cubes selon une grille prédéfinie. Il suffit alors d'adapter les éléments situés au bord du domaine : ceux-ci seront de moins bonne qualité et parfois même pas totalement conformes au bord théorique.

Le maillage quadrangulaire et surtout hexaédrique reste un domaine de recherche actif.

Chapitre 15

Homogénéisation

Résumé — L'idée bien connue de l'homogénéisation est de remplacer un milieu « compliqué » par un milieu équivalent simple afin de simplifier le modèle numérique à résoudre. L'intérêt de ces techniques est donc évident.

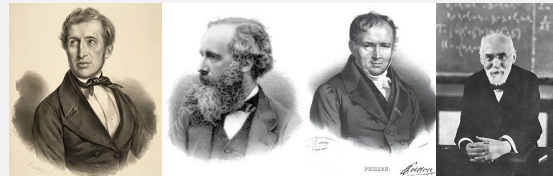
Par exemple, un matériau composite composé de plusieurs plis (couches) constituées chacune de fibres noyées dans une matrice et dont l'orientation diffère d'une couche à l'autre peut être représenté avantageusement par un matériau « homogénéisé » ou « équivalent ».

Nous allons présenter les méthodes d'homogénéisation ainsi que leurs applications en mécanique et acoustique, mais également en modélisation.

Histoire

Les *théories des milieux effectifs* visent à estimer les propriétés *effectives* (i.e. macroscopiques) d'un milieu en fonction des propriétés locales de chacun des constituants ainsi que d'un certain nombre d'informations sur la microstructure.

Les premières théories remontent au XIX^e siècle et sont dues à Mossotti, Maxwell, Poisson ou encore Lorentz. Le but de ces modèles est de fournir soit des bornes pour le comportement effectif, soit des approximations du comportement effectif. Les bornes sont optimales lorsqu'il existe une microstructure particulière qui réalise exactement le modèle physique. On évalue les « bonnes » propriétés de ces théories en les confrontant à d'autres résultats théoriques, calculs analytiques et numériques...



Mossotti

Maxwell

Poisson

Lorentz

Ces théories sont utilisées pour les problèmes de conductivité (milieux diélectriques), en mécanique, magnétique, thermique... lorsque l'on a des phases de conductivité, d'élasticité, des coefficients thermiques... variables. Ces problèmes sont en général très difficiles à résoudre (non-linéaires et anisotropes) alors qu'en même temps, du point de vue des applications pratiques, il n'est pas forcément nécessaire de tenir compte de l'ensemble des degrés de liberté de ces systèmes.

L'existence d'un *comportement effectif* n'est nullement assurée. On montre que, sous certaines hypothèses (en particulier l'existence d'un *volume élémentaire représentatif*), on peut effectivement remplacer un matériau hétérogène par un milieu homogène équivalent.

Enfin, d'un point de vue purement numérique, ces méthodes peuvent être utilisées pour « simplifier » un système à résoudre, que cela ait un sens physique ou non.

Notons que les techniques d'homogénéisation ne sont pas que des artefacts, des trucs et astuces. Certaines des grandeurs physiques que nous utilisons tous les jours ne sont que des moyennes. Le meilleur exemple en est la pression. Bien qu'en un point donné d'un gaz on voit passer des particules allant en tous sens, on constate également, à une échelle plus macroscopique, un schéma d'ensemble qui permet de définir par exemple la pression qu'exerce ledit gaz sur une paroi... pourtant rien de « cohérent » ne se dégage à l'échelle microscopique.

Le but de ce chapitre est donc d'effectuer le passage du niveau microscopique au niveau macroscopique (historiquement les premières homogénéisations, appelées alors moyennisations, utilisaient les moyennes arithmétique et harmonique) en fournissant la justification de ce passage (existence et unicité) ainsi que la formule (l'algorithme) de calcul des coefficients efficaces.

Les premières études asymptotiques des problèmes aux limites non-stationnaires aux coefficients oscillants ont été développés par l'école russe, notamment par Bakhvalov en 1975, pour des résultats sur \mathbb{R}^n tout entier, puis par Panasenko en 1978, qui a étendu ces résultats au cas d'un problème aux limites dans un ouvert borné. Pour ce faire, Panasenko a introduit des correcteurs de couches limites, explicitement construits dans le cas particulier d'une seule couche. La notion de convergence à deux échelles a été introduite en 1989 par Nguetseng et ensuite développée par Allaire avec pour application l'homogénéisation périodique. Elle a été généralisée au cas de quelques problèmes multi-échelles. Les quatre principales méthodes d'homogénéisation périodique sont la méthode des échelles multiples, la méthode des fonctions tests oscillantes, la méthode de la convergence à double échelle et la méthode de l'éclatement périodique.

Les méthodes que nous présenterons dans ce chapitre seront les méthodes de développement régulier, de la couche limite, et de développement asymptotique. Le cas des milieux poreux sera ensuite considéré. Enfin, nous mentionnerons une application de ces méthodes pour réduire la dimension de certains problèmes. La notion de convergence à deux échelles sera utilisée au chapitre 16 pour l'optimisation topologique.

Une dernière remarque avant d'entrer dans le vif du sujet, car elle correspond à des interrogations de nombreux étudiants : il n'est pas nécessaire qu'un problème soit périodique pour pouvoir recourir aux techniques d'homogénéisation. En effet, l'homogénéisation consiste à remplacer un problème « compliqué » par un « plus simple » selon une certaine « mesure » : il s'agit donc généralement de montrer que l'intégrale d'un truc compliqué peut s'écrire comme une autre intégrale d'un truc plus simple... et c'est l'égalité de ces deux intégrales qui fait que l'on dit que les problèmes sont équivalents (conduisent à la même solution), et donc que le second problème correspond à l'homogénéisation du premier. Dans le même ordre d'idée, et pour enfoncer le clou, un matériau homogénéisé n'est pas forcément isotrope. En effet, on peut remplacer un matériau fortement anisotrope par un matériau homogénéisé isotrope, mais cela n'est pas forcément le plus adapté : il peut être préférable de l'homogénéiser en un matériau lui-même anisotrope mais « plus simple ». Un matériau composite composé de plusieurs plis ayant des orientations différentes pourrait être remplacé par exemple par un matériau composite homogénéisé n'ayant qu'une couche (donc plus simple), mais restant anisotrope (qui serait, en quelque sorte, orienté selon une direction « homogénéisée », et il est alors plus « facile » de raisonner avec cette orientation « principale » lorsque l'on souhaite physiquement comprendre le comportement du matériau).

15.1 Méthodes d'homogénéisation

Nous allons considérer le problème de conductivité linéaire donné par l'équation de Poisson avec condition de Dirichlet. Nous rappelons qu'il s'agit du problème suivant (qui décrit le déplacement vertical u d'une membrane plane) :

$$\begin{cases} \operatorname{div}(A(x)\nabla u) = f(x) & \text{pour } x \in \Omega \\ u = 0 & \text{sur } \Gamma = \partial\Omega \end{cases} \quad (15.1)$$

On rappelle également que si Ω est un ouvert connexe borné de \mathbb{R}^n , et $H_0^1(\Omega)$ l'espace de Sobolev des fonctions qui s'annulent sur Γ , alors la solution $u \in H_0^1(\Omega)$ du problème précédent est également la solution du problème faible :

$$\forall v \in H_0^1(\Omega), \quad - \int_{\Omega} A(x) \nabla u \cdot \nabla v = \int_{\Omega} f(x) v \quad (15.2)$$

où $f \in L^2(\Omega)$. Pour que ces deux formulations soient équivalentes, on a supposé que $A(x)$ est régulière : i.e. que pour chaque x de Ω , $A(x)$ est une matrice $n \times n$, dont les éléments sont mesurables et qui est symétrique, définie-positive, bornée, i.e. qu'il existe deux constantes K_1 et $K_2 > 0$ telles que :

$$\forall v \in \mathbb{R}^n, \quad K_1 v \cdot v \leq A(x) v \cdot v \leq K_2 v \cdot v \quad (15.3)$$

Mentionnons le cas particulier où $A(x)$ est régulier dans Ω mais pas sur Γ , alors on a deux conditions d'interface :

$$[u] = 0 \quad \text{et} \quad [n \cdot A(x) \nabla u] = 0 \quad (15.4)$$

Nous savons également que la solution existe et est unique (d'après le théorème de Lax-Milgram, i.e. d'après le théorème de Riesz-Fréchet en remarquant le produit scalaire qui va bien...)

Intéressons nous maintenant d'un peu plus près à $A(x)$. Supposons que $A(x)$ possède une certaine périodicité ε , par exemple, que le matériau constituant le domaine soit constitué de couches successives de deux matériaux différents, mais avec une périodicité, comme illustré sur la figure 15.1. Dans un tel cas, nous dirons que les coefficients A dépendent de x/ε . Cette écriture sous forme

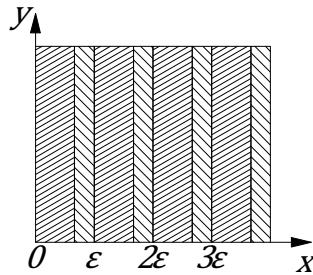


FIGURE 15.1: Matériau périodique

normalisée permet de dire que A est périodique de période 1.

Dans le cas unidimensionnel, le problème devient :

$$\begin{cases} \frac{d}{dx} \left(A \left(\frac{x}{\varepsilon} \right) \right) \frac{du}{dx} = f(x) & \text{pour } x \in [0, l] \\ u = 0 & \text{pour } x = 0 \text{ et } x = l \end{cases} \quad (15.5)$$

A est une fonction de \mathbb{R} dans \mathbb{R} de période 1, et nous la supposons régulière par morceaux et telle qu'il existe K_1 et $K_2 > 0$ telles que $K_1 \leq A(v) \leq K_2$.

La solution du problème est alors :

$$u(x) = \int_0^x A^{-1}(y/\varepsilon) \left(\int_0^y f(v) dv + c_1 \right) dy + c_2 \quad (15.6)$$

avec $c_2 = 0$ et

$$c_1 = \frac{\int_0^l A^{-1}(y/\varepsilon) \int_0^y f(v) dv dy}{\int_0^l A^{-1}(y/\varepsilon) dy} \quad (15.7)$$

Le nombre d'intervalles pour le calcul approché de c_1 est très grand (i.e. très supérieur à l/ε).

Exprimé autrement, ce résultat devient : le coefficient homogénéisé est $1/\langle \frac{1}{A} \rangle$ et non pas $\langle A \rangle$ (où $\langle \cdot \rangle$ désigne la moyenne).

15.1.1 Méthode de développement régulier

Dans la méthode de développement régulier, on se propose de chercher la solution asymptotique $u^{(\infty)}$ sous la forme :

$$u^{(\infty)} \sim \sum_{i=0}^{\infty} \varepsilon^i u_i(x) \quad (15.8)$$

où $u_i(x)$ ne dépend pas de ε .

La solution asymptotique $u^{(\infty)}$ est bien de la forme précédente si en injectant cette expression dans celle du problème alors on obtient une petite erreur au second membre.

Pour le calcul on procède donc comme suit :

- on injecte la forme souhaitée (cette forme s'appelle l'ansatz) ;
- pour avoir un second membre avec un terme d'erreur petit, on obtient la forme des $u_i(x)$;
- on vérifie que les $u_i(x)$ trouvés conviennent bien.

15.1.2 Méthode de la couche limite

Considérons le problème unidimensionnel :

$$\begin{cases} \varepsilon^2 u'' - p(x)u = f(x) & \text{pour } x \in [0; 1] \\ u(0) = A \text{ et } u(1) = B \end{cases} \quad (15.9)$$

où $p \in C([0; 1])$; $p(x) \geq 0$, $x \in [0; 1]$; $f \in C([0; 1])$; $A \in \mathbb{R}$, $B \in \mathbb{R}$ et $\varepsilon > 0$ le petit paramètre. En négligeant le terme $\varepsilon^2 u''$, nous arrivons à l'équation $-p(x)u_r = f(x)$. En d'autres termes,

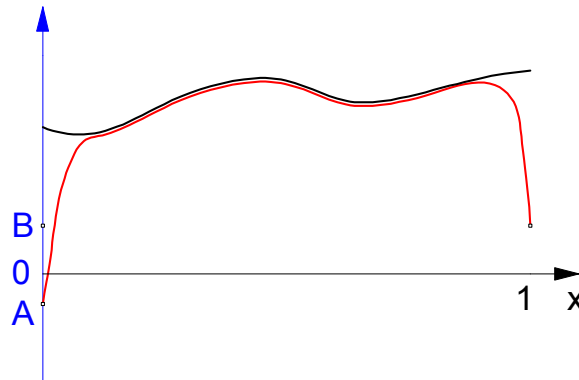


FIGURE 15.2: Méthode de la couche limite

et comme montré sur la figure 15.2, nous disposons de la courbe correspondant à $y = u_r = -f(x)/p(x)$ (en noir) qui approche la solution du problème donnée par la courbe rouge.

La solution exacte se confond avec u_r sur la plus grande partie de l'intervalle $[0; 1]$, mais u et u_r sont fortement différents dans un voisinage des extrémités.

Cherchons une solution asymptotique sous la forme de l'ansatz :

$$u^{(\infty)} \sim \sum_{i=0}^{\infty} \varepsilon^i (u_i^r(x) + u_i^0(x/\varepsilon) + u_i^1((x-1)/\varepsilon)) \quad (15.10)$$

où les $u_i^r(x)$ sont les termes réguliers, et $u_i^0(x/\varepsilon)$ et $u_i^1((x-1)/\varepsilon)$ les termes de couche limite correspondant aux extrémités $x = 0$ et $x = l$.

Pour le calcul on procède donc comme suit :

- on injecte les termes réguliers de l'ansatz dans la formulation du problème ;
- on traite les extrémités pour vérifier les conditions aux limites, ce qui donne les termes de couche limite ;
- on vérifie que la solution trouvée conviennent bien (i.e. que l'erreur commise est négligeable).

15.1.3 Méthode de développement asymptotique infini

À ce niveau du texte, on a dû s'apercevoir que la démarche était la même pour les deux méthodes précédentes, sauf évidemment la forme de l'ansatz. Dans la méthode de développement asymptotique infini, on va encore procéder de la même manière, mais on cherchera une solution sous la forme d'ansatz :

$$u^{(\infty)} \sim \sum_{i=0}^{\infty} \varepsilon^i u_i(x, x/\varepsilon) \quad (15.11)$$

où chaque $u_i(x, x/\varepsilon)$ est de la forme $u_i(x, x/\varepsilon) = N_i(x/\varepsilon)v_\varepsilon^{(i)}(x)$, $\forall i > 0$ et $u_0(x, x/\varepsilon) = v_\varepsilon(x)$, i.e. :

$$u^{(\infty)} \sim \sum_{i=0}^{\infty} \varepsilon^i N_i(x/\varepsilon)v_\varepsilon^{(i)}(x) \quad (15.12)$$

et où les $N_i(x/\varepsilon)$ sont 1-périodiques et $N_0 = 1$.

Pour un détail plus mathématique, on pourra se référer par exemple à [1] qui présente la notion de convergence à deux échelles (« Two-scale convergence »). Cet article, bien qu'un peu technique, reste relativement abordable.

15.1.4 Cas des coefficients discontinus

Il est possible d'appliquer la même approche que lorsque les coefficients sont continus avec quelques modifications : il « suffit » d'ajouter les conditions d'interfaces présentées en début de paragraphe et rappelées ci-dessous :

$$[u] = 0 \quad \text{et} \quad [n \cdot A(x/\varepsilon)\nabla u] = 0 \quad (15.13)$$

partout où cela est nécessaire.

On retrouve les mêmes conditions d'existence de la solution que dans le cas continu et la périodicité de la solution homogénéisée reste inchangée.

15.2 Homogénéisation simplifiée pour les matériaux composites

Après ce petit tour « mathématique » des méthodes d'homogénéisation, nous proposons un petit complément très mécanique, et très « pratique ».

Pour information, la modélisation du muscle cardiaque s'inspire des techniques d'homogénéisation des matériaux composites.

15.2.1 Introduction

D'une manière générale, tous les matériaux, mêmes isotropes, sont hétérogènes en dessous d'une certaine échelle. Il peut sembler naturel d'utiliser des propriétés homogènes équivalentes correspondant à des propriétés mécaniques effectives. Toutefois, comme nous l'avons déjà vu, ces propriétés effectives ne s'obtiennent pas par une simple moyenne des propriétés des constituants, mêmes pondérées par les fractions volumiques.

Les propriétés effectives du milieu homogène équivalent cherché peuvent être obtenues en résolvant un problème aux limites sur un volume élémentaire dV , à condition que celui-ci soit suffisamment grand pour être représentatif de la microstructure du matériau hétérogène. Dans le cas où les constituants présentent une structure périodique, le volume dV peut se réduire à un volume élémentaire.

Une fois le volume élémentaire déterminé, on le soumet à des sollicitations élémentaires pour déterminer la réponse résultante. La difficulté réside en fait dans le choix des conditions aux limites à appliquer au volume élémentaire considéré pour imposer une déformation ou contrainte globale moyenne donnée. Dans le cas linéaire, on peut prouver l'existence et l'unicité pour les différents cas de conditions aux limites existants.

Principalement pour les composites stratifiés ou sandwichs, il y a deux niveaux d'homogénéisation :

- du niveau micromécanique au niveau mésoscopique : Les hétérogénéités de base sont les fibres et la matrice. On effectue ici une étape d'homogénéisation locale.
- du niveau mésoscopique au niveau macroscopique : Les hétérogénéités de base sont les différentes couches du stratifié. Ces couches sont considérées comme « homogènes » (étape précédente). Cette fois, il s'agit d'une homogénéisation dans l'épaisseur du stratifié.

On pourra par exemple se reporter à [111] pour voir comment cela est utilisé pour développer un modèle de plaque homogénéisée équivalente à un matériau fait de carton ondulé entre deux peaux de carton.

15.2.2 Loi des mélanges, bornes de Voigt et de Reuss

On considère un composite UD (unidirectionnel) de repère d'orthotropie (l, t) , constitué de fibres noyées dans une matrice polymère. Soit une cellule élémentaire de fraction volumique $V = 1$ constituée de fibres et de matrice. On note V_m la fraction volumique de matrice, V_f la fraction volumique de fibre, et on a :

$$V = V_m + V_f = 1 \quad (15.14)$$

À l'échelle locale, on fait les hypothèses suivantes :

- Fibres : comportement élastique linéaire fragile isotrope de coefficients E_f et ν_f ;
- Matrice : comportement élastique non-linéaire, isotrope de coefficients E_m et ν_m .

On souhaite déterminer les relations existant entre E_l , E_t , E_f , E_m , V_m et V_f . Pour cela, on fait également les hypothèses suivantes :

- On travaille en élasticité linéaire.
- La liaison fibres/matrice est parfaite.
- Localement, on a : $\sigma_f = E_f \varepsilon_f$ et $\sigma_m = E_m \varepsilon_m$.

Loi des mélanges (ou modèles à bornes ou de Voigt et de Reuss) :

- premier essai — Il s'effectue dans la direction parallèle aux fibres (compression longitudinale)

$$E_{\text{longitudinal}} = E_l = E_f V_f + E_m V_m \quad (15.15)$$

C'est la loi des mélanges, qui est bien vérifiée dans la direction des fibres. Il s'agit de la borne supérieure de Voigt (1887).

- deuxième essai — Il s'effectue dans la direction perpendiculaire aux fibres (compression transversale)

$$\frac{1}{E_{\text{transverse}}} = \frac{1}{E_t} = \frac{V_f}{E_f} + \frac{V_m}{E_m} \quad (15.16)$$

C'est la loi des mélanges en souplesse. Cette relation n'est pas très bien vérifiée transversalement mais donne une indication sur la borne inférieure, dite de Reuss (1929).

- Module de cisaillement et coefficient de Poisson d'un UD par la loi des mélanges :

$$\nu_{lt} = \nu_f V_f + \nu_m V_m \quad \text{et} \quad \frac{1}{G_{lt}} = \frac{V_f}{G_f} + \frac{V_m}{G_m} \quad (15.17)$$

Les modèles à bornes fournissent un encadrement du comportement mécanique du matériau composite par des comportements mécaniques limites (bornes). Ils sont obtenus par la résolution du problème de l'élasticité linéaire sous forme faible. La minimisation de l'énergie potentielle conduit à la borne supérieure de Voigt. la résolution en contrainte conduit à la borne inférieure de Reuss. Pour gagner un peu en généralité, on peut remplacer les termes fibres et matrice par des phases, car ces modèles sont applicables à des mélanges de polymères (matériaux composés) et à des composites chargés par des particules diverses.

Les bornes correspondent aux associations série des deux phases (borne inférieure de Reuss, équivalent au modèle du module transverse équivalent de la loi des mélanges) et parallèle (borne supérieure de Voigt, équivalent au modèle du module longitudinal équivalent de la loi des mélanges).

Aucune hypothèse n'est faite sur la morphologie du matériau. Il est simplement admis que pour le modèle de Reuss, la contrainte est homogène dans les deux phases (continuité de la contrainte) et, pour le modèle de Voigt, la déformation est constante (continuité de la déformation) dans tout le composite. L'intérêt est limité dès que l'écart des caractéristiques des deux phases est important. Évidemment, d'autres modèles existent :

- Hashin et Shtrikman (1963) : resserre les bornes de Reuss et Voigt
- Takayanagi (combinaison de Reuss et Voigt)
- Halpin - Tsai : pour le renforcement par des fibres courtes alignées
- Tsai - Pagano : fibres courtes
- Halpin - Kardos : extension de la précédente

15.3 Homogénéisation des matériaux poreux

Considérons à nouveau notre problème de Poisson avec conditions de Dirichlet pour un milieu poreux. Ω est le domaine (borné) de \mathbb{R}^n qui contient Ω_ε l'ensemble des trous périodiques. Nous avons donc :

$$\begin{cases} \operatorname{div}(A(x)\nabla u) = f(x) & \text{pour } x \in \Omega \setminus \Omega_\varepsilon \\ u = 0 & \text{sur le contour extérieur } \Gamma = \partial\Omega \end{cases} \quad (15.18)$$

Sur le bord des trous, on peut imposer, soit des conditions de Dirichlet :

$$u = 0 \quad (15.19)$$

soit des conditions de Neumann :

$$n \cdot A(x/\varepsilon)\nabla u = 0 \quad (15.20)$$

On supposera encore que les coefficient sont 1-périodiques.

Cherchons des solutions dans le cas où l'on a des conditions de Neumann sur le bords des trous.

La forme choisie pour l'ansatz est le développement au second ordre suivant :

$$u^{(2)} = u_0(x, x/\varepsilon) + \varepsilon u_1(x, x/\varepsilon) + \varepsilon^2 u_2(x, x/\varepsilon) \quad (15.21)$$

où les $u_i(x, x/\varepsilon)$ sont 1-périodiques en x/ε qui sera noté ξ .

On obtient un système du type :

$$\begin{cases} L_{\xi\xi}u_0 & = 0 \\ L_{\xi\xi}u_1 + L_{\xi x}u_0 + L_{x\xi}u_0 & = 0 \\ L_{\xi\xi}u_2 + L_{\xi x}u_1 + L_{x\xi}u_1 + L_{xx}u_0 & = f(x) \end{cases} \quad (15.22)$$

et les conditions de Neumann :

$$n \cdot A(\xi) [\varepsilon^{-1} \nabla_\xi u_0 + \varepsilon^0 (\nabla_\xi u_1 + \nabla_x u_0) + \varepsilon (\nabla_\xi u_2 + \nabla_x u_1) + \varepsilon^2 \nabla_x u_2] = 0 \quad (15.23)$$

d'où :

$$\begin{cases} n \cdot A(\xi) \nabla_{\xi} u_0 & = 0 \\ n \cdot A(\xi) (\nabla_{\xi} u_1 + \nabla_x u_0) & = 0 \\ n \cdot A(\xi) (\nabla_{\xi} u_2 + \nabla_x u_1) & = 0 \end{cases} \quad (15.24)$$

On peut prouver que si v_0 est solution du problème, alors $v_0 + \text{cte}$ aussi. Des conditions d'existence de la solution il vient, tous calculs faits, la matrice des coefficients homogénéisés \hat{A} :

$$\hat{A} = [\hat{a}_{i,j}] \quad \text{avec} \quad \hat{a}_{i,j} = \int_{Q \setminus G_0} \sum_{k=1}^n \left(a_{ik} \frac{\partial N_j}{\partial \xi_k} + a_{ij} \right) d\xi \quad (15.25)$$

où G_0 est un trou de la cellule élémentaire Q . Le même type de calcul pourrait être mené avec les conditions de Dirichlet sur le bord des trous. Cette homogénéisation des milieux poreux est valable pour l'acoustique comme pour la mécanique.

15.4 Homogénéisation des problèmes non stationnaires

Nous n'avons pas encore abordé de manière pratique les problèmes non stationnaires dans ce document, puisqu'ils se trouvent au chapitre 17.

Il est tout à fait possible d'utiliser les méthodes présentées dans ce chapitre aux problèmes dépendant du temps. Il n'y a pas vraiment de précautions supplémentaires à prendre, mais il faut adapter la forme de l'ansatz. Par exemple, l'ansatz du paragraphe précédent, pour le même type de problème non stationnaire serait :

$$u^{(2)} = u_0(x, \xi, t) + \varepsilon u_1(x, \xi, t) + \varepsilon^2 u_2(x, \xi, t) \quad (15.26)$$

avec $u_i(x, \xi, t)$ 1-périodique en ξ .

15.5 Changement de dimension et raccord de maillage

Les méthodes d'homogénéisation peuvent également être utilisées pour « changer (réduire) » la dimension d'un problème.

Considérons un problème qui se pose dans un domaine plan de longueur 1 et de largeur $\pm \varepsilon/2$. Alors, on peut considérer le comportement asymptotique de la solution lorsque $\varepsilon \rightarrow 0$. Une fois cette solution asymptotique trouvée, le problème initial posé sur le pavé $[0; 1] \times [-\varepsilon/2, +\varepsilon/2]$ est remplacé par le problème homogénéisé posé sur le segment $[0; 1]$. On a donc bien réduit la dimension du problème.

Ce genre de chose correspond typiquement à un modèle plaque ou coque (2D) utilisé à la place d'un modèle tridimensionnel (3D), ou encore mieux à un modèle barre ou poutre (1D).

Cela permet également de développer des éléments finis permettant de raccorder des maillages 3D à des maillages 2D ou à des maillages 1D, afin d'alléger les modèles numériques là où ils peuvent l'être.

En mécanique, de nombreuses théories de plaques ou poutres existent (voir la paragraphe 11.2). Elles sont généralement présentées de manières « physique », mais ne sont rien d'autre que des méthodes d'homogénéisation.

Libre à chacun de préférer une présentation plutôt mécanicienne ou plutôt mathématicienne, le résultat est finalement le même. Mais il nous semble, et c'est la motivation même à l'origine de ce document, que le fait de connaître les deux aide à mieux cerner à la fois les hypothèses sur lesquelles ces développements sont faits (et que l'on oublie parfois, ayant pour conséquence des résultats que l'on peut qualifier de surprenants) ainsi que les potentialités qui s'offrent à nous.

Chapitre 16

Optimisation

Résumé — Ce chapitre va nous permettre de revenir sur les multiplicateurs de Lagrange dont nous avons déjà eu l'occasion de parler en plusieurs endroits de ce document, et dont l'intérêt doit être assez clair pour le lecteur à ce niveau du document.

Nous allons donc prendre un peu de temps pour mieux détailler leur utilisation, en particulier pour transformer une formulation variationnelle assortie d'un critère objectif à atteindre (le tout soumis à un certain nombre de contraintes) à l'aide de ce que l'on appelle le lagrangien, et que nous avons très brièvement évoqué au paragraphe 7.6).

Nous nous intéresserons aussi bien au cas où la fonction objectif est définie par des paramètres, qu'au cas de l'optimisation de forme.

Un problème d'optimisation nécessite de fournir :

- un modèle décrivant le problème : il s'agit généralement d'une équation aux dérivées partielles ;
- un critère (ou des critères) que l'on cherche à minimiser ou à maximiser (c'est la fonction objectif, ou fonction coût) ;
- un ensemble admissible de variables d'optimisation. Cet ensemble prend en compte les éventuelles contraintes que l'on impose aux variables.

Dans ce document, visant un public plutôt mécanicien, nous nous intéresserons essentiellement à de l'optimisation de « structures », et nous envisagerons les cas suivants :

- l'optimisation paramétrique où l'on dispose d'un nombre réduit de variables qui paramètrent la structure (par exemple l'épaisseur d'une plaque, l'orientation des plis d'un composite...). Nous verrons que des problèmes d'homogénéisation entrent dans ce cadre ;
- l'optimisation géométrique : il s'agit du cas où l'on souhaite faire varier toute ou partie de la forme des frontières, mais sans changer la topologie de la pièce, i.e. sans ajouter ou supprimer de « trous » ;
- l'optimisation topologique : c'est le cas le plus général d'optimisation de forme d'une structure. On cherche à trouver la meilleure forme possible sans restriction, i.e. quitte à modifier la topologie.

Si l'on s'en sort en 2D en remarquant qu'une topologie est caractérisée par le nombre de composantes connexes des bords, cela est plus compliqué en 3D où il faut en plus tenir compte du nombre d'anses ou de boucles de la structure.

Notre présentation de l'optimisation se concentrera sur l'approche continue, l'approche discrète n'étant que très brièvement mentionnée au paragraphe 16.1.7 afin de justifier notre choix de l'optimisation continue.

Histoire

Le mot « optimiser » a pour origine latine « optimum » qui signifie le meilleur. Il s'agit donc de faire les choses de la meilleure façon qui soit, ou encore de concevoir la meilleure des solutions possibles en terme d'objectif prescrit (minimisation ou maximisation) et répondant à un ensemble de contraintes. L'optimisation commence avec le calcul des variations, même si des exemples plus

anciens peuvent être trouvés.

Lié à la mythologie de la création de Carthage, résolu élégamment par la reine Didon, ce problème d'isopérimétrie, connu sous le nom de « Problème de Didon », est sans doute l'un des plus anciens problèmes d'optimisation dont on trouve trace.

Dans *Catoptrica*, Héron d'Alexandrie étudie la lumière et ses réflexions. Il énonce ainsi les principes de réflexion de la lumière, principes guidés par la règle selon laquelle la nature choisit toujours le chemin le plus court.

Cette étude de la propagation de la lumière s'est poursuivie par l'énoncé des principes variationnels par Pierre de Fermat et Christian Huygens. Le français énonce une règle fondamentale dans la recherche d'optimum : « Lorsqu'une grandeur, par exemple l'ordonnée d'une courbe, est parvenue à son maximum ou son minimum, dans une situation infiniment voisine, son accroissement ou sa diminution est nulle. »

En juin 1696 dans *Acta Eruditorum*, Jean Bernoulli reprend un problème initialement posé par Galilée. Il s'agit du problème de la courbe brachistochrone que l'on peut formuler comme suit : quelle est la forme de courbe joignant deux points donnés dans un plan vertical, telle qu'un point matériel, soumis uniquement à la pesanteur et initialement sans vitesse, la parcourt en un temps minimal ? [Galilée pensait avoir résolu le problème et que la solution était un arc de cercle] Très rapidement, Leibniz propose une solution à Jean Bernoulli, mais sans qu'il reconnaisse la courbe en question. C'est Jean Bernoulli, qui dispose de deux solutions, qui reconnaît un arc de cycloïde commençant avec une tangente verticale. Tous deux décident de différer la publication de leurs solutions pour laisser à d'autres la possibilité d'aborder le problème. Celui-ci fut également résolu par Jacques Bernoulli, Newton, L'Hôpital et Tschirnhaus. Les méthodes imaginées pour sa résolution amenèrent à développer la branche des mathématiques qu'on appelle le calcul des variations.



J. Bernoulli Euler Lagrange

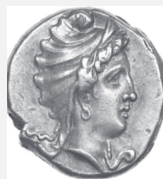
La solution de Jean Bernoulli était fondée sur une analogie avec la propagation de la lumière et le principe de Fermat, ainsi que la loi de Descartes. Celle de Leibniz, était fondée sur l'approximation de la courbe par des lignes brisées et était le premier pas vers l'équation d'Euler-Lagrange. Le second pas a été accompli par Euler qui a ébauché, à partir de considérations géométriques, la méthode des « petites variations » ; vers le milieu du dix-huitième siècle Joseph-

Louis Lagrange a donné sa forme actuelle à la solution d'Euler. Legendre a complété l'équation d'Euler-Lagrange, qui est une condition du premier ordre, par la condition du second ordre qui porte son nom.

Ces résultats ont été rassemblés par Lagrange dans sa *Théorie des fonctions analytiques*, parue en 1797, et dans laquelle on lit : « On peut les réduire à ce principe général. Lorsqu'une fonction de plusieurs variables doit être un maximum ou minimum, et qu'il y a entre ces variables une ou plusieurs équations, il suffira d'ajouter à la fonction proposée les fonctions, qui doivent être nulles, multipliées chacune par une quantité indéterminée, et de chercher ensuite le maximum ou minimum comme si les variables étaient indépendantes ; les équations que l'on trouvera combinées avec les équations données, serviront à déterminer toutes les inconnues ». La démarche est on ne peut plus claire, et c'est celle que nous suivrons Ces « quantités indéterminées » sont bien les multiplicateurs de Lagrange. Tout est dit !

Il revenait à Weierstrass, en 1879, de définir la notion d'extremum fort et d'établir la condition qui porte son nom. Les travaux de Jacobi et Hamilton, contemporains de ceux de Weierstrass, ont permis de donner sa forme définitive à la solution de Jacques Bernoulli déjà mentionnée. Les principaux résultats du calcul des variations classique avaient dès lors été obtenus.

La formulation intégrale quant à elle a été peaufinée à la fin du XIX^e siècle et au début du XX^e siècle. Les économistes s'intéresseront également à ces théories. En 1939, Leonid Kantorovich invente la programmation linéaire. La formulation générale des problèmes de programmation linéaire quant à elle sera finalisée en 1947 par George Dantzig, qui invente par ailleurs la méthode du simplexe. On parle aussi de « recherche opérationnelle » pendant la seconde guerre mondiale, puis de programmation linéaire, et enfin de programmation non-linéaire. Le calcul des variations a connu un profond renouveau dans les années 1950 avec le développement de la théorie de la commande optimale, sous l'impulsion de Lev Pontriaguine et Richard Bellman.



Didon



Héron d'Alexandrie



Fermat

Le calcul des variations reste en mathématiques un domaine fort actif, et les mathématiciens qui ont contribué à son développement sont extrêmement nombreux.

16.1 Théorie de l'optimisation

16.1.1 Existence et unicité d'un minimum

Définition 67 — Minima. Soient V un espace de Banach et $K \subset V$ un sous-ensemble non vide. Soit $J : V \rightarrow \mathbb{R}$ le critère (i.e. la fonction objectif). On considère le problème :

$$\inf_{v \in K} J(v) \quad (16.1)$$

On dit que u est un minimum local de J sur K si :

$$u \in K \quad \text{et} \quad \exists \delta > 0, \forall v \in K, \|v - u\| < \delta \Rightarrow J(v) \geq J(u) \quad (16.2)$$

On dit que u est un minimum global de J sur K si :

$$u \in K \quad \text{et} \quad J(v) \geq J(u), \forall v \in K \quad (16.3)$$

Définition 68 — Suite minimisante. Une suite minimisante du critère J sur K est une suite $(u^n)_{n \in \mathbb{N}}$ telle que :

$$u^n \in K, \forall n \quad \text{et} \quad \lim_{n \rightarrow \infty} J(u^n) = \inf_{v \in K} J(v) \quad (16.4)$$

Par définition de l'infimum de J sur K , une telle suite existe toujours.

Évidemment, nous allons nous intéresser à l'optimisation en dimension finie, et pour cela, nous nous servirons du résultats suivant :

Théorème 53 — Optimisation en dimension finie $V = \mathbb{R}^N$. Soit K un ensemble fermé non vide de \mathbb{R}^N , et J une fonction continue sur K à valeurs dans \mathbb{R} vérifiant la propriété infinie à l'infini suivante :

$$\text{Quelle que soit la suite } (u^n)_{n \in \mathbb{N}} \text{ dans } K, \quad \lim_{n \rightarrow \infty} \|u^n\| = +\infty \Rightarrow \lim_{n \rightarrow \infty} J(u^n) = +\infty \quad (16.5)$$

alors il existe au moins un point de minimisation de J sur K .

De plus, on peut extraire de toute suite minimisante de J sur K une sous-suite convergeant vers un point de minimum sur K .

Remarque. Le théorème précédent n'est pas toujours vérifié en dimension infinie : une fonction continue sur un fermé borné n'atteint pas toujours son minimum.

Afin d'obtenir des résultats d'existence, on rajoute une hypothèse de convexité.

Définition 69 — Convexité. Un ensemble $K \subset V$ est dit convexe si :

$$\forall x, y \in K, \forall \theta \in [0; 1], \text{ l'élément } (\theta x + (1 - \theta)y) \in K \quad (16.6)$$

Une fonction J définie sur un ensemble convexe non vide K et à valeurs dans \mathbb{R} est dite

convexe sur K si et seulement si :

$$J(\theta x + (1 - \theta)y) \leq \theta J(x) + (1 - \theta)J(y), \quad \forall x, y \in K, \forall \theta \in [0; 1] \quad (16.7)$$

Si de plus l'inégalité précédente est stricte lorsque $x \neq y$ et $\theta \in]0; 1[$, alors la fonction J est dite strictement convexe.

On dispose alors du résultat suivant d'existence :

Théorème 54 — Existence du minimum. Soient K un convexe fermé non vide d'un espace de Banach réflexif V , et J une fonction convexe continue sur K qui est infinie à l'infini dans K , i.e. qui vérifie l'équation (16.5), alors il existe un minimum de J sur K .

Remarques. En toute généralité, ce théorème reste valable si V est le dual d'un espace de Banach séparable. Le fait que V soit un espace de Banach réflexif correspond au fait que $(V')' = V$ (avec V' le dual de V). Enfin et surtout, ce théorème est vrai dans tous les espaces que l'on rencontre habituellement dans nos problématiques, en particulier pour les espace $L^p(\Omega)$ pour $1 < p \leq \infty$.

Théorème 55 — Unicité du minimum. Si de plus la fonction J est strictement convexe, alors il existe au plus un point de minimum.

Théorème 56 Si la fonction J est convexe sur un ensemble convexe K , alors tout point de minimum local de J sur K est un minimum global.

On voit donc tout l'intérêt de vérifier que notre fonction objectif est convexe. Un exemple d'un tel problème en mécanique est la minimisation de l'énergie.

16.1.2 Différentiabilité et optimalité

Nous avons déjà présenté au paragraphe 3.5, à l'équation (3.7), la différentielle. Si une fonction peut être approchée par une forme linéaire (la différentielle), alors on dit également qu'elle est différentiable au sens de Fréchet.

Réécrivons la définition de la différentielle dans le cas de la fonction objectif J .

Définition 70 — Différentiabilité au sens de Fréchet. Soit V un espace de Banach. On dit que la fonction J , définie au voisinage de $u \in V$ et à valeurs dans \mathbb{R} , est différentiable au sens de Fréchet en u s'il existe une forme linéaire $L \in V'$, continue sur V , telle que :

$$J(u + v) = J(u) + L(v) + o(v), \quad \text{avec} \quad \lim_{v \rightarrow 0} \frac{|o(v)|}{\|v\|} = 0 \quad (16.8)$$

On appelle L la différentielle (ou la dérivée, ou le gradient) de J en u que l'on note $L = J'(u)$, ou encore $L(v) = \langle J'(u), v \rangle_{V', V}$.

Remarques. Si V est un espace de Hilbert, alors le théorème de représentation de Riesz-Fréchet, donné par l'équation (7.5), permet d'identifier V à son dual V' . Il existe donc un unique $p \in V$ tel que $\langle p, v \rangle = L(v)$. On note $p = J'(u)$.

Cette identification $V = V'$ est utilisée si $V = \mathbb{R}^N$ ou $V = L^2(\Omega)$.

En pratique, il est plus simple de calculer la dérivée directionnelle :

$$j'(0) = \langle J'(u), v \rangle_{V', V} \quad (16.9)$$

avec $j(t) = J(u + tv)$.

Lien entre minimisation et formulation variationnelle. Nous nous proposons de retrouver les résultats du paragraphe 7.6.

Soient $a(.,.)$ une forme bilinéaire symétrique (continue coercitive) et $f(.)$ une forme linéaire continue. On considère la fonction $J(u) = \frac{1}{2}a(u, u) - f(u)$. On pose, comme ci-dessus, $j(t) = J(u + tv)$. On a donc :

$$j(t) = \frac{t^2}{2}a(v, v) + t(a(u, v) - f(v)) + J(u)$$

On dérive j par rapport à t , ce qui donne :

$$j'(t) = ta(v, v) + a(u, v) - f(v)$$

Par définition, on a $j'(0) = \langle J'(u), v \rangle_{V', V}$, donc :

$$\langle J'(u), v \rangle_{V', V} = a(u, v) - f(v)$$

La condition $J'(u) = 0$ est une formulation variationnelle. On peut ainsi démontrer l'équivalence entre la minimisation de l'énergie $J(u)$ et la résolution de la formulation variationnelle.

Ce résultat est fondamental (et souvent utilisé, donc à connaître). La question que l'on est en droit de se poser est alors de savoir si une telle équivalence tient toujours dans le cas général. La réponse est oui, et nous allons maintenant présenter ces résultats.

Théorème 57 — Inéquation d'Euler. Soient K un ensemble connexe et u un point de K . On suppose que la fonction J est différentiable en u . Si u est un point de minimum local de J sur K , alors :

$$\langle J'(u), v - u \rangle \geq 0, \quad \forall v \in K \quad (16.10)$$

Si u vérifie cette inéquation et si J est convexe, alors u est un minimum global de J sur K .

Si u est intérieur à K , on obtient l'équation d'Euler :

$$J'(u) = 0 \quad (16.11)$$

Dans le cas général, l'inéquation d'Euler est une condition nécessaire. Pour les fonctions convexes, elle est nécessaire et suffisante.

16.1.3 Lagrangien

Dans un premier temps, nous allons nous intéresser au problème de minimisation avec contraintes d'égalité. Le cas des contraintes d'inégalités sera traité juste à la suite.

Définition 71 — Lagrangien (du problème de minimisation avec contraintes d'égalité). Nous travaillons toujours dans notre espace V . Soit $F(v)$ une application dérivable de V dans \mathbb{R}^M (i.e. $F(v)$ est un vecteur dont les composantes sont $F_1(v), \dots, F_M(v)$). Nous cherchons à trouver le minimum de la fonction objectif $J(v)$ où v satisfait aux contraintes $F(v) = 0$, i.e. :

$$\inf_{v \in V, F(v)=0} J(v) \quad (16.12)$$

On appelle Lagrangien du problème, la fonction :

$$\mathcal{L}(v, \mu) = J(v) + \mu.F(v) = J(v) + \sum_{i=1}^M \mu_i F_i(v), \quad \forall (v, \mu) \in V \times \mathbb{R}^M \quad (16.13)$$

La variable $\mu \in \mathbb{R}^M$ est appelée multiplicateur de Lagrange pour la contrainte $F(v) = 0$.

Un premier résultat est que le problème de minimisation précédent est équivalent à :

$$\inf_{v \in V, F(v)=0} J(v) = \inf_{v \in V} \sup_{\mu \in \mathbb{R}^M} \mathcal{L}(v, \mu) \quad (16.14)$$

et l'on dispose du théorème suivant :

Théorème 58 — Stationnarité du Lagrangien. On suppose que J et F sont continûment dérivables au voisinage de $u \in V$ tel que $F(u) = 0$. Si u est un minimum local, et si les vecteurs $(F'_i(u))_{1 \leq i \leq M}$ sont linéairement indépendants, alors il existe des multiplicateurs de Lagrange $\lambda_1, \dots, \lambda_M \in \mathbb{R}$ tels que :

$$\frac{\partial \mathcal{L}}{\partial v}(u, \lambda) = J'(u) + \lambda \cdot F'(u) = 0 \quad \text{et} \quad \frac{\partial \mathcal{L}}{\partial \mu}(u, \lambda) = F(u) = 0 \quad (16.15)$$

Intéressons-nous maintenant au problème de minimisation avec contraintes d'inégalité. Il s'agit du même problème que précédemment, mais la relation (16.12) est remplacée par la relation :

$$\inf_{v \in V, F(v) \leq 0} J(v) \quad (16.16)$$

où $F(v) \leq 0$ signifie que pour tout $1 \leq i \leq M$, $F_i(v) \leq 0$

Définition 72 Soit u tel que $F(u) \leq 0$. L'ensemble $I(u) = \{i \in \{1, \dots, M\}, F_i(u) = 0\}$ est appelé ensemble des contraintes actives en u . On dit que les contraintes d'inégalité sont qualifiées en $u \in K$ si la famille $(F'_i(u))_{i \in I(u)}$ est libre.

Définition 73 — Lagrangien (du problème de minimisation avec contraintes d'inégalité). On appelle Lagrangien du problème, la fonction :

$$\mathcal{L}(v, \mu) = J(v) + \mu \cdot F(v) = J(v) + \sum_{i=1}^M \mu_i F_i(v), \quad \forall (v, \mu) \in V \times (\mathbb{R}^+)^M \quad (16.17)$$

La variable positive $\mu \in (\mathbb{R}^+)^M$ est appelée multiplicateur de Lagrange pour la contrainte $F(v) \leq 0$.

Synthèse méthodologique : Le lagrangien (pour le problème de minimisation avec contraintes d'égalité, comme d'inégalité) est la somme de la fonction objectif et de la formulation variationnelle de l'équation d'état du problème considérée comme une contrainte (i.e. nous avons suivi à la lettre la méthode décrite par Lagrange lui-même).

Un premier résultat est que le problème de minimisation précédent est équivalent à :

$$\inf_{v \in V, F(v) \leq 0} J(v) = \inf_{v \in V} \sup_{\mu \in (\mathbb{R}^+)^M} \mathcal{L}(v, \mu) \quad (16.18)$$

et l'on dispose du théorème suivant :

Théorème 59 — Stationnarité du Lagrangien. On suppose que les contraintes sont qualifiées en u tel que $F(u) \leq 0$. Si u est un minimum local, alors il existe des multiplicateurs de Lagrange $\lambda_1, \dots, \lambda_M \in \mathbb{R}^+$ tels que :

$$\frac{\partial \mathcal{L}}{\partial v}(u, \lambda) = J'(u) + \lambda \cdot F'(u) = J'(u) + \sum_{i=1}^M \lambda_i F'_i(u) = 0, \quad \lambda_i \geq 0, \quad \forall i \in \{1, \dots, M\} \quad (16.19)$$

Notons que $\lambda_i = 0$ si $F_i(u) < 0$.

En plus de la stationnarité du Lagrangien, notons que la condition $\lambda \geq 0$, $F(u) \leq 0$, $\lambda \cdot F(u) = 0$ est équivalente à l'inéquation d'Euler pour la maximisation par rapport à μ dans le convexe fermé qu'est $(\mathbb{R}^+)^M$:

$$\frac{\partial \mathcal{L}}{\partial \mu}(u, \lambda) \cdot (\mu - \lambda) = F(u) \cdot (\mu - \lambda) \leq 0, \quad \forall \mu \in (\mathbb{R}^+)^M \quad (16.20)$$

D'un point de vue pratique, λ donne la dérivée (sans la calculer) du minimum par rapport à la contrainte.

Illustration. Considérons le cas de la minimisation de $J(u)$ sous la contrainte $F(u) = c$. Le Lagrangien est :

$$\mathcal{L}(v, \mu, c) = J(v) + \mu \cdot (F(v) - c)$$

et on se propose d'étudier la sensibilité du Lagrangien par rapport à c .

Notons $u(c)$ et $\lambda(c)$ le point de minimum et le multiplicateur de Lagrange correspondant. On supposera évidemment qu'ils sont dérivables par rapport à c . Il vient alors :

$$\nabla_c (J(u(c))) = -\lambda(c)$$

et ainsi λ donne bien la dérivée (sans la calculer) du minimum par rapport à c .

16.1.4 Dualité et point selle

Dans ce paragraphe, on va regarder comment supprimer les contraintes portant sur le problème. Pour cela, on va adjoindre au problème dit primal, un problème dual. Ce dernier est souvent plus simple que le problème primal, notamment parce qu'il ne comporte pas de contraintes. Évidemment, connaissant la solution du problème dual, on devra être en mesure de remonter à celle du problème primal.

Définition 74 Soit un Lagrangien $\mathcal{L}(v, q)$. On dit que $(u, p) \in U \times P$ est un point selle de \mathcal{L} sur $U \times P$ si :

$$\forall q \in P, \quad \mathcal{L}(u, q) \leq \mathcal{L}(u, p) \leq \mathcal{L}(v, p), \quad \forall v \in U \quad (16.21)$$

Pour $v \in U$ et $q \in P$, posons $\mathcal{J}(v) = \sup_{q \in P} \mathcal{L}(v, q)$ et $\mathcal{G}(q) = \inf_{v \in U} \mathcal{L}(v, q)$. On appelle problème primal :

$$\inf_{v \in U} \mathcal{J}(v) \quad (16.22)$$

et problème dual :

$$\sup_{q \in P} \mathcal{G}(q) \quad (16.23)$$

Pour faire le lien avec ce qui précède, plaçons nous dans le cas où $U = V$, $P = \mathbb{R}^M$, et $\mathcal{L}(v, q) = J(v) + q \cdot F(v)$. Il vient alors : $\mathcal{J}(v) = J(v)$ si $F(v) = 0$ et $\mathcal{J}(v) = +\infty$ sinon. Par contre, il n'y a pas de contrainte pour le problème dual, si ce n'est $q \in P = \mathbb{R}^M$.

Comme annoncé, le problème dual est plus simple que le problème primal, car il n'a pas de contrainte.

Théorème 60 — Dualité forte. Le couple (u, p) est un point selle de \mathcal{L} sur $U \times P$ si et seulement si :

$$\mathcal{J}(u) = \min_{v \in U} \mathcal{J}(v) = \max_{q \in P} \mathcal{G}(q) = \mathcal{G}(p) \quad (16.24)$$

Ainsi, si l'on sait résoudre le problème dual, alors on obtient la solution du problème primal grâce à une minimisation sans contrainte.

16.1.5 Maximisation

Jusqu'à présent, nous n'avons parlé que de problème de minimisation. Si l'on a à traiter un problème de maximisation, il suffit alors de le transformer en problème de minimisation :

$$\sup_{v \in V} J(v) = - \inf_{v \in V} (-J(v)) \quad (16.25)$$

16.1.6 Ensemble admissible

Nous avons été assez vagues sur le domaine dans lequel peuvent varier les variables d'optimisation. Nous avons parlé d'espace, d'ensemble connexe...

D'un point de vue pratique, les variables (ou paramètres) ne peuvent varier qu'à l'intérieur d'un domaine admissible. Cet ensemble admissible sera noté \mathcal{U}_{ad} .

Cet ensemble peut aussi bien comprendre :

- des domaines discrets : par exemple, 1) le nombre de plis dans un stratifié doit être un entier compris entre 1 et n , où n est le nombre de plis maximum au-delà duquel on estime que le processus de fabrication ne sera plus performant (techniquement et/ou économiquement), 2) l'épaisseur d'une tôle peut n'avoir qu'un nombre fini de valeurs discrètes correspondant à ce qui existe en stock ou dans le commerce ;
- que des domaines continus : par exemple, 1) l'angle d'orientation des plis dans un stratifié est une variable continue sur $[-\pi ; \pi]$, 2) l'épaisseur d'une structure peut être vue comme une variable continue allant d'une valeur min à une valeur max données.

Lorsque nous parlerons d'optimisation géométrique (ou optimisation de forme), alors nous aurons à manipuler des contraintes du type : telle partie de la frontière peut bouger, mais sa longueur doit être constante, ou le volume total délimité doit respecter telle ou telle condition.

Nous pourrions également avoir des critères liés à l'état des contraintes ou des déformations de la structure, des critères de moindre carrés (pour atteindre par exemple un déplacement cible), une rigidité, un déplacement maximal...

Pour prendre en compte des variables ayant des valeurs discrètes, on peut se servir de fonctions caractéristiques. Par exemple : si une variable v ne peut prendre que la valeur α dans un sous-domaine $\omega \subset \Omega$ et la valeur β ailleurs, alors on pourra écrire :

$$v = \alpha \chi(x) + \beta(1 - \chi(x)) \quad (16.26)$$

où x est la position du point considéré et $\chi(x)$ la fonction caractéristique qui vaut 1 si $x \in \omega$ et 0 sinon.

On peut également faire varier le critère en fonction de la zone de la structure considérée. Pour cela, notre domaine Ω peut être décomposé en une partition $(\omega_i)_{1 \leq i \leq n}$, et l'on est amené à considérer l'ensemble admissible :

$$\mathcal{U}_{ad}^n = \{x \in \mathcal{U}_{ad}, x \text{ vérifie la condition } i \text{ dans } \omega_i, 1 \leq i \leq n\} \quad (16.27)$$

On peut également avoir des contraintes de régularité. C'est par exemple ce que l'on peut imposer à une épaisseur variable : non seulement celle-ci doit être comprises entre une valeur mini et une valeur maxi, mais on peut également vouloir que sa variation soit bornée, i.e. qu'elle ne varie pas trop vite, ce qui correspond à une contrainte pratique du processus de fabrication. Ainsi, on se fixe une constante positive R , et l'on travaille sur l'ensemble admissible :

$$\mathcal{U}_{ad}^{reg} = \{x \in \mathcal{U}_{ad} \cap W^{1,\infty}(\Omega), \|x\|_{W^{1,\infty}(\Omega)} \leq R\} \quad (16.28)$$

Le problème reste de choisir la valeur de R . En général, il n'y a d'ailleurs pas convergence quand R tend vers l'infini. Numériquement, on préfère la contrainte de régularité :

$$\|x\|_{H^1(\Omega)} \leq R \quad (16.29)$$

16.1.7 L'optimisation discrète

On peut également discrétiser des contraintes d'optimisation. L'ensemble admissible devient alors :

$$\mathcal{U}_{ad}^{disc} = \left\{ \begin{array}{l} \text{contraintes sur } x \text{ idem celles sur } \mathcal{U}_{ad}; \dots \\ \dots \text{ contraintes sur } \int x \dots = \dots \text{ remplacées par } \sum_{i=1}^n c_i x_i = \dots \end{array} \right\} \quad (16.30)$$

où $\sum_{i=1}^n c_i x_i$ est une approximation de $\int x \dots$

Le problème est de calculer le gradient de J^{disc} . Si l'on procède de manière directe, cela est impraticable numériquement. On passe alors par l'état adjoint (voir exemples plus bas).

Il n'y a pas de simplification à utiliser une approche discrète plutôt que continue :

- certes « discrétiser d'abord puis optimiser » garantit une parfaite compatibilité, mais cela nécessite une parfaite connaissance du solveur numérique ;
- alors qu'« optimiser en continu puis discrétiser », tel que présenté dans ce chapitre, est beaucoup plus simple et ne nuit pas en précision numérique si l'on choisit bien les espaces d'éléments finis.

16.1.8 Remarques sur la mise en œuvre

Des méthodes dites « stochastiques » telles que Monte-Carlo, recuit simulé ou des algorithmes génétiques sont envisageables. Elles sont coûteuses en CPU et nous ne les aborderons pas.

Nous allons dire deux mots sur les méthodes déterministes, telles que les méthodes de gradient (gradient optimal, gradient à pas fixe, gradient projeté) ou de Newton. Notons qu'elles fournissent un minimum local, et que, pour fonctionner, il faut disposer du gradient. Ainsi, nous devons pouvoir calculer J' . C'est d'ailleurs sur le calcul de J' que nous insisterons dans les exemples traités.

La méthode la plus simple (et la plus sûre) pour calculer le gradient est :

$$\lim_{\varepsilon \rightarrow 0} \frac{J(x + \varepsilon y) - J(x)}{\varepsilon} = \langle J'(x), y \rangle \quad (16.31)$$

Par linéarité on écrit : $u(x + \varepsilon y) = u(x) + \varepsilon \tilde{u}(y)$, les relations satisfaites par $\tilde{u}(y)$ dépendant du problème considéré. Toutefois, remarquons qu'avec cette manière de procéder $J'(x)$ n'est pas explicite.

Pour calculer explicitement l'expression du gradient, on utilise l'état adjoint p , qui nécessite de résoudre deux problèmes aux limites (u et p). Cette méthode est très efficace en pratique, c'est la meilleure possible.

D'une manière générale, on peut dire que l'état adjoint est obtenu comme stationnarité d'un lagrangien dont p est le multiplicateur de Lagrange et l'équation d'état la contrainte. De même, J' est obtenu comme stationnarité de ce même Lagrangien par rapport à la variable de contrôle optimal.

Si l'équation d'état est non-autoadjointe (i.e. si la forme bilinéaire du problème n'est pas symétrique), alors l'opérateur principal de l'équation adjointe est le transposé ou l'adjoint de celui de l'équation d'état. Par contre, même si l'équation d'état est non linéaire, l'équation adjointe est linéaire (c'est assez normal, puisque finalement, on ne s'intéresse qu'aux fluctuations en somme).

Nous n'en dirons pas plus ici, quelques exemples plus bas permettront de voir le calcul effectif de l'état adjoint et de J' .

16.2 Optimisation de forme

Maintenant que nous avons présenté la théorie de l'optimisation continue, nous nous proposons de dire quelques mots sur sa mise en œuvre dans le cadre de l'optimisation de forme. D'un point de vue vraiment pratique, c'est au paragraphe suivant que se trouveront les exemples, traités dans le cas d'une plaque.

Le problème principal qui se pose est de savoir comment poser le problème, i.e. comment représenter des formes, et par suite comment définir le calcul des variations sur les formes.

16.2.1 Optimisation paramétrique

On est dans le cas le plus simple. La géométrie ne peut pas « vraiment » varier : elle est paramétrée par des variables. L'ensemble \mathcal{U}_{ad} est donc celui défini par l'ensemble de ces variables munies de leur domaine de variation (valeur mini et maxi).

En somme, ce cas a déjà été traité auparavant, et il n'y a pas de différence dans le traitement d'un paramètre, que celui-ci représente un élément de la géométrie ou un autre type de variable.

16.2.2 Optimisation géométrique

On rappelle que dans ce cas, la géométrie peut varier mais pas la topologie. On n'est donc pas autorisé à ajouter ou supprimer des trous.

Approche « libre »

Généralement, le contour $\Gamma = \partial\Omega$ est partitionné comme suit :

$$\Gamma = \Gamma_D \cup \Gamma_N \cup \Gamma_R \quad (16.32)$$

où Γ_D est la partie où les déplacements sont imposés, considérée comme immuable ; Γ_N est la partie où les efforts extérieurs sont prescrits, également considérée comme fixe ; et Γ_R est le reste du contour, seule partie qui peut varier (tout ou partie, auquel cas on partitionne encore en une partie fixe et une partie mobile). On commence donc par imposer que Γ_D et Γ_N sont fixes :

$$\mathcal{U}_{ad} = \{\Gamma_D \cup \Gamma_N \subset \Gamma\} \quad (16.33)$$

Ne rien imposer de plus ne permet pas de résoudre le problème. Il nous faut des critères d'optimisation. On veut trouver la forme optimale qui permet, par exemple, de maximiser la rigidité, ou d'obtenir un déplacement cible. Mais cela ne suffit pas encore. Si toutes les formes sont admissibles, alors on va arriver à des solutions où l'on a des volumes nuls ou infinis, car cela répondra au mieux... Il est donc nécessaire de contraindre la forme : tout d'abord, on va restreindre la géométrie à ne varier que dans un domaine D_0 borné défini, ce qui se traduit par l'ajout de la condition :

$$\mathcal{U}_{ad} = \{\Omega \subset D_0\} \quad (16.34)$$

Ensuite, par exemple, on peut décider de travailler à volume fixé V_0 , ce qui se traduit par :

$$\mathcal{U}_{ad} = \{|\Omega| = V_0\} \quad (16.35)$$

Nous en sommes donc à un ensemble admissible qui se présente sous la forme :

$$\mathcal{U}_{ad} = \{\Omega \subset D_0 / \Gamma_D \cup \Gamma_N \subset \Gamma \text{ et } |\Omega| = V_0\} \quad (16.36)$$

Mais cela n'est pas encore suffisant pour obtenir une solution. Il est nécessaire d'ajouter que Ω vérifie la propriété du cône uniforme. Cette condition impose à la fois une condition sur le rayon de courbure du bord (c'est une condition pour le bord varie de manière assez lisse) et empêche la formation de petits trous.

Théorème 61 — D. Chenaïs. Si l'ensemble admissible est :

$$\mathcal{U}_{\text{ad}} = \left\{ \Omega \subset D_0 \text{ tq : } \begin{array}{l} (i) \quad \Omega \text{ vérifie la propriété du cône uniforme} \\ (ii) \quad \Gamma_D \cup \Gamma_N \subset \Gamma \text{ et } |\Omega| = V_0 \end{array} \right\} \quad (16.37)$$

alors le problème d'optimisation :

$$\inf_{\Omega \in \mathcal{U}_{\text{ad}}} J(\Omega)$$

admet au moins une solution optimale.

Approche « régulière »

Une autre voie est possible pour réaliser une optimisation géométrique : c'est celle où l'on déforme progressivement le domaine initial pour arriver jusqu'au domaine final. Nous avons déjà mentionné cette manière de faire à propos des maillages au chapitre 14. Nous allons pour cela utiliser la notion de transformation par difféomorphisme du domaine initial Ω_0 de \mathbb{R}^N . Une telle approche a pour avantage d'assurer qu'au cours de la transformation le domaine reste bien topologiquement identique au domaine initial. Elle a également le gros avantage de permettre de calculer un gradient.

Définition 75 — Espace des difféomorphismes. L'espace des difféomorphismes \mathcal{T} (ou espace des bijections régulières) sur \mathbb{R}^N est défini par :

$$\mathcal{T} = \{T \text{ tel que } (T - Id) \text{ et } (T + Id) \in W^{1,\infty}(\mathbb{R}^N; \mathbb{R}^N)\} \quad (16.38)$$

L'espace de Sobolev $W^{1,\infty}(\mathbb{R}^N; \mathbb{R}^N)$ est également appelé espace des vecteurs Lipschitziens. Il est muni de :

$$\|\varphi\|_{W^{1,\infty}(\mathbb{R}^N; \mathbb{R}^N)} = \sup_{x \in \mathbb{R}^N} (|\varphi(x)|_{\mathbb{R}^N} + |\nabla \varphi(x)|_{\mathbb{R}^N \times \mathbb{R}^N}) < \infty \quad (16.39)$$

Ainsi, φ est continue mais ses dérivées sont seulement bornées.

L'espace des formes admissibles est alors :

$$\mathcal{C}(\Omega_0) = \{\Omega \text{ tel que : } \exists T \in \mathcal{T}, \Omega = T(\Omega_0)\} \quad (16.40)$$

Attention, chaque domaine Ω est bien paramétré par un difféomorphisme T , mais celui-ci n'est pas unique.

On définit la pseudo-distance :

$$d(\Omega_1, \Omega_2) = \inf_{T \in \mathcal{T} / T(\Omega_1) = \Omega_2} (\|T - Id\| + \|T^{-1} - Id\|)_{W^{1,\infty}(\mathbb{R}^N; \mathbb{R}^N)} \quad (16.41)$$

Si Ω_0 est borné, alors on peut utiliser $C^1(\mathbb{R}^N; \mathbb{R}^N)$ au lieu de $W^{1,\infty}(\mathbb{R}^N; \mathbb{R}^N)$.

L'ensemble admissible est finalement :

$$\mathcal{U}_{\text{ad}} = \{\Omega \in \mathcal{C}(\Omega_0) / \Gamma_D \cup \Gamma_N \subset \Gamma \text{ et } |\Omega| = V_0\} \quad (16.42)$$

Comme précédemment, on fixe une constante positive R (contrainte de faisabilité), et on introduit le sous-espace régulier :

$$\mathcal{U}_{\text{ad}}^{\text{reg}} = \{\Omega \in \mathcal{U}_{\text{ad}} / d(\Omega, \Omega_0) \leq R\} \quad (16.43)$$

On dispose alors du résultat suivant :

Théorème 62 Si l'ensemble admissible est $\mathcal{U}_{\text{ad}}^{\text{reg}}$ défini juste avant, alors le problème d'optimisation :

$$\inf_{\Omega \in \mathcal{U}_{\text{ad}}^{\text{reg}}} J(\Omega)$$

admet au moins une solution optimale.

Notons que les bords dans $\mathcal{U}_{\text{ad}}^{\text{reg}}$ ne peuvent pas trop osciller.

Comme $J(\Omega)$ est paramétré par le difféomorphisme T , il est possible de calculer sa dérivée.

En se limitant aux difféomorphismes de la forme $T = Id + \theta$, avec $\theta \in W^{1,\infty}(\mathbb{R}^N; \mathbb{R}^N)$, on peut dériver $\theta \rightarrow J((Id + \theta)\Omega_0)$ en 0, qui est strictement contractante. Ainsi, $T = Id + \theta$ est une bijection de \mathbb{R}^N . Il suffit d'appliquer la définition de la dérivée au sens de Fréchet. La forme linéaire est cette fois-ci appelée dérivée de forme.

De la même manière, il est possible de définir la dérivée d'une fonction dépendant du domaine, que ce soit la dérivée eulérienne ou la dérivée lagrangienne. Dans la mesure où la dérivée eulérienne est délicate à manipuler, on préférera utiliser la dérivée lagrangienne, quitte à exprimer les résultats de la dérivation eulérienne à partir de la dérivation lagrangienne. Enfin, on peut également dériver une équation par rapport au domaine, ou une formulation variationnelle... mais nous en resterons là pour l'instant.

De toutes façons, ce n'est pas cette méthode (calculatoire) qui est utilisée en pratique, mais la méthode du Lagrangien... dont nous avons déjà parlé auparavant. En effet, la dérivée du Lagrangien par rapport au domaine, dans la direction θ , évaluée sur le domaine Ω_0 , est égale à la dérivée directionnelle $J'(\Omega_0)(\theta)$. Ainsi, grâce à ce calcul, on peut obtenir de bons résultats pour $J'(\Omega_0)$ sans passer par les dérivées de forme ou matérielle. Ce calcul nécessite toutefois que l'on connaisse la dérivabilité de u par rapport au domaine.

Un problème implicite non évoqué est que pour changer de forme, il faut pouvoir remailler ou déformer le maillage à chaque itération. Comme un remaillage est beaucoup plus coûteux qu'une déformation de maillage, on procède par déformation de maillage (avec toutes les précautions nécessaires sur le respect des formes d'éléments, sur le croisement des bords...) avec de temps en temps un remaillage (on ne peut pas y couper...). C'est pourquoi la mise en œuvre informatique de l'optimisation géométrique reste compliquée (surtout en 3D).

Une « Boîte à outils FREEFEM++ pour l'optimisation de formes (géométrique et topologique) » est disponible sur le site de Grégoire Allaire du CMAP [4].

16.2.3 Optimisation topologique par homogénéisation

La restriction majeur de l'optimisation géométrique dont nous venons de parler est que l'on ne modifie pas la topologie de la structure, ce qui limite considérablement les formes possibles. Nous allons voir maintenant comment s'affranchir de cette limitation, en présentant une méthode d'homogénéisation. Nous ne présenterons pas d'approche évolutionnaire.

La méthode d'homogénéisation est une méthode qui permet de rendre les problèmes bien posés en élargissant l'espace des formes admissibles. Ces formes généralisées seront les limites de suites minimisantes de formes classiques (ce qui augmente les possibilités, mais sans trop d'excentricité). Ainsi, la méthode d'homogénéisation autorise comme formes admissibles des matériaux composites obtenus par perforation (macroscopique ou microscopique) de la matière : c'est ainsi que sont introduits les « trous » dans la structure.

Ainsi, au lieu de représenter un trou par une fonction caractéristique (qui vaut 1 dans la matière et 0 dans les trous), on va considérer que les trous peuvent être aussi bien macroscopiques que microscopiques. Dans une forme généralisée, on ne représente donc plus les trous, mais on modélise une densité de matière, i.e. un paramètre continu θ variant dans $[0; 1]$, ainsi qu'une microstructure,

i.e. la forme des trous. En effet, la forme des trous est importante car elle influe directement sur le comportement effectif du matériau composite, déterminé par homogénéisation selon les techniques présentées au chapitre 15.

Nous disposerons donc de deux variables d'optimisation : θ la densité de matière et $\hat{A}(x)$ le comportement effectif.

D'un point de vue pratique, les trous, avec bords libres, sont soumis à une condition de Neumann. Ils sont remplacés par un matériau ayant une rigidité très faible au regard de celle du reste du domaine. Par exemple, le module d'Young sera représenté par :

$$E = E_{\text{fort}}\chi(x) + E_{\text{faible}}(1 - \chi(x)) \quad (16.44)$$

où E_{fort} et E_{faible} sont le module d'Young du matériau constituant la structure, et celui, choisi beaucoup plus faible, représentant les trous ; la fonction $\chi(x)$ étant la fonction caractéristique associée. L'ensemble admissible s'écrit alors :

$$\mathcal{U}_{\text{ad}} = \left\{ \chi \in L^\infty(\Omega; \{0, 1\}), \int_{\Omega} \chi(x) = V \right\} \quad (16.45)$$

Le problème d'optimisation de forme de la structure s'écrit :

$$\inf_{\chi \in \mathcal{U}_{\text{ad}}} J(\chi) \quad (16.46)$$

avec par exemple $J(\chi) = \int_{\Omega} f u_{\chi}$ pour un chargement $f \in L^2(\Omega)$, ou $J(\chi) = \int_{\Omega} |u_{\chi} - u_0|^2$ pour un déplacement u_0 prescrit (moindres carrés).

Notons, en complément du chapitre 15, que les méthodes d'homogénéisation fonctionnent également pour les milieux non périodiques.

En effet, on montre que, si dans l'équation (16.44), la fonction $\chi(x)$ est remplacée par une suite $\chi_{\epsilon}(x)$ (ϵ n'est pas la période), alors il existe une sous-suite, une densité $0 \leq \theta \leq 1$, et un tenseur homogénéisé $A^*(x)$ tels que χ_{ϵ} converge en moyenne faiblement vers θ , \hat{A}_{ϵ} converge au sens de l'homogénéisation vers A^* , i.e. que u_{ϵ} converge dans $L^2(\Omega)$ vers la solution u du problème homogénéisé.

C'est ce résultat qui sert à la théorie mathématique de l'optimisation topologique par la méthode d'homogénéisation. Il reste évidemment encore un peu de travail mathématique pour tout justifier jusqu'au bout, mais ce n'est pas le but de ce document. Les quelques indications données permettent de comprendre comment la méthode fonctionne et sur quoi elle repose. Des simplifications, des conditions d'optimalité, la vérification de convexité... permettent d'aboutir à des algorithmes efficaces. Nous renvoyons une fois encore à la « Boîte à outils FREEFEM++ pour l'optimisation de formes (géométrique et topologique) » disponible sur le site de Grégoire Allaire du CMAP [4].

Remarque sur la mise en œuvre de l'optimisation topologique. Des codes de calculs permettent aujourd'hui de réaliser de l'optimisation topologique, mais tous ne le permettent pas.

Il existe une manière simple, mais moins élégante et surtout moins performante que ce qui a été présenté, de s'en sortir avec un code classique non prévu pour cela (c'est par exemple ce que je fais avec mon propre code éléments finis qui n'est pas du tout prévu pour cela à la base). 1) On commence par mailler le domaine maximal dans lequel la structure optimale doit se trouver, 2) puis on effectue le calcul éléments finis, 3) en fonction d'un critère prédéfini (par exemple un critère portant sur une contrainte équivalente inférieure à un seuil), on supprime les éléments négligeables (si le code dispose de méthodes de type « birth and death ») ou on remplace le matériau lié à ces éléments par un matériau très faible (on est plus proche de l'approche par homogénéisation), 4) puis on reboucle le processus à partir du deuxième pas.

Évidemment, il est possible d'améliorer cette stratégie rudimentaire : par exemple ne supprimer/modifier que les éléments dont tous les points vérifient la condition, sinon diviser l'élément (et alors comment le diviser au mieux). On peut être amené à rajouter des éléments. On doit veiller à ne pas introduire de singularité artificielle qui nuirait à l'évaluation de la condition...

16.3 Exemples d'optimisation d'une membrane

Dans ce paragraphe, nous considérerons une membrane élastique occupant au repos un domaine plan Ω , ouvert borné régulier de \mathbb{R}^2 , et nous nous intéresserons au déplacement vertical de la membrane $u(x)$ solution de l'équation :

$$-\operatorname{div}(K(x)\nabla u) = f \text{ dans } \Omega \quad (16.47)$$

avec $f \in L^2(\Omega)$ les forces appliquées. Les conditions aux limites dépendront des cas traités et seront donc spécifiées pour chacun des exemples traités.

16.3.1 Optimisation du matériau élastique anisotrope en rigidité

La membrane est soumise aux conditions aux limites :

$$u = 0 \text{ sur } \partial\Omega$$

Elle est par ailleurs constituée d'un matériau élastique anisotrope dont le tenseur, dans un repère adapté, est :

$$K_0 = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}$$

avec $0 < \alpha < \beta$ les rigidités dans chacune des directions principales de ce matériau. Le tenseur d'élasticité $K(x)$ est obtenu par rotation du matériau anisotrope, en fonction du point courant x , par :

$$K(x) = R^*(x)K_0R(x), \quad \text{avec } R(x) = \begin{bmatrix} \cos \theta(x) & -\sin \theta(x) \\ \sin \theta(x) & \cos \theta(x) \end{bmatrix}$$

Le but est d'optimiser la membrane en faisant varier la rotation du matériau anisotrope, paramétrée par θ .

L'ensemble admissible des rotations est :

$$\mathcal{U}_{\text{ad}} = \{\theta \in L^\infty(\Omega; [0; \pi])\}$$

La fonction objectif est la compliance (la rigidité). On cherche à obtenir :

$$\inf_{\theta \in \mathcal{U}_{\text{ad}}} J(\theta) = \int_{\Omega} f(x)u(x)$$

En écrivant la compliance comme le minimum de l'énergie complémentaire (i.e. en contraintes τ), on obtient :

$$\int_{\Omega} f(x)u(x) = \min_{\substack{\tau \in L^2(\Omega)^2 \\ -\operatorname{div} \tau = f \text{ dans } \Omega}} K^{-1}(x)\tau(x) \cdot \tau(x)$$

i.e. on est ramené au problème de double minimisation :

$$\min_{\theta \in \mathcal{U}_{\text{ad}}} \min_{\substack{\tau \in L^2(\Omega)^2 \\ -\operatorname{div} \tau = f \text{ dans } \Omega}} K^{-1}(x)\tau(x) \cdot \tau(x)$$

Si $(\theta, \tau) \in \mathcal{U}_{\text{ad}} \times L^2(\Omega)^2$ est un minimum, alors $\tau = K\nabla u$ et en tout point l'angle $\theta(x)$ minimise $K^{-1}\tau \cdot \tau$.

Considérons donc la fonction : $\theta \mapsto R^*K_0^{-1}R\tau \cdot \tau = K_0(R\tau) \cdot (R\tau)$. Si $\tau = 0$ alors n'importe quel angle θ réalise le minimum. Si $\tau \neq 0$, alors, en notant $e_2 = (0, 1)$ le vecteur propre

correspondant à la valeur propre β^{-1} de K_0^{-1} , le minimum est atteint pour un angle θ tel que $R\tau$ soit parallèle à e_2 . Dans ce cas, il existe un unique minimum $\theta \in [0; \pi[$ caractérisé par $\tan \theta = \tau_1/\tau_2$.

Cette valeur de θ est optimale aussi bien pour la fonction introduite que pour notre problème d'optimisation.

D'après la condition d'optimalité, on a $R\tau = \pm|\tau|e_2$ si (θ, τ) est optimal, ce qui se traduit par :

$$K^{-1}\tau.\tau = |\tau|^2 K_0^{-1}e_2.e_2 = \beta^{-1}|\tau|^2$$

i.e. le matériau anisotrope optimal s'aligne en permanence avec le vecteur des contraintes τ qui ne voit donc que la rigidité β .

De même, $R\nabla u$ est toujours parallèle à e_2 , ce qui implique que $K\nabla u = \beta\nabla u$ dans Ω .

La solution du problème de minimisation s'obtient en résolvant le système :

$$\begin{cases} -\operatorname{div}(\beta\nabla u) = f & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega \end{cases}$$

puis en calculant, pour tout point x , l'angle $\theta(x)$ tel que $\tan \theta(x) = \tau_1/\tau_2$ si $\tau \neq 0$ (si $\tau = 0$, n'importe quel angle convient).

16.3.2 Optimisation du matériau élastique anisotrope pour un déplacement cible

On reprend le cas précédent, i.e. on a les conditions aux limites :

$$u = 0 \text{ sur } \partial\Omega$$

et on cherche encore à optimiser la membrane en faisant varier la rotation du matériau anisotrope, paramétrée par θ .

$$\mathcal{U}_{\text{ad}} = \{\theta \in L^\infty(\Omega; [0; \pi])\}$$

La fonction objectif est un déplacement cible $u_0(x) \in L^2(\Omega)$. On utilise la méthode des moindres carrés, et on cherche donc à obtenir :

$$\inf_{\theta \in \mathcal{U}_{\text{ad}}} J(\theta) = \int_{\Omega} C(x)|u(x) - u_0(x)|^2$$

avec $C(x)$ une fonction régulière donnée.

Une condition nécessaire d'optimalité pour θ est que la dérivée $J'(\theta)$ s'annule. Calculons donc le gradient de la fonction objectif dans la direction $\eta \in L^\infty(\Omega)$. Elle est donnée par :

$$\langle J'(\theta), \eta \rangle = \lim_{t \rightarrow 0} \frac{J(\theta + t\eta) - J(\theta)}{t}$$

Soit $u(\theta + t\eta)$ la solution de l'équation d'état pour l'angle $\theta + t\eta$. En notant $v = \langle u'(\theta), \eta \rangle$ la dérivée en $t = 0$ de la fonction $t \mapsto u(\theta + t\eta)$, on obtient :

$$\langle J'(\theta), \eta \rangle = 2 \int_{\Omega} C(x)(u(x) - u_0(x))v(x)$$

En dérivant l'équation d'état par rapport à t , on obtient l'équation vérifiée par v :

$$\begin{cases} -\operatorname{div}(K\nabla v) = \operatorname{div}(K'(\theta)\eta\nabla u) & \text{dans } \Omega \\ v = 0 & \text{sur } \partial\Omega \end{cases}$$

avec K' la dérivée de la fonction $\theta \mapsto K(\theta)$.

Pour éliminer v , on introduit l'état adjoint p , solution unique de :

$$\begin{cases} -\operatorname{div}(K\nabla p) = 2C(u - u_0) & \text{dans } \Omega \\ p = 0 & \text{sur } \partial\Omega \end{cases}$$

puis, comme indiqué plus haut, on multiplie l'équation en v par p et vice versa. En intégrant, il vient :

$$\int_{\Omega} K\nabla p \cdot \nabla v = 2 \int_{\Omega} C(u - u_0)v = - \int_{\Omega} K'(\theta)\eta \nabla u \cdot \nabla p$$

d'où l'on déduit que :

$$\langle J'(\theta), \eta \rangle = - \int_{\Omega} K'(\theta)\eta \nabla u \cdot \nabla p$$

Comme $K'(\theta)\eta \nabla u \cdot \nabla p$ est dans $L^1(\Omega)$, qui lui-même est un sous-espace des formes linéaires continues sur $L^\infty(\Omega)$, on peut conclure que :

$$J'(\theta) = K'(\theta)\eta \nabla u \cdot \nabla p$$

Revenons à notre problème d'optimisation. Nous avons déjà mentionné qu'une condition nécessaire d'optimalité pour θ est que $J'(\theta) = 0$. Autrement dit, en fixant les vecteurs ∇u et ∇p , l'application $\theta \mapsto K(\theta)\nabla u \cdot \nabla p$ admet un extremum en tout point x . Si soit ∇u soit ∇p est nul, alors on ne peut rien dire de l'angle θ . Sinon, on introduit deux vecteurs orthogonaux :

$$a = \frac{\nabla u}{|\nabla u|} + \frac{\nabla p}{|\nabla p|} \quad \text{et} \quad b = \frac{\nabla u}{|\nabla u|} - \frac{\nabla p}{|\nabla p|}$$

et on réécrit :

$$\begin{aligned} K(\theta)\nabla u \cdot \nabla p &= \frac{|\nabla u||\nabla p|}{4} (K(\theta)a \cdot a - K(\theta)b \cdot b) \\ &= \frac{|\nabla u||\nabla p|}{4} (K_0(Ra) \cdot (Ra) - K_0(Rb) \cdot (Rb)) \end{aligned}$$

dont les extrema sont des minima ou des maxima (pas de point selle) qui s'obtiennent en alignant les vecteurs Ra et Rb avec les vecteurs propres du tenseur K_0 , que nous noterons e_1 et e_2 (i.e. Ra aligné avec e_1 ou e_2 , et donc Rb aligné avec l'autre). On a donc obtenu les angles minimisant le problème.

16.3.3 Optimisation de l'épaisseur en rigidité

On s'intéresse toujours au déplacement vertical u de la membrane, solution de l'équation (16.47), dans le cas où $K(x) = h(x)$, avec conditions de Dirichlet sur le bord :

$$\begin{cases} -\operatorname{div}(h(x)\nabla u) = f & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega \end{cases}$$

avec $f \in L^2(\Omega)$.

L'épaisseur $h(x)$ appartient à l'ensemble admissible :

$$\mathcal{U}_{\text{ad}} = \{h \in L^\infty(\Omega), 0 < h_{\min} \leq h(x) \leq h_{\max} \text{ dans } \Omega\}$$

et l'on s'intéresse à la fonction objectif :

$$\inf_{h \in \mathcal{U}_{\text{ad}}} J(h) = \int_{\Omega} g(x)u(x)$$

où $g(x) \in L^2(\Omega)$ est une fonction donnée.

La formulation variationnelle du problème est : trouver $u \in H_0^1(\Omega)$ tel que :

$$\int_{\Omega} h \nabla u \cdot \nabla v = \int_{\Omega} f v, \quad \forall v \in H_0^1(\Omega)$$

Le Lagrangien (qui rappelons le est la somme de la fonction objectif et de la formulation variationnelle considérée comme une contrainte) est :

$$\mathcal{L}(h, v, q) = \int_{\Omega} g v + \int_{\Omega} h \nabla v \cdot \nabla q - \int_{\Omega} f q$$

où $(h, v, q) \in \mathcal{U}_{\text{ad}} \times H_0^1(\Omega) \times H_0^1(\Omega)$.

La formulation variationnelle de l'équation adjointe est donnée, par définition, par :

$$\left\langle \frac{\partial \mathcal{L}}{\partial v}(h, u, p), \varphi \right\rangle = 0, \quad \forall \varphi \in H_0^1(\Omega)$$

i.e. :

$$\int_{\Omega} h \nabla p \cdot \nabla \varphi + \int_{\Omega} g \varphi = 0, \quad \forall \varphi \in H_0^1(\Omega)$$

Par conséquent, l'état adjoint $p \in H_0^1(\Omega)$ est solution de :

$$\begin{cases} -\operatorname{div}(h \nabla p) = -g & \text{dans } \Omega \\ p = 0 & \text{sur } \partial\Omega \end{cases}$$

La dérivée de la fonction objectif $J(h)$ est donnée, pour tout $k \in L^\infty(\Omega)$, par :

$$\langle J'(h), k \rangle = \int_{\Omega} J'(h) k = \left\langle \frac{\partial \mathcal{L}}{\partial h}(h, u, p), k \right\rangle = \int_{\Omega} k \nabla u \cdot \nabla p$$

soit : $J'(h) = \nabla u \cdot \nabla p$.

16.3.4 Optimisation de l'épaisseur pour un déplacement cible

On s'intéresse au même problème qu'au paragraphe précédent :

$$\begin{cases} -\operatorname{div}(h(x) \nabla u) = f & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega \end{cases}$$

avec $f \in L^2(\Omega)$.

L'épaisseur $h(x)$ appartient à l'ensemble admissible :

$$\mathcal{U}_{\text{ad}} = \{h \in L^\infty(\Omega), 0 < h_{\min} \leq h(x) \leq h_{\max} \text{ dans } \Omega\}$$

et l'on s'intéresse à la fonction objectif :

$$\inf_{h \in \mathcal{U}_{\text{ad}}} J(h) = \int_{\Omega} |\nabla u(x) - \nabla u_0(x)|^2$$

où $u_0(x) \in H^1(\Omega)$ est un déplacement cible donné.

La formulation variationnelle a été donnée au paragraphe précédent, et le Lagrangien est défini par :

$$\mathcal{L}(h, v, q) = \int_{\Omega} |\nabla v - \nabla u_0|^2 + \int_{\Omega} h \nabla v \cdot \nabla q - \int_{\Omega} f q$$

où $(h, v, q) \in \mathcal{U}_{\text{ad}} \times H_0^1(\Omega) \times H_0^1(\Omega)$.

La formulation variationnelle de l'équation adjointe est donnée, par définition, par :

$$\left\langle \frac{\partial \mathcal{L}}{\partial v}(h, u, p), \varphi \right\rangle = 0, \quad \forall \varphi \in H_0^1(\Omega)$$

i.e. :

$$\int_{\Omega} h \nabla p \cdot \nabla \varphi + 2 \int_{\Omega} \nabla(u - u_0) \cdot \nabla \varphi = 0, \quad \forall \varphi \in H_0^1(\Omega)$$

Par conséquent, l'état adjoint $p \in H_0^1(\Omega)$ est solution de :

$$\begin{cases} -\operatorname{div}(h \nabla p) = 2 \Delta(u - u_0) & \text{dans } \Omega \\ p = 0 & \text{sur } \partial\Omega \end{cases}$$

La dérivée de la fonction objectif $J(h)$ est donnée, pour tout $k \in L^\infty(\Omega)$, par :

$$\langle J'(h), k \rangle = \int_{\Omega} J'(h) k = \left\langle \frac{\partial \mathcal{L}}{\partial h}(h, u, p), k \right\rangle = \int_{\Omega} k \nabla u \cdot \nabla p$$

soit : $J'(h) = \nabla u \cdot \nabla p$. Compte-tenu du paragraphe précédent, ne pouvait-on se douter du résultat ?

16.3.5 Optimisation de l'épaisseur pour une contrainte cible

On va maintenant considérer que le bord du domaine est partitionné en deux parties de mesures non nulles $\partial\Omega = \Gamma_D \cup \Gamma_N$. Le problème à résoudre est donc :

$$\begin{cases} -\operatorname{div}(h(x) \nabla u) = f & \text{dans } \Omega \\ u = 0 & \text{sur } \Gamma_D \\ h \frac{\partial u}{\partial n} = 0 & \text{sur } \Gamma_N \end{cases}$$

avec $f \in L^2(\Omega)$.

L'épaisseur $h(x)$ appartient à l'ensemble admissible :

$$\mathcal{U}_{\text{ad}} = \{h \in L^\infty(\Omega), 0 < h_{\min} \leq h(x) \leq h_{\max} \text{ dans } \Omega\}$$

et l'on s'intéresse à la fonction objectif :

$$\inf_{h \in \mathcal{U}_{\text{ad}}} J(h) = \frac{1}{2} \int_{\Omega} |\sigma(x) - \sigma_0(x)|^2$$

où $\sigma = h \nabla u$ est le vecteur des contraintes, et $\sigma_0(x) \in L^2(\Omega)^N$ une contrainte cible donnée.

Pour écrire la formulation variationnelle du problème, on introduit l'espace de Hilbert $V = \{u \in H^1(\Omega) / u = 0 \text{ sur } \Gamma_D\}$. Il s'agit donc de trouver $u \in V$ tel que :

$$\int_{\Omega} h \nabla u \cdot \nabla v = \int_{\Omega} f v, \quad \forall v \in V$$

Le Lagrangien est :

$$\mathcal{L}(h, v, q) = \frac{1}{2} \int_{\Omega} |\sigma - \sigma_0|^2 + \int_{\Omega} h \nabla v \cdot \nabla q - \int_{\Omega} f q$$

où $(h, v, q) \in \mathcal{U}_{\text{ad}} \times V \times V$.

La formulation variationnelle de l'équation adjointe est donnée, par définition, par :

$$\left\langle \frac{\partial \mathcal{L}}{\partial v}(h, u, p), \varphi \right\rangle = 0, \quad \forall \varphi \in V$$

i.e. :

$$\int_{\Omega} h \nabla p \cdot \nabla \varphi + \int_{\Omega} (\sigma - \sigma_0) \cdot h \nabla \varphi = 0, \quad \forall \varphi \in V$$

Par conséquent, l'état adjoint $p \in V$ est solution de :

$$\begin{cases} -\operatorname{div}(h \nabla p) = \operatorname{div}(h(\sigma - \sigma_0)) & \text{dans } \Omega \\ p = 0 & \text{sur } \Gamma_D \\ h \frac{\partial p}{\partial n} = -h(\sigma - \sigma_0) & \text{sur } \Gamma_N \end{cases}$$

Cette dernière condition pouvant se réécrire :

$$h \frac{\partial p}{\partial n} = h \sigma_0 \cdot n \quad \text{sur } \Gamma_N$$

car $\partial u / \partial n = 0$ sur Γ_N .

La dérivée de la fonction objectif $J(h)$ est donnée, pour tout $k \in L^\infty(\Omega)$, par :

$$\langle J'(h), k \rangle = \int_{\Omega} J'(h) k = \left\langle \frac{\partial \mathcal{L}}{\partial h}(h, u, p), k \right\rangle = \int_{\Omega} (\sigma - \sigma_0) \cdot k \nabla u + \int_{\Omega} k \nabla u \cdot \nabla p$$

soit : $J'(h) = (\sigma - \sigma_0) \cdot \nabla u + \nabla u \cdot \nabla p$.

16.3.6 Optimisation géométrique en rigidité sous tension prescrite

On reprend le cas précédent, mais on impose cette fois la contrainte normale sur Γ_N : On considère une membrane dont le bord est partitionné en deux parties de mesures non nulles $\partial\Omega = \Gamma_D \cup \Gamma_N$. Cette membrane est d'épaisseur constante non variable (et donc on pourra la faire disparaître des équations), et n'est soumise à aucune force dans Ω . Elle est fixée sur Γ_D et est soumise à une contrainte normale prescrite sur le bord Γ_N . Le problème à résoudre est donc :

$$\begin{cases} -\Delta u = 0 & \text{dans } \Omega \\ u = 0 & \text{sur } \Gamma_D \\ \frac{\partial u}{\partial n} = g \cdot n & \text{sur } \Gamma_N \end{cases}$$

avec $g \in C^1(\mathbb{R}^2)^2$ un champ de vecteur donné.

Le bord Γ_D est fixe, mais le bord Γ_N peut varier. L'ensemble des formes admissibles à volume constant est :

$$\mathcal{U}_{\text{ad}} = \{\Omega \in \mathbb{R}^2 / \Gamma_D \subset \partial\Omega, |\Omega| = V_0\}$$

et l'on optimise la rigidité de la membrane en minimisant sa compliance :

$$\inf_{h \in \mathcal{U}_{\text{ad}}} J(h) = \int_{\Gamma_N} u g \cdot n$$

Pour écrire la formulation variationnelle du problème, on introduit l'espace de Hilbert $V = \{u \in H^1(\Omega) / u = 0 \text{ sur } \Gamma_D\}$. Il s'agit donc de trouver $u \in V$ tel que :

$$\int_{\Omega} \nabla u \cdot \nabla v = \int_{\Gamma_N} g \cdot n v, \quad \forall v \in V$$

Le Lagrangien est :

$$\mathcal{L}(\Omega, v, q) = \int_{\Gamma_N} g \cdot n v + \int_{\Omega} \nabla v \cdot \nabla q - \int_{\Gamma_N} g \cdot n q$$

où les paramètres v et q sont dans $X = \{\varphi \in H^1(\mathbb{R}^2) / \varphi = 0 \text{ sur } \Gamma_D\}$. On remarquera que les trois variables (Ω, v, q) sont bien indépendantes.

La formulation variationnelle de l'équation adjointe est donnée, par définition, par :

$$\left\langle \frac{\partial \mathcal{L}}{\partial v}(\Omega, u, p), \varphi \right\rangle = 0, \quad \forall \varphi \in X$$

i.e. :

$$\int_{\Omega} \nabla \varphi \cdot \nabla p + \int_{\Gamma_N} g \cdot n \varphi = 0, \quad \forall \varphi \in X$$

Par conséquent, l'état adjoint $p \in H^1(\mathbb{R}^2)$ est solution de :

$$\begin{cases} -\Delta p = 0 & \text{dans } \Omega \\ p = 0 & \text{sur } \Gamma_D \\ \frac{\partial p}{\partial n} = -g \cdot n & \text{sur } \Gamma_N \end{cases}$$

Cela implique que $p = -u$ (le problème est auto-adjoint).

La dérivée de la fonction objectif $J(\Omega)$ va faire intervenir la normale n qui dépend de Ω , on élimine cela en réécrivant le Lagrangien sous la forme :

$$\mathcal{L}(\Omega, v, q) = \int_{\Omega} \operatorname{div}(v g) + \int_{\Omega} \nabla u \cdot \nabla q - \int_{\Omega} \operatorname{div}(q g)$$

La dérivée de forme s'obtient donc formellement par :

$$\langle J'(\Omega), \theta \rangle = \int_{\Omega} J'(h) \theta = \left\langle \frac{\partial \mathcal{L}}{\partial \Omega}(\Omega, u, p), \theta \right\rangle = \left\langle \frac{\partial \mathcal{L}}{\partial \Omega}(\Omega, u, -u), \theta \right\rangle$$

En n'oubliant pas que Γ_D est fixe, et que seul Γ_N peut varier, on obtient :

$$J'(\Omega)(\theta) = - \int_{\Gamma_N} \theta \cdot n (|\nabla u|^2 - 2 \operatorname{div}(u g))$$

16.3.7 Optimisation de l'épaisseur pour la première fréquence propre

On s'intéresse toujours au déplacement vertical u de la membrane d'épaisseur variable $h(x)$ avec conditions de Dirichlet sur le bord, mais on va chercher cette fois à ce que ce déplacement vertical $u \in H_0^1(\Omega)$ soit une fonction propre non nulle associée à la valeur propre $\lambda(h) \in \mathbb{R}$. Notre problème est donc :

$$\begin{cases} -\operatorname{div}(h \nabla u) = \lambda(h) u & \text{dans } \Omega \\ u = 0 & \text{sur } \partial \Omega \end{cases}$$

On s'intéresse au mode fondamental de la membrane, i.e. au cas où $\lambda(h)$ est la plus petite valeur propre. Or celle-ci est donnée par :

$$\lambda(h) = \min_{v \in H_0^1(\Omega), v \neq 0} \frac{\int_{\Omega} h |\nabla v|^2}{\int_{\Omega} v^2}$$

L'épaisseur $h(x)$ appartient à l'ensemble admissible :

$$\mathcal{U}_{\text{ad}} = \{h \in L^\infty(\Omega), 0 < h_{\min} \leq h(x) \leq h_{\max} \text{ dans } \Omega\}$$

et l'on veut maximiser la première fréquence propre de vibration. On s'intéresse à la fonction objectif :

$$\inf_{h \in \mathcal{U}_{\text{ad}}} J(h) = -\lambda(h) + \ell \int_{\Omega} h(x)$$

où $\ell \geq 0$ est le multiplicateur de Lagrange fixé pour une contrainte sur l'aire de la membrane.

Intéressons nous d'abord au cas où $\ell = 0$. Dans ce cas, la fonction objectif peut se réécrire :

$$- \sup_{h \in \mathcal{U}_{\text{ad}}} \min_{v \in H_0^1(\Omega), v \neq 0} \frac{\int_{\Omega} h |\nabla v|^2}{\int_{\Omega} v^2}$$

que l'on transforme en problème de minimisation (en sortant le signe $-$).

De plus, comme $h \leq h_{\max}$, on a :

$$\min_{v \in H_0^1(\Omega), v \neq 0} \frac{\int_{\Omega} h |\nabla v|^2}{\int_{\Omega} v^2} \leq \min_{v \in H_0^1(\Omega), v \neq 0} \frac{\int_{\Omega} h_{\max} |\nabla v|^2}{\int_{\Omega} v^2}$$

pour toute épaisseur $h \in \mathcal{U}_{\text{ad}}$. Autrement dit, la fonction constante $h = h_{\max}$ minimise la fonction objectif. Cela signifie simplement qu'en l'absence de contrainte de volume ou de poids, la membrane qui maximise la première fréquence propre est la plus épaisse.

Revenons maintenant au cas où $\ell \neq 0$. Le Lagrangien est :

$$\mathcal{L}(h, \mu, v, q) = -\mu + \ell \int_{\Omega} f(x) + \int_{\Omega} (h \nabla v \cdot \nabla q - \mu v q)$$

avec $(h, \mu, v, q) \in \mathcal{U}_{\text{ad}} \times \mathbb{R} \times H_0^1(\Omega) \times H_0^1(\Omega)$.

Comme on le faisait remarquer, l'équation d'état se retrouve en dérivant le Lagrangien par rapport à q . La formulation variationnelle associée s'écrit :

$$\left\langle \frac{\partial \mathcal{L}}{\partial q}(h, \lambda, u, p), \varphi \right\rangle = 0, \quad \forall \varphi \in H_0^1(\Omega)$$

Le Lagrangien étant symétrique en v et q , sa dérivée par rapport à v est identique à celle par rapport à q . L'équation d'état pour l'état adjoint est la même que pour l'équation d'état (le problème est auto-adjoint). Néanmoins, la fonction propre n'étant définie qu'à une constante multiplicative près, on peut simplement dire que p et u sont proportionnels. Pour connaître cette constante de proportionnalité, on dérive le Lagrangien par rapport à μ et, à l'optimum, cette dérivée étant nulle, on obtient :

$$-1 - \int_{\Omega} u p = 0$$

d'où :

$$p = -\frac{u}{\int_{\Omega} u^2}$$

La dérivée de la fonction objectif $J(h)$ est donnée, pour tout $k \in L^\infty(\Omega)$, par :

$$\langle J'(h), k \rangle = \int_{\Omega} J'(h) k = \left\langle \frac{\partial \mathcal{L}}{\partial h}(h, \lambda, u, p), k \right\rangle = \int_{\Omega} k(\ell + \nabla u \cdot \nabla p)$$

$$\text{soit : } J'(h) = \ell - \frac{|\nabla u|^2}{\int_{\Omega} u^2}.$$

Chapitre 17

Problèmes non stationnaires

Résumé — Dans ce chapitre, nous nous intéresserons au cas non stationnaire. Toutefois, nous n'aborderons pas les éléments finis espace-temps, car il s'agit d'une formulation gourmande en ressources, d'où sa très faible utilisation (bien que la méthode soit en elle-même intéressante).

Dans ce chapitre, nous aurons besoin de « dériver numériquement », i.e. de construire des schémas numériques approchant des dérivées. Le chapitre C en annexe permettra à certains de se rafraîchir la mémoire en regardant comment on résout les équations différentielles et aux dérivées partielles... puis numériquement. Nous utiliserons en effet la méthode de Newmark décrite au paragraphe C.2.5.

17.1 Équation non stationnaire de la dynamique

Considérons l'équation de la dynamique, sous forme non stationnaire (i.e. dépendant du temps) :

$$M\ddot{u}(t) + C\dot{u}(t) + Ku(t) = F(t) \quad (17.1)$$

qui sous forme discrétisée par éléments finis sera réécrite :

$$M\ddot{\mathbf{q}}(t) + C\dot{\mathbf{q}}(t) + K\mathbf{q}(t) = \mathbf{F}(t) \quad (17.2)$$

avec les conditions initiales $\mathbf{q}(t = 0) = \mathbf{q}_0$ et $\dot{\mathbf{q}}(t = 0) = \dot{\mathbf{q}}_0$.

On cherche à obtenir la discrétisation \mathbf{q} du champ u et ses dérivées temporelles $\dot{\mathbf{q}}$ et $\ddot{\mathbf{q}}$ à différents instants t et vérifiant l'équation de la dynamique.

Pour résoudre un tel problème, on se propose de réaliser une discrétisation temporelle qui pourra être explicite ou implicite.

Le schéma généralement utilisé pour la discrétisation temporelle est celui de Newmark, qui est présenté dans le cas de la résolution des équations différentielles au paragraphe C.2.5. Le schéma le plus général est, Δ_t représentant le pas de temps :

$$\begin{aligned} \mathbf{q}_{t+\Delta_t} &= \mathbf{q}_t + \Delta_t \dot{\mathbf{q}}_t + \frac{\Delta_t^2}{2} ((1-2\beta)\ddot{\mathbf{q}}_t + 2\beta\ddot{\mathbf{q}}_{t+\Delta_t}) \\ \dot{\mathbf{q}}_{t+\Delta_t} &= \dot{\mathbf{q}}_t + \Delta_t ((1-\gamma)\ddot{\mathbf{q}}_t + \gamma\ddot{\mathbf{q}}_{t+\Delta_t}) \end{aligned} \quad (17.3)$$

que nous avons écrit sous la forme discrétisée, mais qui serait valable pour l'approximation continue (comme présenté au chapitre au paragraphe C.2.5).

Les différents schémas de Newmark correspondent à des valeurs particulières de β et γ .

17.2 Schéma explicite : différences finies centrées

Dans le cas où $\beta = 0$ et $\gamma = 1/2$, on retombe sur le schéma des différences finies centrées.

On obtient :

$$\begin{aligned}\ddot{\mathbf{q}}_t &= \frac{1}{\Delta_t^2} (\mathbf{q}_{t+\Delta_t} - 2\mathbf{q}_t + \mathbf{q}_{t-\Delta_t}) \\ \dot{\mathbf{q}}_t &= \frac{1}{2\Delta_t} (\mathbf{q}_{t+\Delta_t} - \mathbf{q}_{t-\Delta_t})\end{aligned}\quad (17.4)$$

et l'équation de la dynamique discrétisée s'écrit sous la forme d'une équation modifiée :

$$\bar{\mathbf{K}}\mathbf{q}_{t+\Delta_t} = \mathbf{R} \quad (17.5)$$

avec :

$$\bar{\mathbf{K}} = \frac{1}{\Delta_t^2} \mathbf{M} + \frac{1}{2\Delta_t} \mathbf{C} \quad (17.6)$$

et :

$$\mathbf{R} = \mathbf{F}_t - \mathbf{K}\mathbf{q}_t + \frac{1}{\Delta_t^2} \mathbf{M} (2\mathbf{q}_t - \mathbf{q}_{t-\Delta_t}) + \frac{1}{2\Delta_t} \mathbf{C}\mathbf{q}_{t-\Delta_t} \quad (17.7)$$

qui ne dépendent bien que de données disponibles. Le calcul se déroule donc comme suit :

- Solution initiale à $t = 0$:
 - connaissant \mathbf{q}_0 et $\dot{\mathbf{q}}_0$, on calcule $\ddot{\mathbf{q}}_0$ en résolvant l'équation « classique » de la dynamique ;
 - on calcule ensuite $\mathbf{q}_{-\Delta_t} = \mathbf{q}_0 - \Delta_t \dot{\mathbf{q}}_0 + \frac{\Delta_t^2}{2} \ddot{\mathbf{q}}_0$.
- On dispose alors de tous les éléments pour, à chaque pas de temps $t + \Delta_t$:
 - résoudre l'équation modifiée, qui nous fournit $\mathbf{q}_{t+\Delta_t}$;
 - puis obtenir $\dot{\mathbf{q}}_t$ et $\ddot{\mathbf{q}}_t$ par le schéma de Newmark adopté, ici celui des différences finies centrées.

On peut faire les remarques suivantes sur ce type de calcul :

- Pour un pas de temps donné, \mathbf{q}_t ne dépend que des données du temps passé, on a donc une résolution vectorielle rapide.
- Si les matrices \mathbf{M} et \mathbf{C} sont diagonales, alors cette méthode est très efficace même pour les problèmes de grande taille.
- Ce schéma est inconditionnellement stable si $\Delta_t \leq T_{\min}/\pi$, avec T_{\min} la plus petite période du système correspondant à l'équation de la dynamique (classique).
- la précision est de l'ordre de Δ_t^2 .
- L'amortissement numérique est nul.
- Il est possible d'introduire un amortissement numérique pour contrôler les hautes fréquences. Dans ce cas, il faut considérer le schéma avec $\beta = 0$ et $\gamma > 1/2$. Il n'y a pas automatiquement stabilité du schéma, celle-ci est à calculer pour chaque schéma.

17.3 Schéma implicite : schéma de Newmark classique

Dans le cas où $\beta = 1/4$ et $\gamma = 1/2$, on obtient le schéma implicite de Newmark, qui est celui qui est utilisé généralement pour l'analyse dynamique des structures.

Il vient alors :

$$\begin{aligned}\ddot{\mathbf{q}}_{t+\Delta_t} &= \frac{4}{\Delta_t^2} (\mathbf{q}_{t+\Delta_t} - \mathbf{q}_t) - \frac{4}{\Delta_t} (\dot{\mathbf{q}}_t - \dot{\mathbf{q}}_t) \\ \dot{\mathbf{q}}_{t+\Delta_t} &= \dot{\mathbf{q}}_t + \frac{\Delta_t}{2} (\ddot{\mathbf{q}}_{t+\Delta_t} + \ddot{\mathbf{q}}_t)\end{aligned}\quad (17.8)$$

et l'équation de la dynamique discrétisée s'écrit encore sous la forme d'une équation modifiée :

$$\bar{\mathbf{K}}\mathbf{q}_{t+\Delta_t} = \mathbf{R} \quad (17.9)$$

mais cette fois avec :

$$\bar{\mathbf{K}} = \mathbf{K} + \frac{4}{\Delta_t^2}\mathbf{M} + \frac{2}{\Delta_t}\mathbf{C} \quad (17.10)$$

et :

$$\mathbf{R} = \mathbf{F}_{t+\Delta_t} + \mathbf{M}\left(\frac{4}{\Delta_t^2}\mathbf{q}_t + \frac{4}{\Delta_t}\dot{\mathbf{q}}_t + \ddot{\mathbf{q}}_t\right) + \mathbf{C}\left(\frac{2}{\Delta_t}\mathbf{q}_t + \dot{\mathbf{q}}_t\right) \quad (17.11)$$

qui dépendent également des données au même pas de temps. Le calcul se déroule donc comme suit :

- Solution initiale à $t = 0$:
 - connaissant \mathbf{q}_0 et $\dot{\mathbf{q}}_0$, on calcule $\ddot{\mathbf{q}}_0$ en résolvant l'équation « classique » de la dynamique ;
 - on construit $\bar{\mathbf{K}}$ et, si \mathbf{M} , \mathbf{C} , \mathbf{K} et Δ_t sont constants (ce qui est généralement le cas), la triangulariser.
- À chaque pas de temps $t + \Delta_t$:
 - calculer \mathbf{R} (que l'on appelle le résidu, d'où le choix de la notation) ;
 - calculer $\bar{\mathbf{K}}$ et triangulariser si nécessaire ;
 - résoudre l'équation modifiée, qui nous fournit $\mathbf{q}_{t+\Delta_t}$;
 - puis obtenir $\dot{\mathbf{q}}_t$ et $\ddot{\mathbf{q}}_t$ par le schéma de Newmark adopté, ici celui de Newmark implicite.

On peut faire les remarques suivantes sur ce type de calcul :

- Pour un pas de temps donné, \mathbf{q}_t dépend également des données du même pas de temps, on a donc une résolution matricielle coûteuse.
- Ce schéma est inconditionnellement stable, et donc comme on peut utiliser de plus grands pas de temps, on réduit le coût mentionné à la ligne précédente.
- la précision est de l'ordre de Δ_t^2 , et donc comme on ne peut pas utiliser de trop grands pas de temps sans réduire la précision...
- L'amortissement numérique est nul.
- Il est possible d'introduire un amortissement numérique pour contrôler les hautes fréquences. Dans ce cas, il faut considérer le schéma avec $\gamma > 1/2$ et $\beta = (\gamma + \frac{1}{2})^2$. On obtient encore un schéma stable.

17.4 Comparaison des méthodes explicite et implicite

Concernant la méthode explicite, on retiendra d'abord que cette méthode nécessite moins de mémoire, est donc plus rapide et mieux adaptée aux problèmes de grandes tailles. Comme cette méthode nécessite moins de mémoire et des petits pas de temps (pour la stabilité), elle est bien adaptée au cas des chocs. Comme elle n'est que conditionnellement stable, elle est plutôt adaptée à la résolution élément par élément, donc au traitement local. Enfin, cette méthode est très robuste numériquement et permet de traiter le cas de non linéarités couplées.

Concernant la méthode implicite, on retiendra que puisque cette méthode est inconditionnellement stable, elle est bien adaptée à la résolution de problèmes globaux (qui nécessitent qu'il y ait convergence). De plus, cette méthode est bien moins robuste numériquement que la précédente (pivots nuls, divergence)... Enfin et surtout, on se souviendra que c'est une méthode coûteuse en mémoire et en temps...

La figure 17.1 donne, de manière imagée, les domaines d'utilisation des méthodes implicite et explicite.

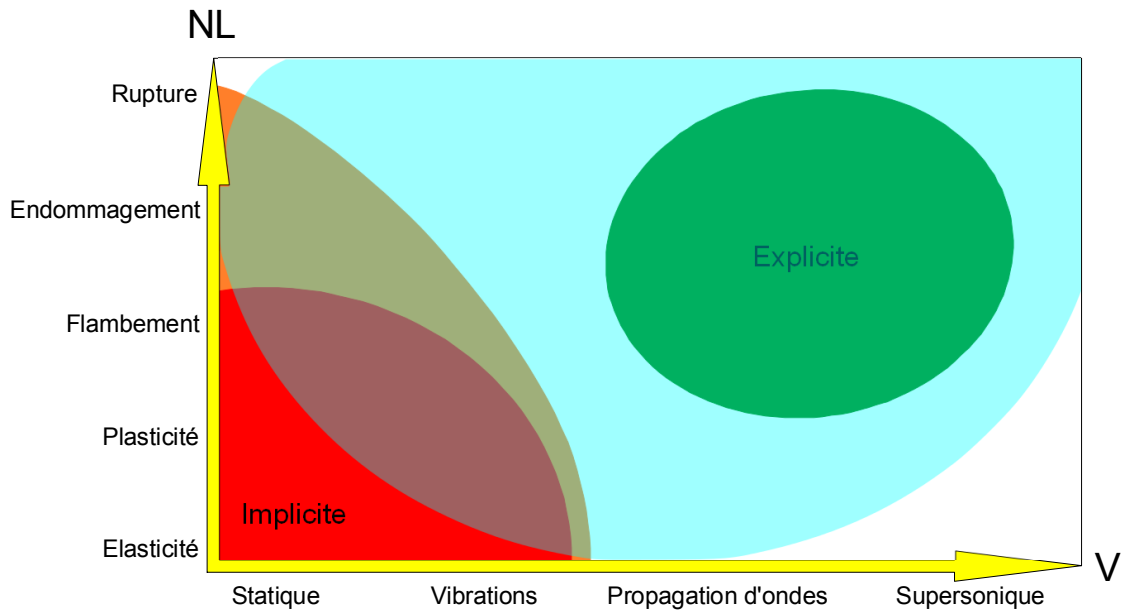


FIGURE 17.1: Domaines d'utilisation des méthodes implicite et explicite

17.5 Exemple : un calcul de propagation avec FREEFEM++

Quelque soit le logiciel utilisé¹, il y a toujours un grand nombre d'exemples traités fournis avec, sans compter le nombre d'ouvrages disponibles.

Nous avons souhaité néanmoins présenter un listing (qui sera discuté oralement) correspondant à un calcul réalisé sous FREEFEM++. Ce logiciel gratuit vous permettra de faire vos calculs par éléments finis... mais c'est également un outil pédagogique sans équivalent : sa manipulation se fait en des termes proches de la formulation mathématique, et en cela, il est une passerelle parfaite entre la théorie et la pratique de codes plus « industriels ». Le fait de pouvoir discrétiser à loisir une formulation variationnelle que l'on rentre soi-même est également un argument qui nous semble un vrai plus, même si cela commence à exister dans des codes commerciaux (mais reste confidentiel).²

Voici donc un petit listing sur lequel nous pourrions revenir de vive voix. Commençons donc, comme il se doit, par définir la géométrie du problème, paramétrée par a et b , que nous maillons :

```
1 real mu=1.;
2 real ro=0. , ro=1.;
3 // taille du maillage
4 int ncuts = 20;
5
6 // construction du domaine
7 real a=pi*0.8, b=1.2*pi ;
8
9 border gam1(t=0.,a){x=t;y=0; label =1;};
10 border gam2(t=0.,b){x=a;y=t; label =2;};
11 border gam3(t=a,0.){x=t;y=b; label =3;};
12 border gam4(t=b,0.){x=0.;y=t; label =4;};
13
14 // Maillage
15 mesh Th=buildmesh(gam1(ncuts)+gam2(ncuts)+gam3(ncuts)+gam4(ncuts));
16 // plot(Th, ps="mesh.eps") ;
```

1. Lorsque l'on parle de codes éléments finis, on cite les grands noms du domaine... et ils correspondent à des codes extraordinairement puissants, mais souvent chers.

Or un certain nombre de codes éléments finis professionnels sont disponibles gratuitement (au téléchargement et à l'utilisation). Les deux plus connus sont CAST3M (anciennement castem) et CODE ASTER développés et maintenus par le CEA et par EDF respectivement.

2. Notons que pour ceux qui préfèrent, il existe aussi RHEOLEF, un environnement C++ pour le calcul par éléments finis, qui reste proche, dans sa philosophie, de ce que propose FREEFEM++. Nous ne le connaissons pas assez pour fournir des exemples dans ce document.

```
17 // plot(Th, wait=1) ;
```

Nous définissons quelques valeurs afin de pouvoir définir le chargement (le chargement défini par la fonction w_1 ne sera pas utilisée dans le calcul effectué). Notons comme il est aisé de définir des fonctions :

```
18 // conditions aux limites essentielles
19 real r0=a/10., r, xc=a/3., yc=3.*b/4.;
20
21 // chargement initial sinusoidal
22 func real pertb(real r)
23 { if(r<r0) {return (1.+cos(pi*r/r0))
24   ;}
25   else return 0.;}
26 //func real w1( real x, real y)
27 func real w11( real x, real y)
28 {
29   return pertb(sqrt((x-xc)^2+(y-yc)^2)
30   );
31 }
32 // chargement constant sur un carre

33 // pas utilise sur cet exemple
34 func real w1( real c, real d)
35 {
36   if(c>(0.8*xc))
37     {if (c<(1.2*xc))
38       {if (d>(0.8*yc))
39         {if (d<(1.2*yc)) {return 0.8;}
40         else return 0.;}
41       else return 0.;}
42     else return 0.;}
43   else return 0.;
44 }
45
46 func real w0( real x, real y)
47 {
48   return 0.*sin(x)*sin(y) ;
49 }
```

Ensuite, nous définissons l'espace des éléments finis utilisés, et les variables appartenant à ces espaces. L'analogie avec la « formulation mathématique » saute aux yeux... et c'est pourquoi cet outil nous semble particulièrement intéressant sur le plan pédagogique.

```
50 // Espaces elements finis
51 fespace Vh(Th,P2);
52 fespace Wh(Th,P1dc);
53
54 Vh w, wa, wd, wda, wdd, wdda, fi ;
55 Wh sxz, syz ;
```

Il s'agit d'un exemple non stationnaire, nous allons donc faire une boucle sur le temps. Dans FREEFEM++, nous entrons directement la formulation variationnelle du problème sous forme « mathématique »... le lien entre pratique des éléments finis et théorie est beaucoup plus clair. La boucle de temps est basée sur un schéma de Newmark implicite. Le temps « courant » est le temps $t + \Delta_t$ et le temps précédent est le temps t pour rester cohérent avec les notations du chapitre 17.

Il s'agit d'un problème de déplacement hors plan, dont le champ inconnu est traditionnellement noté w . Ainsi, dans le listing, w représente $q_{t+\Delta_t}$, $wa \equiv q_t$, $wd \equiv \dot{q}_{t+\Delta_t}$, $wda \equiv \dot{q}_t$, $wdd \equiv \ddot{q}_{t+\Delta_t}$ et $wdda \equiv \ddot{q}_t$. $fi \equiv \varphi$ est la fonction test. La formulation variationnelle, en n'indiquant pas le temps actuel et en utilisant l'indice a pour le pas de temps précédent (a comme avant), est :

$$\begin{aligned}
 \text{antiplane}(\dot{w}, \varphi) = & \int_{Th} \rho \dot{w} \varphi + \frac{\Delta_t^2}{4} \int_{Th} \mu \left(\frac{\partial \dot{w}}{\partial x} \frac{\partial \varphi}{\partial x} + \frac{\partial \dot{w}}{\partial y} \frac{\partial \varphi}{\partial y} \right) \\
 & - \frac{\Delta_t}{2} \int_{Th} \rho_g \varphi - \int_{Th} \rho \left(\dot{w}_a + \frac{\Delta_t}{2} \ddot{w}_a \right) \varphi \\
 & + \frac{\Delta_t}{2} \int_{Th} \mu \left(\frac{\partial w_a}{\partial x} \frac{\partial \varphi}{\partial x} + \frac{\partial w_a}{\partial y} \frac{\partial \varphi}{\partial y} \right) \\
 & + \frac{\Delta_t^2}{4} \int_{Th} \mu \left(\frac{\partial \dot{w}_a}{\partial x} \frac{\partial \varphi}{\partial x} + \frac{\partial \dot{w}_a}{\partial y} \frac{\partial \varphi}{\partial y} \right) \\
 & + \text{Condition } (\dot{w} = 0 \text{ sur les lignes } 1, 2, 3, 4)
 \end{aligned} \tag{17.12}$$

```

56 int n, k, Ntemps=100;
57 real T=2.*pi/sqrt(2.), pastemps=T/Ntemps
58 real dpt=0.5*pastemps;
59
60 problem antiplane(wd,fi, init=1)=int2d(Th)(ro*wd*fi)+
61     int2d(Th)(dpt^2*mu*(dx(wd)*dx(fi)+dy(wd)*dy(fi))) +
62     int2d(Th)(-dpt*rog*fi)+
63     int2d(Th)(ro*(-wda-dpt*wdda)*fi)+
64     int2d(Th)(dpt*mu*(dx(wa)*dx(fi)+dpt*dx(wda)*dx(fi)+
65     dy(wa)*dy(fi)+dpt*dy(wda)*dy(fi)))+
66     on(1,2,3,4,wd=0.);
67
68 // formulation du probleme en temps
69 // conditions initiales en temps
70 w=w0(x,y);
71 wd=w11(x,y);
72 wdd=0.;
73
74 real errorw, errorwd, temps, enerC, enerP, enerT;
75 real[int] visoS(20);
76 int ivi;
77 for (ivi=0;ivi<10;ivi++){
78     visoS[ivi]=-1+0.1*ivi;
79     visoS[ivi+10]=(ivi+1)*0.1;
80 }

```

Enfin, pour tester la qualité des résultats, il nous faut quelques indicateurs... Sans entrer dans le détail, des noms comme `enerC`, `enerP`, et `enerT` doivent vous mettre sur la piste.

```

81 real[int] Ec(Ntemps), Ep(Ntemps), Et(Ntemps), tt(Ntemps);
82
83 for (n=0;n<Ntemps;n++){
84     wa=w; wda=wd; wdda=wdd;
85     temps=n*pastemps;
86
87     enerC=0.5*int2d(Th)(ro*wd^2);
88     enerP=0.5*int2d(Th)(mu*(dx(w)*dx(w)+ dy(w)*dy(w)));
89     enerT=enerC+enerP;
90     Ec(n)=enerC; Ep(n)=enerP; Et(n)=enerT; tt(n)=temps;
91
92     cout << " iteration n= " << n << " enerP= " << enerP <<
93     "enerC= " << enerC << " enerTotale= " << enerT
94     << endl;
95
96     temps=(n+1.)*pastemps;
97
98 // resolution du probleme
99     antiplane;
100
101     w=wa+dpt*(wda+wd);
102     wdd=(wd-wda)/dpt-wdda;
103
104     sxz=mu*dx(w);
105     syz=mu*dy(w);
106     plot(Th,wd,fill=true, value=1, viso=visoS, nbiso=visoS.n,
107     ps=n, wait=0);
108 }
109
110 // Energie
111 // plot([tt,Ec],[tt,Ep],[tt,Et], ps="energie.eps");

```

À la figure 17.2 se trouve une illustration de ce que nous venons de calculer.

17.6 Décomposition modale

Pour le public visé, le lecteur devrait, à ce moment du document, se demander pourquoi il n'a pas encore été question de projection sur les modes propres.

On peut montrer que les modes de vibration de la structure forment une base. Il peut paraître intéressant de chercher la solution en temps du problème sous forme d'une approximation par

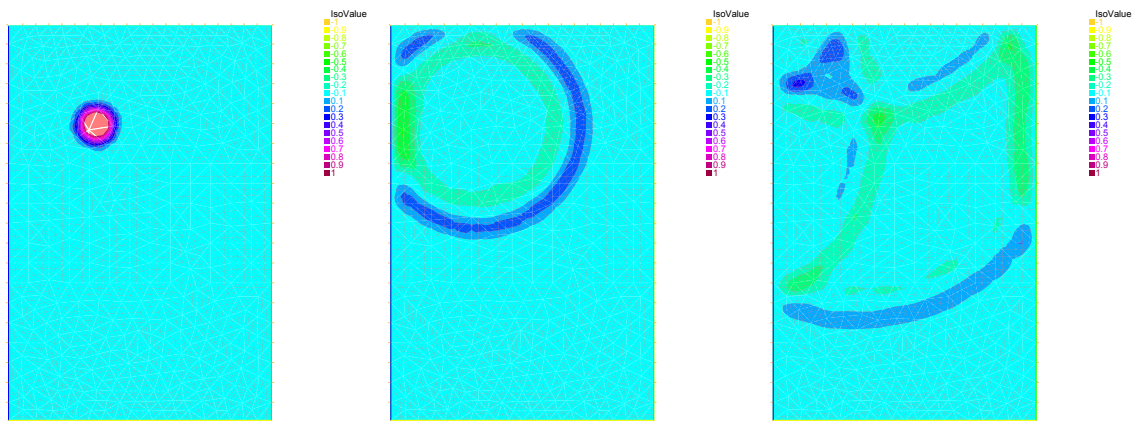


FIGURE 17.2: Propagation : pas de temps 0, 20 et 40

projection sur la base des quelques N premiers modes propres, i.e. de rechercher N fonctions scalaires du temps.

Lorsque l'on dispose déjà des N premiers modes et lorsque le nombre n de degrés de liberté est important, cette approche peut s'avérer particulièrement efficace. Par ailleurs, lorsque les modes sont associés à des mouvements simples de la structure, il peut être facile pour un ingénieur un peu expérimenté d'imaginer le nombre de modes à utiliser pour représenter le phénomène.

La méthode de la décomposition modale sera exposée au paragraphe 18.3.4, lorsque nous aurons fait quelques rappels sur les modes propres.

Chapitre 18

Les ondes

Résumé — Jusqu'ici, nous n'avons quasiment pas parlé de mode propre cela est plus ou moins évoqué à travers, par exemple, « la plus petite période du système » au chapitre précédent... mais cela a été fait à dessein.

En effet, nous avons voulu, dans ce chapitre, regrouper les approches modales, car cela nous a semblé plus en cohérence.

18.1 Introduction

Dans la mesure où se document s'adresse à des ingénieurs en mécanique (ou en acoustique), la notion de mode propre est une notion relativement maîtrisée.

Un système mécanique atteint un mode propre de vibration lorsque tous les points de ce système sont à une fréquence donnée appelée fréquence propre du système. Une fréquence propre est fondamentale si elle n'est pas le multiple d'une autre fréquence propre ; dans le cas contraire, c'est une harmonique.

On appelle résonance le phénomène selon lequel certains systèmes physiques sont particulièrement sensibles à certaines fréquences. Un système résonant peut accumuler une énergie, si celle-ci est appliquée sous forme périodique, et proche d'une fréquence propre. Soumis à une telle excitation, le système est alors le siège d'oscillations de plus en plus importantes, jusqu'à atteindre un régime d'équilibre qui dépend des éléments dissipatifs du système, ou bien jusqu'à rupture d'un composant du système.

Si l'on soumet un système résonant à une percussion (pour les systèmes mécaniques) ou à une impulsion (pour les systèmes électriques), et non plus à une excitation périodique, alors le système sera le siège d'oscillations amorties sur des fréquences proches de ses fréquences propres et retournera progressivement à son état stable. Physiquement, c'est le coup de marteau de choc donné sur une structure pour en déterminer les modes.

Un système susceptible d'entrer en résonance, i.e. susceptible d'être le siège d'oscillations amorties, est un oscillateur. Un tel système a la particularité de pouvoir emmagasiner temporairement de l'énergie sous deux formes : potentielle ou cinétique. L'oscillation est le phénomène par lequel l'énergie du système passe d'une forme à l'autre, de façon périodique.

Si l'on injecte une énergie potentielle au moment où l'énergie potentielle déjà emmagasinée est maximale, l'énergie ainsi injectée s'ajoute à l'énergie déjà emmagasinée et l'amplitude de l'oscillation va augmenter, ainsi que l'énergie totale. Idem pour l'énergie cinétique. Ainsi, si l'on apporte de l'énergie avec une périodicité égale (ou proche) de la périodicité propre du système, l'énergie totale va augmenter régulièrement et l'amplitude des oscillations du système va ainsi croître. L'exemple le plus simple est celui d'une balançoire : l'énergie de chaque poussée s'ajoute à l'énergie totale, à condition de pousser au bon moment...

Le phénomène de résonance n'est rien d'autre que cet effet d'accumulation de l'énergie en injectant celle-ci au moment où elle peut s'ajouter à l'énergie déjà accumulée, i.e. « en phase » avec cette dernière.

Quand l'excitation aura cessé, le système résonant sera le siège d'oscillations amorties : il va revenir plus ou moins vite à son état d'équilibre stable. En effet, l'énergie de départ sera peu à peu absorbée par les éléments dissipatifs du système (amortisseur visqueux en mécanique, résistances en électricité...). Un système peu amorti sera le siège d'un grand nombre d'oscillations qui diminueront lentement avant de disparaître complètement.

La représentation modale est pertinente dans le domaine des basses fréquences, i.e. pour les premiers modes propres. Dans les domaines moyennes et hautes fréquences, on utilise des méthodes adaptées à la densité spectrale élevée.

Les domaines moyennes fréquences et hautes fréquences sont définis par la densité spectrale. En effet, l'expression en fréquences n'a pas de sens pour définir ces domaines, une similitude sur un système physique modifie les fréquences propres mais le spectre reste semblable, à un facteur près. Dans le cas de fréquences multiples, il existe un sous-espace propre donc les modes propres sont arbitraires dans ce sous espace. Dans le cas de fréquences voisines (densité spectrale élevée), la représentation modale n'est pas robuste car de faibles perturbations du domaine physique vont entraîner un changement important des modes propres associés à ces fréquences. Donc la représentation modale n'est pertinente que pour le domaine des basses fréquences, domaine défini par la densité spectrale. Le domaine basses fréquences s'étendra jusqu'à quelques Hz en génie civil, jusqu'à des milliers de Hz pour de petites structures mécaniques.

Le phénomène de synchronisation, ou accrochage de fréquences est un phénomène par lequel deux systèmes excités chacun selon une fréquence se mettent à osciller selon la même fréquence. On trouve de nombreux exemples de ce phénomène dans la nature :

- Le plus connu et le plus facilement observable concerne la Lune : celle-ci présente toujours la même face à la Terre. Cela signifie que la période de rotation de la Lune sur elle-même T_0 est égale à la période de rotation de la lune autour de la Terre $T \cong 28$ jours. Il s'agit d'une résonance 1 : 1. L'analyse montre que ce n'est pas une coïncidence, et que cela est dû à un faible couplage gravitationnel entre ces deux mouvements.
- Le règne animal n'est pas en reste et fournit lui-aussi des exemples de synchronisation, auxquels on ne pense pas spontanément. Citons le vol d'oiseaux ou le clignotement des lucioles...

Histoire

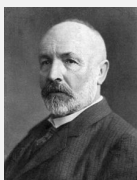
On trouve d'autres exemples historiques relatifs à ce phénomène de synchronisation. En voici deux parmi les plus connus.

Le premier est celui de la synchronisation des balanciers de deux pendules accrochées au même mur d'une pièce. Un faible couplage par les vibrations transmises dans le mur, et une dissipation faible, expliquent cet accrochage de fréquences et cette résonance 1 : 1.

C'est Huygens qui a remarqué et expliqué ce phénomène : le système composé des deux balanciers et du mur a deux fréquences voisines faiblement couplées, et il possède, par le couplage, deux modes propres correspondant aux mouvements en phase et en opposition de phase des deux pendules. C'est sur le premier mode que se produit la synchronisation.



Huygens



Cantor

L'escalier de Cantor, ou escalier du diable est un exemple mathématique incontournable en analyse. Il correspond au graphe d'une fonction continue f , sur $[0; 1]$, telle que $f(0) = 0$, $f(1) = 1$, qui est dérivable presque partout, la dérivée étant presque partout nulle.

L'escalier de Cantor peut également être vu comme la fonction de répartition d'une variable aléatoire réelle continue qui n'est pas à densité, et qui est même étrangère à la mesure de Lebesgue.

Enfin, l'escalier du diable peut aussi être vu comme résultant d'un phénomène de synchronisation. Si l'on change un paramètre extérieur du système de façon lente et continue, par exemple l'amplitude $\alpha_0 \in \mathbb{R}$, alors la valeur de l'accrochage $a = p/q$ va dépendre de ce paramètre. On obtient alors une fonction $a(\alpha_0)$ de $\mathbb{R} \rightarrow \mathbb{Q}$. Cette fonction très étrange comporte une multitude de paliers plus ou moins larges... et correspond à l'escalier du diable.

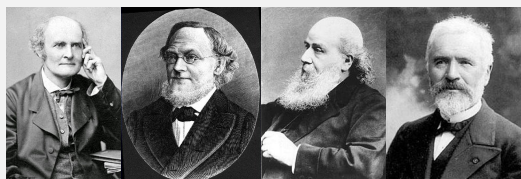
En mathématiques, le concept de vecteur propre est une notion portant sur une application linéaire d'un espace dans lui-même. Il correspond à l'étude des axes privilégiés, selon lesquels l'application se comporte comme une dilatation, multipliant les vecteurs par une même constante. Ce rapport de dilatation est appelé valeur propre ; les vecteurs auxquels il s'applique s'appellent vecteurs propres, réunis en un espace propre.

Histoire

Bien qu'existant sous une forme non formalisée depuis longtemps, il aura fallu attendre l'invention des structures algébriques nécessaires pour vraiment pouvoir parler des valeurs propres (issues par exemple de la clôture algébrique de \mathbb{C} démontrée par Gauß).

L'exemple immédiat qui vient à l'esprit est le traitement de l'équation de la chaleur par Fourier qui utilise déjà une base de vecteurs propres, bien que le concept n'ait pas encore été défini. Hamilton introduira la notion de polynôme caractéristique, ce qui permet de déterminer ce que l'on appelle maintenant les valeurs propres associées à l'endomorphisme d'une équation différentielle linéaire.

Plusieurs aller-retour permettront de définir les notions d'espace vectoriel (Cayley, Grassmann, Cauchy), de matrice (Sylvester, Cayley) et de valeurs propres (Sylvester, Jordan). Hilbert finalement fera prendre conscience de la profondeur de la notion de valeur propre. L'analyse fonctionnelle naît dans la foulée, et elle est l'objet de la première partie de ce document.



Cayley

Grassmann

Sylvester

Jordan

18.2 Notions de valeur, vecteur, mode et fréquence propres

Étant donné une matrice carrée \mathbf{A} d'ordre n (à coefficients dans un anneau commutatif), on cherche un polynôme dont les racines sont précisément les valeurs propres de \mathbf{A} . Ce polynôme est appelé polynôme caractéristique de \mathbf{A} et est défini par :

$$p_{\mathbf{A}}(X) := \det(X\mathbf{I}_n - \mathbf{A}) \quad (18.1)$$

avec X l'indéterminée du polynôme et \mathbf{I}_n la matrice identité d'ordre n . Si λ est une valeur propre de \mathbf{A} , alors il existe un vecteur propre \mathbf{V} non nul tel que $\mathbf{A}\mathbf{V} = \lambda\mathbf{V}$, i.e. tel que l'on ait $(\lambda\mathbf{I}_n - \mathbf{A})\mathbf{V} = \mathbf{0}$. Puisque \mathbf{V} est non nul, cela implique que la matrice $(\lambda\mathbf{I}_n - \mathbf{A})$ est singulière, donc de déterminant nul. Cela montre que les valeurs propres de \mathbf{A} sont des zéros de la fonction $\lambda \mapsto \det(\lambda\mathbf{I}_n - \mathbf{A})$ i.e. des racines du polynôme $\det(X\mathbf{I}_n - \mathbf{A})$. La propriété la plus importante des polynômes caractéristiques est que les valeurs propres de \mathbf{A} sont exactement les racines du polynôme $p_{\mathbf{A}}(X)$.

Quelques propriétés importantes :

- $p_{\mathbf{A}}(X)$ est un polynôme unitaire (coefficient dominant égal à 1) et son degré est égale à n .
- \mathbf{A} et sa transposée ont le même polynôme caractéristique.
- Deux matrices semblables ont le même polynôme caractéristique. (\mathbf{A} et \mathbf{B} sont semblables s'il existe une matrice inversible \mathbf{P} telle que $\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}$). Attention, la réciproque n'est pas vraie en général.

- Si $p_A(X)$ peut être décomposé en produit de facteurs de degré 1, alors A est semblable à une matrice triangulaire (et même à une matrice de Jordan).

Pour aller un peu plus loin. Le théorème de Cayley-Hamilton (dont la première démonstration est due à Frobenius) affirme que tout endomorphisme d'un espace vectoriel de dimension finie n sur un corps commutatif quelconque annule son propre polynôme caractéristique.

En termes de matrice, cela signifie que : si A est une matrice carrée d'ordre n et si $p_A(X)$ est son polynôme caractéristique, alors en remplaçant formellement X par la matrice A dans le polynôme, le résultat est la matrice nulle, i.e. :

$$p_A(A) = A^n + p_{n-1}A^{n-1} + \dots + p_1A + p_0I_n = \mathbb{O}_n \quad (18.2)$$

Cela signifie que le polynôme caractéristique est un polynôme annulateur de A . Les applications sont importantes car le polynôme minimal (qui est l'unique polynôme unitaire qui engendre l'idéal annulateur de l'ensemble des polynômes qui annulent l'endomorphisme dont A est la représentation) cache une décomposition en somme directe de sous-espaces stables.

18.3 Vibration des structures

Revenons sur l'équation de la dynamique sous forme matricielle :

$$M\ddot{q} + C\dot{q} + Kq = F \quad (18.3)$$

qui peut être vue comme l'équation de la statique $Kq = F$ à laquelle on ajoute des forces *extérieures* d'inertie $-M\ddot{q}$ et des forces *extérieures* visqueuses $-C\dot{q}$. D'un point de vue pratique, on distingue trois types de problèmes :

- détermination d'une réponse libre : dans ce cas, la sollicitation est nulle $F = 0$;
- détermination d'une réponse périodique : dans ce cas, la sollicitation F est périodique ;
- détermination d'une réponse transitoire : dans ce cas, la sollicitation F est quelconque.

Dans les deux premiers cas, les conditions initiales du système n'ont aucune importance. On cherche à déterminer une solution générale. On pourrait considérer que le chapitre 17 répond à ces trois types de problèmes, ce qui n'est pas fondamentalement faux. Toutefois, il est plus judicieux de considérer que seul le cas de la dynamique transitoire y a été explicité, et encore uniquement pour l'aspect non modal (qui sera développé dans ce chapitre un peu plus loin).

En effet, dans les cas des vibrations libres (amorties ou non) ou périodique forcées, il est possible d'utiliser la notion de mode, qui n'avait pas été abordée dans les chapitres précédents.

Définition 76 — Méthode spectrale. Une méthode spectrale consiste à transformer le problème considéré en un problème nécessitant de calculer des valeurs et fonctions propres d'un opérateur.

Si l'opérateur en question est linéaire, la fonction dont on cherche à calculer les valeurs (i.e. la solution du problème considéré) peut être exprimée comme combinaison linéaire des fonctions sur lesquelles l'opérateur agit de façon facilement calculable : si ce sont les fonctions propres de l'opérateur on parle vraiment de méthode spectrale, si ce sont d'autres fonctions on parle de méthodes pseudo-spectrales. C'est pourquoi la méthode des éléments finis stochastiques présentée au chapitre 22 est bien elle-aussi une méthode spectrale.

18.3.1 Vibrations libres non amorties

En l'absence de sollicitation et d'amortissement, l'équation de la dynamique devient :

$$M\ddot{q} + Kq = 0 \quad (18.4)$$

dont la solution générale est harmonique et s'écrit :

$$q = \bar{q}e^{i\omega t} \quad (18.5)$$

En injectant la forme de la solution générale dans l'équation de la dynamique, on voit que la pulsation ω est solution du problème de valeurs propres suivant :

$$\mathbf{K}\bar{\mathbf{q}} = \omega^2 \mathbf{M}\bar{\mathbf{q}} \quad (18.6)$$

ce qui conduit à :

$$\det(\mathbf{K} - \omega^2 \mathbf{M}) = 0 \quad (18.7)$$

On obtient ainsi les n valeurs propres $\omega_1, \dots, \omega_n$, où n est la taille du système (i.e. les matrices \mathbf{M} et \mathbf{K} sont $n \times n$). On trouve également les n vecteurs $\bar{\mathbf{q}}_i$ appelés modes propres du système et que l'on norme par rapport à la masse, i.e. tels que :

$${}^T\bar{\mathbf{q}}_i \mathbf{M} \bar{\mathbf{q}}_i = 1 \quad \forall i \in [1, n] \quad (18.8)$$

La détermination des valeurs propres ω_i se fait rarement en cherchant les zéros de l'équation du déterminant en raison de la très grande taille du système dans le cas général, et des considérables différences d'ordre de grandeur entre les valeurs propres.

De toutes façons, ce sont les premières fréquences qui déterminent le comportement du système.

Rappel du cas unidimensionnel L'équation différentielle à résoudre est $M\ddot{u} + Ku = 0$, dont la solution s'écrit :

$$u = A \sin(\omega_0 t) + B \cos(\omega_0 t) \quad (18.9)$$

La pulsation propre du système ω_0 , sa fréquence propre f_0 et sa période propre T_0 sont définies et reliées par les relations :

$$\omega_0 = \sqrt{\frac{K}{M}} \quad f_0 = \frac{\omega_0}{2\pi} \quad T_0 = \frac{1}{f_0} \quad (18.10)$$

Les constantes d'intégration sont déterminées à l'aide des conditions aux limites sur le déplacement u_0 et la vitesse \dot{u}_0 . Il vient :

$$A = \frac{\dot{u}_0}{\omega_0} \quad \text{et} \quad B = u_0 \quad (18.11)$$

Des méthodes permettant de trouver les premiers zéros d'un polynôme de degré n ont donc été mises au point. La majeure partie de ces méthodes consiste à écrire la relation du déterminant sous la forme suivante :

$$\mathbf{H}\mathbf{X} = \lambda \mathbf{X} \quad (18.12)$$

où \mathbf{H} est une matrice définie positive. On se sert pour cela de la décomposition de Cholesky de \mathbf{K} , i.e. en l'écrivant à partir d'une matrice triangulaire inférieure \mathbf{L} sous la forme $\mathbf{K} = \mathbf{L}^T \mathbf{L}$. L'équation du déterminant conduit alors à :

$$\mathbf{L}^{-1} \mathbf{M}^T \mathbf{L}^{-1T} \mathbf{L} \bar{\mathbf{q}} = \frac{1}{\omega^2} {}^T \mathbf{L} \bar{\mathbf{q}} \quad (18.13)$$

et l'on obtient la forme cherchée en posant $\lambda = \omega^{-2}$, $\mathbf{X} = {}^T \mathbf{L} \bar{\mathbf{q}}$ et $\mathbf{H} = \mathbf{L}^{-1} \mathbf{M}^T \mathbf{L}^{-1}$, où \mathbf{H} est bien symétrique.

Si \mathbf{M} et \mathbf{K} sont définies positives (ce qui est le cas habituel des problèmes en dynamique des structures), il existe n valeurs propres réelles positives. Ces solutions sont appelées pulsations propres du système.

Remarque. Si \mathbf{K} est singulière (elle ne possède pas d'inverse), alors, afin de pouvoir utiliser les méthodes précédentes, on utilise un artifice qui consiste à introduire un paramètre $\alpha \in \mathbb{R}$ du même ordre de grandeur que ω^2 . On doit alors résoudre :

$$(\mathbf{K} + \alpha \mathbf{M}) \bar{\mathbf{q}} = (\omega^2 + \alpha) \mathbf{M} \bar{\mathbf{q}} \quad (18.14)$$

La nouvelle matrice $\mathbf{K} + \alpha \mathbf{M}$ est alors inversible et la solution cherchée est $\omega^2 + \alpha$.

18.3.2 Vibrations libres amorties

Problèmes du premier ordre

Si $\mathbf{M} = \mathbb{O}$, l'équation de la dynamique se transforme en celle de la chaleur :

$$\mathbf{C}\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{0} \quad (18.15)$$

dont on cherche une solution générale sous la forme :

$$\mathbf{q} = \bar{\mathbf{q}}e^{-\omega t} \quad (18.16)$$

ce qui conduit au problème de valeurs propres :

$$(\mathbf{K} - \omega\mathbf{C})\bar{\mathbf{q}} = \mathbf{0} \quad (18.17)$$

Les matrices \mathbf{C} et \mathbf{K} sont généralement définies positives, donc ω est réelle positive. La solution présente un terme de décroissance exponentielle qui ne correspond pas réellement à un état de régime permanent.

Problèmes du second ordre

Dans le cas général ($\mathbf{M} \neq \mathbb{O}$), on doit donc résoudre l'équation de la dynamique sans sollicitation :

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{C}\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{0} \quad (18.18)$$

dont on cherche une solution générale sous la forme :

$$\mathbf{q} = \bar{\mathbf{q}}e^{-\alpha t} \quad (18.19)$$

avec $\alpha \in \mathbb{C}$. Cela conduit au problème de valeurs propres :

$$(\alpha^2\mathbf{M} + \alpha\mathbf{C} + \mathbf{K})\bar{\mathbf{q}} = \mathbf{0} \quad (18.20)$$

où $\bar{\mathbf{q}} \in \mathbb{C}$.

La partie réelle de la solution représente une vibration amortie. Ce problème est plus délicat à résoudre que les précédents si bien que la résolution explicite est peu courante.

Rappel du cas unidimensionnel L'équation différentielle à résoudre est : $M\ddot{u} + C\dot{u} + Ku = 0$, que l'on réécrit, en introduisant le coefficient d'amortissement ξ :

$$\ddot{u} + 2\xi\omega_0\dot{u} + \omega_0^2 u = 0 \quad (18.21)$$

La solution s'écrit :

$$u = [A \sin(\omega_D t) + B \cos(\omega_D t)] e^{-\xi\omega_0 t} \quad (18.22)$$

où ω_D , la pseudo-pulsation, est définie par :

$$\omega_D = \omega_0 \sqrt{1 - \xi^2} \quad (18.23)$$

On remarquera que ω_D n'est définie que pour $\xi < 1$, i.e. dans le cas des systèmes sous-critiques ou sous-amortis.

Les constantes d'intégration sont déterminées à l'aide des conditions aux limites sur le déplacement u_0 et la vitesse \dot{u}_0 . Il vient :

$$A = \frac{\dot{u}_0 + u_0 \xi \omega_0}{\omega_0} \quad \text{et} \quad B = u_0 \quad (18.24)$$

Le système amorti oscille à une pulsation ω_D (légèrement) inférieure à la pulsation du système non amorti ω_0 . Si l'amortissement est positif (ce qui n'est parfois pas le cas pour des systèmes instables), l'amplitude du mouvement décroît dans le temps de façon exponentielle (en atteignant une amplitude nulle mais pour un temps infini).

Dans le cas d'un système sur-amorti ($\xi > 1$), alors, en posant $\omega'_D = \sqrt{\xi^2 - 1}$, la solution est du type :

$$u = \left[A e^{\omega'_D t} + B e^{-\omega'_D t} \right] e^{-\xi \omega_0 t} \quad (18.25)$$

avec les constantes d'intégration :

$$A = \frac{\dot{u}_0 + (\omega'_D + \xi \omega_0) u_0}{2\omega'_D} \quad \text{et} \quad B = \frac{-\dot{u}_0 + (\omega'_D - \xi \omega_0) u_0}{2\omega'_D} \quad (18.26)$$

Un système sur-amorti n'est pas un oscillateur...

Dans le cas d'un système critique ($\xi = 1$), alors la solution s'écrit :

$$u = (u_0 + \omega_0 u_0 t + \dot{u}_0 t) e^{-\omega_0 t} \quad (18.27)$$

Ce système n'est pas lui non plus un oscillateur...

18.3.3 Vibrations périodiques forcées

Il s'agit du cas où la sollicitation est périodique. Nous l'écrivons sous la forme :

$$\mathbf{F} = \bar{\mathbf{F}} e^{\alpha t} \quad (18.28)$$

avec $\alpha = \alpha_1 + i \alpha_2 \in \mathbb{C}$.

La solution générale s'écrit :

$$\mathbf{q} = \bar{\mathbf{q}} e^{\alpha t} \quad (18.29)$$

En substituant cette forme de solution dans l'équation de la dynamique, il vient :

$$(\alpha^2 \mathbf{M} + \alpha \mathbf{C} + \mathbf{K}) \bar{\mathbf{q}} = \mathbf{D} \bar{\mathbf{q}} = -\bar{\mathbf{F}} \quad (18.30)$$

qui n'est pas un problème de valeurs propres, mais ce système peut être résolu comme un problème statique, i.e. en inversant la matrice \mathbf{D} . Attention, la solution appartient à \mathbb{C} .

On sépare alors les parties réelle et imaginaire en notant : $e^{\alpha t} = e^{\alpha_1 t} (\cos(\alpha_2 t) + i \sin(\alpha_2 t))$, $\bar{\mathbf{F}} = \bar{\mathbf{F}}_1 + i \bar{\mathbf{F}}_2$ et $\bar{\mathbf{q}} = \bar{\mathbf{q}}_1 + i \bar{\mathbf{q}}_2$.

On obtient alors le système suivant :

$$\begin{bmatrix} (\alpha_1^2 - \alpha_2^2) \mathbf{M} + \alpha_1 \mathbf{C} + \mathbf{K} & -2\alpha_1 \alpha_2 \mathbf{M} - \alpha_2 \mathbf{C} \\ 2\alpha_1 \alpha_2 \mathbf{M} + \alpha_2 \mathbf{C} & (\alpha_1^2 - \alpha_2^2) \mathbf{M} + \alpha_1 \mathbf{C} + \mathbf{K} \end{bmatrix} \begin{pmatrix} \bar{\mathbf{q}}_1 \\ \bar{\mathbf{q}}_2 \end{pmatrix} = - \begin{pmatrix} \bar{\mathbf{F}}_1 \\ \bar{\mathbf{F}}_2 \end{pmatrix} \quad (18.31)$$

dans lequel toutes les quantités sont réelles.

Il est ainsi possible de déterminer la réponse à toute excitation périodique par résolution directe.

Pour une excitation périodique, la réponse après une phase transitoire initiale n'est plus influencée par les conditions initiales. La solution obtenue représente la réponse qui s'établit. Ceci est valable aussi bien pour les problèmes en dynamique des structures que pour les problèmes de conduction de chaleur.

Rappel du cas unidimensionnel L'équation différentielle à résoudre est : $M\ddot{u} + C\dot{u} + Ku = F(t)$, que l'on réécrit :

$$\ddot{u} + 2\xi\omega_0\dot{u} + \omega_0^2 u = F_0/M \cos(\omega t) \quad (18.32)$$

en supposant que $F(t)$ est un chargement mono fréquentiel.

La solution est somme d'une solution particulière, appelée régime permanent ou forcé, et d'une combinaison linéaire de l'équation sans second membre, dit régime transitoire.

On voit alors, de manière intuitive, que :

- La fréquence du régime permanent (ou forcé) est celle de la fréquence d'excitation (qui « force » le système) ;
- La fréquence du régime transitoire est la fréquence propre du système (puisque'il n'y a pas de second membre).

La solution s'écrit :

$$u = \frac{F_0}{M} \frac{\cos(\omega t - \theta)}{\sqrt{\left(1 - \frac{\omega^2}{\omega_0^2}\right)^2 + \left(2\xi \frac{\omega}{\omega_0}\right)^2}} + [A \sin(\omega_D t) + B \cos(\omega_D t)] e^{-\xi\omega_0 t} \quad (18.33)$$

En présence d'amortissement, le régime transitoire disparaît après quelques périodes d'oscillation.

18.3.4 Régimes transitoires

Le lecteur aura sans doute remarqué que dans les méthodes présentées ci-dessus, les conditions initiales du problème ne sont pas prises en compte. Par exemple le comportement sismique des structures ou l'évolution transitoire d'un problème de conduction de chaleur nécessitent de prendre en compte à la fois les conditions initiales et le caractère non périodique des sollicitations.

L'obtention d'une solution à ce genre de problème nécessite soit l'utilisation d'une discrétisation dans le domaine temporel (voir chapitre 17), soit l'utilisation de méthodes adaptées. Dans ce dernier cas, il existe deux approches :

- la méthode de réponse en fréquence ;
- la méthode d'analyse modale.

Nous allons maintenant présenter cette dernière méthode.

Décomposition modale

La méthode de décomposition modale est sans doute l'une des plus importantes et des plus employées. Nous partons toujours de notre équation de la dynamique sous la forme :

$$M\ddot{\mathbf{q}} + C\dot{\mathbf{q}} + K\mathbf{q} = \mathbf{F} \quad (18.34)$$

Nous avons vu qu'en réponse libre ($\mathbf{F} = \{0\}$), la solution s'écrit :

$$\mathbf{q} = \bar{\mathbf{q}} e^{-\alpha t} = \sum_{i=1}^n \bar{\mathbf{q}}_i e^{-\alpha_i t} \quad (18.35)$$

où α_i sont les valeurs propres et $\bar{\mathbf{q}}_i$ les vecteurs propres. Pour la réponse forcée, l'idée consiste à chercher la solution sous la forme d'une combinaison linéaire des modes propres, i.e. sous la forme :

$$\mathbf{q} = \sum_{i=1}^n \bar{\mathbf{q}}_i y_i(t) \quad (18.36)$$

où la quantité $y_i(t)$ représente la contribution de chaque mode. En injectant cette forme de solution dans l'équation de la dynamique (puis en composant à gauche par ${}^T\bar{\mathbf{q}}_i$), on obtient un ensemble d'équations scalaires indépendantes :

$$m_i \ddot{y}_i + c_i \dot{y}_i + k_i y_i + f_i = 0 \quad (18.37)$$

dont les paramètres sont, grâce à l'orthogonalité des modes :

$$\begin{cases} m_i = {}^T\bar{\mathbf{q}}_i \mathbf{M} \bar{\mathbf{q}}_i \\ c_i = {}^T\bar{\mathbf{q}}_i \mathbf{C} \bar{\mathbf{q}}_i \\ k_i = {}^T\bar{\mathbf{q}}_i \mathbf{K} \bar{\mathbf{q}}_i \\ f_i = {}^T\bar{\mathbf{q}}_i \mathbf{F} \end{cases} \quad (18.38)$$

Les équations scalaires se résolvent ensuite par des méthodes élémentaires indépendamment les unes des autres. Le vecteur total est ensuite obtenu par superposition. Toutefois, pour effectuer cette superposition, il n'aura pas échappé au lecteur qu'il faut avoir résolu le problème des valeurs propres. Dans le cas général, le calcul des valeurs et des vecteurs propres complexes est loin d'être facile. La méthode habituelle consiste à déterminer les valeurs propres réelles du problème vu précédemment :

$$\omega^2 \mathbf{M} \bar{\mathbf{q}} = \mathbf{K} \bar{\mathbf{q}} \quad (18.39)$$

On montre que le problème est découplé en y seulement si on a la propriété d'orthogonalité de \mathbf{C} :

$${}^T\bar{\mathbf{q}}_i \mathbf{C} \bar{\mathbf{q}}_i = 0 \quad (18.40)$$

Or ceci n'est pas vrai en général car les vecteurs propres assurent uniquement l'orthogonalité de \mathbf{M} et \mathbf{K} . En revanche, si la matrice d'amortissement \mathbf{C} est une combinaison linéaire des matrices \mathbf{M} et \mathbf{K} , la propriété d'orthogonalité est alors évidemment satisfaite. C'est l'hypothèse de Basile.

Dans la suite on suppose que la propriété d'orthogonalité de \mathbf{C} est satisfaite. L'équation sur ω devient alors :

$$\omega_i^2 \mathbf{M} \bar{\mathbf{q}}_i = \mathbf{K} \bar{\mathbf{q}}_i \quad (18.41)$$

et par suite :

$$\omega_i^2 m_i = k_i \quad (18.42)$$

En supposant que les modes sont normalisés de telle sorte que $m_i = 1$ et en posant $c_i = 2\omega_i^2 c'_i$ (où c'_i correspond au pourcentage d'amortissement par rapport à sa valeur critique), on montre que les équations scalaires se mettent sous forme d'une équation différentielle du second ordre :

$$\ddot{y}_i + 2\omega_i^2 c'_i \dot{y}_i + \omega_i^2 y_i + f_i = 0 \quad (18.43)$$

dont la solution générale est :

$$y_i = \int_0^t f_i e^{-c'_i \omega_i (t-\tau)} \sin \omega_i (t-\tau) d\tau \quad (18.44)$$

Une intégration numérique permet de déterminer une réponse, puis la superposition de ces termes donne la réponse transitoire totale (en principe !). On rappelle que la méthode de décomposition modale nécessiterait la détermination de l'ensemble des valeurs et modes propres, ce qui représenterait des calculs considérables. D'un point de vue pratique, on ne prend en compte qu'un nombre limité de modes étant donné que les réponses à des fréquences élevées sont souvent très amorties et prennent par conséquent des valeurs négligeables. Par ailleurs le problème des hautes fréquences n'est souvent abordé que de manière statistique.

18.3.5 Calcul des modes propres et méthodes de réduction modale

Les méthodes de réduction modale ont pour but d'effectuer un changement de base dans l'étude d'une structure : on souhaite remplacer l'espace vectoriel initial, dont la dimension est égale au nombre de degrés de liberté, par un autre, dont la taille sera inférieure. En d'autres termes, on cherche une base plus optimale afin de diminuer la taille de cet espace vectoriel, tout en s'assurant que ce qui n'est pas pris en compte est bien négligeable. Or, physiquement, on s'aperçoit que les modes propres (et surtout les premiers modes propres) réalisent cet optimal.

Il existe deux principaux types de méthodes :

- les méthodes à interfaces libres (Craig...);
- et les méthodes à interfaces fixes (Craig-Bampton).

Dans ce document, nous ne présenterons que cette dernière, qui, de plus, est particulièrement adaptée au cas de sous-structuration, i.e. lorsque le système considéré est scindé en sous-structures. Mais tout d'abord, commençons par présenter succinctement quelques méthodes de calcul des modes propres, ce qui n'est pas si aisé que cela, et peut s'avérer coûteux selon les méthodes et le nombre de modes calculés.

Quotient de Rayleigh

Définition 77 — Matrice hermitienne. Une matrice hermitienne, ou auto-adjointe, est une matrice \mathbf{A} carrée à éléments complexes telle que cette matrice est égale à la transposée de sa conjuguée, i.e. :

$$\mathbf{A} = \overline{\mathbf{A}}^T \quad (18.45)$$

En particulier, une matrice à éléments réels est hermitienne si et seulement si elle est symétrique. Une matrice hermitienne est orthogonalement diagonalisable et toutes ses valeurs propres sont réelles. Ses sous-espaces propres sont deux à deux orthogonaux.

Définition 78 — Quotient de Rayleigh. Soit \mathbf{A} une matrice hermitienne et \mathbf{x} un vecteur non nul, on appelle quotient de Rayleigh $R(\mathbf{A}, \mathbf{x})$ le scalaire :

$$R(\mathbf{A}, \mathbf{x}) = \frac{{}^T\overline{\mathbf{x}}\mathbf{A}\mathbf{x}}{{}^T\overline{\mathbf{x}}\mathbf{x}}. \quad (18.46)$$

où ${}^T\overline{\mathbf{x}}$ désigne le vecteur adjoint de \mathbf{x} , c'est-à-dire le conjugué du vecteur transposé.

Dans le cas où \mathbf{A} et \mathbf{x} sont à coefficients réels, alors ${}^T\overline{\mathbf{x}}$ se réduit au vecteur transposé, i.e. ${}^T\mathbf{x}$. Le quotient de Rayleigh atteint un minimum λ_{\min} (qui n'est autre que la plus petite valeur propre de \mathbf{A}) lorsque \mathbf{x} est un vecteur propre \mathbf{v}_{\min} associé à cette valeur. De plus, quel que soit le vecteur \mathbf{x} , le quotient de Rayleigh $R(\mathbf{A}, \mathbf{x}) \leq \lambda_{\max}$ (où λ_{\max} est la plus grande valeur propre de \mathbf{A} de vecteur propre associé \mathbf{v}_{\max}) et $R(\mathbf{A}, \mathbf{v}_{\max}) = \lambda_{\max}$. Ainsi, le quotient de Rayleigh, combiné au théorème du minimax de von Neumann, permet de déterminer une à une toutes les valeurs propres d'une matrice.

On peut également l'employer pour calculer une valeur approchée d'une valeur propre à partir d'une approximation d'un vecteur propre. Ces idées forment d'ailleurs la base de l'algorithme d'itération de Rayleigh.

Méthode itérative de Rayleigh

On utilise l'algorithme suivant :

1. choix d'un vecteur initial \mathbf{x}_i ;
2. résolution de $\mathbf{K}\mathbf{x}_{i+1} = \mathbf{M}\mathbf{x}_i$;
3. test de convergence :

- si $|\mathbf{x}_{i+1} - \mathbf{x}_i| < \varepsilon$, alors aller au point 4 ;
 - sinon retourner au point 1 pour choisir un nouveau vecteur ;
4. $\varphi = \mathbf{x}_{i+1}$ et $\omega^2 = \frac{\langle \Phi, \mathbf{K}\Phi \rangle}{\langle \Phi, \mathbf{M}\Phi \rangle}$

Ce processus converge vers le mode propre fondamental. Lorsque l'on veut déterminer le mode le plus proche d'une pulsation donnée ϖ , il suffit de remplacer la matrice de rigidité par $\tilde{\mathbf{K}} = \mathbf{K} - \varpi^2 \mathbf{M}$. Le processus converge alors vers le mode de pulsation $\tilde{\omega}^2 = \omega^2 - \varpi^2$.

Itérations des sous-espaces

La méthode précédente peut être étendue en prenant plusieurs vecteurs initiaux et en se plaçant dans le sous-espace qu'ils définissent. Les pulsations propres doivent alors être calculées à chaque itération en calculant tous les modes propres du système réduit au sous-espace étudié.

Sous-structuration : méthode de Craig et Bampton

Considérons une structure comportant n degrés de liberté et ayant une matrice masse \mathbf{M} et une matrice de rigidité \mathbf{K} . L'utilisation de la méthode de Craig-Bampton impose de décomposer les degrés de liberté de la structure en deux parties :

- les degrés de liberté « frontière » : on considère que ces degrés de liberté sont ceux pouvant potentiellement être chargés au cours du temps et ceux sur lesquels s'appliquent des conditions aux limites (encastrement, appui simple...). Les chargements volumiques (tel le poids) n'influent pas sur la détermination des degrés de liberté frontière, sans quoi cette décomposition n'aurait pas de sens. Ces degrés de liberté sont rangés dans le vecteur \mathbf{q}_L (L comme liaison).

Dans le cas de la sous-structuration, les degrés de liberté frontières correspondent trivialement aux degrés de liberté aux interfaces entre les différentes sous-structures.

- les degrés de liberté « intérieurs » : il s'agit de tous les autres degrés de liberté (un chargement volumique peut éventuellement s'appliquer sur ces degrés de liberté). Ces degrés de liberté constituent le vecteur noté \mathbf{q}_I (I comme intérieur).

La base de réduction se compose de deux types de modes :

- Les modes encastrés : il s'agit des modes propres de la structure calculés en considérant les degrés de liberté frontière encastrés.
- Les modes statiques : ces modes sont obtenus en calculant la déformée statique de la structure lorsqu'un degré de liberté frontière est imposé à 1, tous les autres étant imposés à 0.

Avantages de la méthode :

- elle est facile à programmer
- sa stabilité est connue.
- elle apporte de bons résultats pour des structures de taille raisonnable.
- elle permet d'obtenir les degrés de liberté frontière dans le vecteur réduit ce qui peut s'avérer très utile dans le cas de problèmes de contacts par exemple.

Inconvénients de la méthode :

- cette méthode n'est pas celle qui permet d'obtenir la meilleure réduction du système et peut donc s'avérer coûteuse en temps de calcul.
- sans amortissement structural, la méthode peut diverger au voisinage des fréquences propres de la structure (à cause du gain infini à la résonance sans amortissement).

18.4 Remarques sur l'amortissement

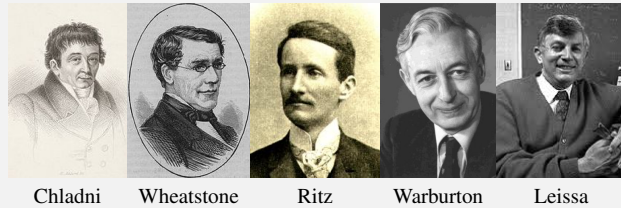
Vous avez peut-être déjà rencontré des cas pour lesquels ont été développés des modèles vibratoires adaptés au problème à traiter. En voici quelques exemples :

- vibration transversale des cordes ;
- vibration longitudinale dans les barres ;
- vibration de torsion dans les barres ;
- vibration de flexion dans les poutres ;
- vibration des membranes ;
- vibration des plaques ;
- propagation des ondes quasi-longitudinales dans les barres ;
- propagation des ondes de flexion dans les poutres...

Histoire

Il ne faut pas négliger ces modèles simplifiés. Ne serait-ce que c'est eux qui ont vu le jour en premier...

C'est en 1787 à Leipzig que Chladni met en évidence expérimentalement la formation de lignes nodales sur une plaque libre avec du sable. Wheatstone et Rayleigh, respectivement en 1833 et 1873, utiliseront des modes de poutre libre pour essayer de comprendre et d'expliquer les figures de Chladni.



Chladni

Wheatstone

Ritz

Warburton

Leissa

En 1909, Ritz, toujours sur ce problème de la plaque libre, utilisera pour la première fois la méthode qui porte son nom. Les premiers résultats concernant la plaque encastree ne viendront qu'en 1931, et sont dus à Sezawa.

En 1939, Igushi développe une méthode pour obtenir certains résultats analytiques, mais les premières synthèses complètes sur les méthodes utilisables pour calculer les fréquences naturelles et les déformées modales de plaques ne viendront qu'en 1954 par Warburton, et 1969 par Leissa.

La méthodologie est toujours la même et se base sur la technique de séparation des variables qui permet de dire que les variables d'espace et de temps peuvent être séparées. On écrira donc un déplacement transversal $w(x, y)$ comme produit d'une fonction dépendant de l'espace $X(x)$ et d'une fonction dépendant du temps $T(t)$: $w(x, t) = X(x)T(t)$. Ainsi, il sera « aisé » de résoudre le problème (en ayant pris en compte les conditions aux limites évidemment).

Généralement, dans un premier temps, lors du développement de ces modèles de cordes, barres, poutres, membranes et plaques, l'amortissement n'est pas pris en compte. Les solutions obtenues pour les réponses libres ne présentent donc pas de décroissance de l'amplitude des mouvements dans le temps. Il est possible d'intégrer cet amortissement de plusieurs façons :

Facteur d'amortissement modal : il s'agit de la manière la plus simple d'introduire l'amortissement en incluant un terme dissipatif correspondant à un modèle d'amortissement visqueux sur la fonction dépendant du temps uniquement. Dans ce cas, la fonction de dépendance temporelle $T(t)$, qui avait pour équation différentielle $\ddot{T}_n(t) + \omega_n^2 T_n(t) = 0$, pour les modes $n \geq 1$, devra désormais répondre à l'équation $\ddot{T}_n(t) + 2\xi_n \omega_n \dot{T}_n(t) + \omega_n^2 T_n(t) = 0$, où ξ_n est le facteur d'amortissement modal ;

Coefficient d'amortissement dans l'équation d'onde : un terme dissipatif est introduit directement dans l'équation des ondes. Celui-ci peut être proportionnel au milieu externe dans lequel se produit le phénomène (par exemple proportionnel à la vitesse de déformation hors plan d'une membrane vibrant dans un fluide, comme l'air), ou proportionnel au milieu interne considéré (par exemple proportionnel à la vitesse de fluctuation des contraintes dans une poutre : modèle de Kelvin-Voigt) ;

Dissipation aux limites : l'amortissement peut également être introduit dans la définition des conditions aux limites, par exemple pour prendre en compte le mode de fixation de la structure. (Par exemple, dans le cas des ondes longitudinales dans une barre, on introduit un ressort et un amortisseur à chaque bout de la barre).

Ces remarques, bien que générales, nécessitent d'être correctement prises en compte pour ces modèles simplifiés.

18.5 Pour aller plus loin : cas des chocs large bande

On s'intéresse au cas des chocs, car il constituent un cas plus compliqué que de la « dynamique lente », sans rien enlever en généralité aux méthodes.

En dynamique transitoire, lorsque le contenu fréquentiel de la sollicitation est large, on montre que les erreurs numériques faites sur chaque longueur d'onde se cumulent, d'où une dégradation de la qualité attendue du résultat plus rapide que prévue. De plus, une erreur sur la périodicité des oscillations due à un trop faible nombre de pas de temps et un déphasage des oscillations à cause de cette accumulation des erreurs numériques au cours des pas de temps peuvent être observées.

Dans de tels cas, il est possible de se placer dans le domaine fréquentiel (à l'aide de la FFT), ce qui conduit à résoudre un problème de vibrations forcées sur une très large bande de fréquences incluant à la fois les basses et les moyennes fréquences pour l'étude des chocs. La solution temporelle est ensuite reconstruite par transformée de Fourier inverse.

Les deux domaines fréquentsiels ayant des propriétés différentes, on recourt à des outils de résolution différents.

Dans le domaine des **basses fréquences**, les phénomènes vibratoires générés par l'excitation ont une longueur d'onde grande face à la structure donc uniquement quelques oscillations sont observables. De plus la structure présente un comportement modal (modes bien distincts les uns des autres). La modélisation est maîtrisée : éléments finis sur base modale, complété si besoin de modes statiques.

Pour les **hautes fréquences**, les longueurs d'ondes sont petites et une centaine d'oscillations est présente sur une dimension de la structure. Il n'est pas approprié de regarder les grandeurs locales, mais plutôt les grandeurs moyennées en espace et en fréquence. On utilise généralement la SEA qui donne un niveau énergétique vibratoire moyen par sous-structure. Cette méthode ne permet pas d'obtenir une solution prédictive car elle requiert la connaissance *a priori* de facteurs de couplages mesurés.

En **moyennes fréquences**, plusieurs dizaines d'oscillations apparaissent sur une dimension de la structure, et la déformée est très sensible aux conditions aux limites et aux paramètres matériaux de la structure. Si un comportement modal est encore visible, les modes sont moins bien séparés (par exemple plusieurs modes présents par Hertz, ces modes étant couplés par l'amortissement) : la méthode des éléments finis est mal adaptée à cause du raffinement de maillage nécessaire, et le calcul de la base modal est également hors de portée. Les méthodes énergétiques quant à elles sont trop globales et ne permettent pas une description satisfaisante de la solution.

Si la structure est divisible en sous-structures homogènes, on peut utiliser la Théorie Variationnelle des Rayons Complexes (TVRC) introduite par Ladevèze en 1996 [51] : les conditions de continuité en déplacements et en efforts aux interfaces entre sous-structures n'ont pas besoin d'être vérifiées *a priori*, mais uniquement au sens faible par une formulation variationnelle.

La TVRC permet l'utilisation d'approximations indépendantes par sous-structure. La solution est supposée bien décrite par la superposition d'un nombre infini de modes locaux, appelés rayons, issus de la vérification des équations d'équilibre dynamique et des relations de comportement par sous-structure. Ces rayons sont à deux échelles : une lente et une rapide. L'échelle rapide est traitée analytiquement (sinon coût numérique élevé), et l'échelle lente numériquement, car elle conduit à un problème à faible nombre d'inconnues.

On profitera de la rapide dispersion des modes moyennes fréquences dans les milieux dispersifs amortis ainsi que de la version large bande de la TVRC développée en 2004 et 2005 [24]

18.5.1 Approches temporelles

Discrétisation spatiale

Une discrétisation par la méthode des éléments finis est mal adaptée aux phénomènes à fort gradient tels que les chocs, car il faut soit un maillage très fin, soit une interpolation par des polynômes de degré élevé, ce qui dans les deux cas augmente considérablement l'effort de calcul.

Étant donné le caractère très localisé des ondes propagatives en dynamique transitoire, la méthode des éléments finis adaptatifs répond au besoin d'enrichir le modèle localement en raffinant le maillage uniquement sur les fronts d'onde de manière contrôlée et automatique.

Ces méthodes recourent à un estimateur d'erreur *a priori* :

- estimateur construit sur les résidus d'équilibre pour l'adaptation de maillage dans le cadre de la propagation d'ondes ;
- estimateur utilisant le lissage des contraintes ;
- estimateur basé sur l'erreur en relation de comportement.

Décomposition de domaine en dynamique transitoire

Le domaine est décomposé en sous-domaines plus petits à calculer. On pourra se servir de la parallélisation du problème. Le problème est donc condensé sur les quantités d'interface entre sous-domaines, ce qui conduit à un problème de taille réduite.

La plupart des méthodes utilisées sont des méthodes sans recouvrement ; elles peuvent être primales, duales ou mixtes. Le problème d'interface est résolu de façon itérative, ce qui évite la construction explicite du complément de Schur¹, mais nécessite un taux de convergence élevé pour être efficace.

Dans la méthode duale, les efforts sont privilégiés : la méthode propose à priori des efforts en équilibre aux interfaces et cherche à écrire la continuité en déplacements. L'inconnue principale, i.e. les inter-efforts entre sous-structures, sont les multiplicateurs de Lagrange aux interfaces.

Discrétisation temporelle

Les méthodes d'intégration directe sont nombreuses et mieux adaptées que les techniques de bases réduites pour les chocs relativement rapides qui mettent en jeu des fréquences élevées.

Notons qu'il existe des méthodes qui s'affranchissent de la discrétisation temporelle et s'appuient sur une méthode asymptotique numérique pour déterminer la réponse transitoire de la structure ; ces méthodes demandent encore à être développées pour les variations temporelles rapides comme les chocs.

Parmi les méthodes d'intégration directes, on utilise classiquement :

- les schémas de Newmark (voir chapitre précédent) pour une intégration d'ordre 2 : les schémas précis au second ordre des différences centrées et de l'accélération moyenne sont privilégiés pour les faibles erreurs d'amplitude et de périodicité qu'ils engendrent.

1. En algèbre linéaire et plus précisément en théorie des matrices, le complément de Schur est défini comme suit. Soit :

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \quad (18.47)$$

une matrice de dimension $(p + q) \times (p + q)$, où les blocs \mathbf{A} , \mathbf{B} , \mathbf{C} , \mathbf{D} sont des matrices de dimensions respectives $p \times p$, $p \times q$, $q \times p$ et $q \times q$, avec \mathbf{D} inversible. Alors, le complément de Schur du bloc \mathbf{D} de la matrice \mathbf{M} est constitué par la matrice de dimension $p \times p$ suivante :

$$\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} \quad (18.48)$$

Lorsque \mathbf{B} est la transposée de \mathbf{C} , la matrice \mathbf{M} est symétrique définie-positive si et seulement si \mathbf{D} et son complément de Schur dans \mathbf{M} le sont.

Le schéma des différences centrées est explicite et adapté pour les chargements de dynamique rapide et les problèmes non linéaires (car la matrice dynamique à inverser est diagonale), mais nécessite de vérifier que le signal ne se propage pas de plus d'un élément pendant un pas de temps (condition de Courant, ou condition CFL dont nous reparlerons au chapitre suivant).

Le schéma de l'accélération moyenne est implicite mais inconditionnellement stable, bien adapté pour les chargement peu rapides.

- la méthode de Galerkin discontinue : Elle autorise les variables du problème déplacement et vitesse à être discontinues en temps. À l'ordre zéro (champs constants sur chaque intervalle de temps), elle permet de s'affranchir des oscillations numériques occasionnées lors du traitement d'un front d'onde. Toutefois, ce schéma dissipe énormément et demande une discrétisation très fine pour bien représenter les irrégularités.
- la TXFEM (Time eXtended FEM) : Elle utilise une base de fonctions de forme en temps enrichie formant une partition de l'unité. Le schéma est équivalent à certaines méthodes de Galerkin discontinues, le nombre de pas de temps inférieur à Newmark, et les oscillations numériques atténuées. Elle est bien adaptée pour le traitement des discontinuités en temps et notamment les chocs. Voir le paragraphe 23.4 pour une courte description.

18.5.2 Approches fréquentielles

Afin de s'affranchir de l'intégration temporelle et des soucis numériques associés, il est possible de réécrire le problème temporelle en un problème fréquentiel (grâce à la FFT).

L'approche fréquentielle est également plus adaptée dans les situations pour lesquelles des paramètres physiques dépendent de la fréquence.

En appliquant la transformée de Fourier à toutes les quantités dépendant du temps, on obtient des quantités qui dépendent de la fréquence. Ce faisant, le problème à résoudre devient un problème de vibration forcée sur une bande de fréquence. Il faut alors calculer des fonctions de réponse en fréquence (FRF) sur une large plage de fréquences.

Cas des basses fréquences

Compte tenu de la grande taille des longueurs d'ondes et du fait que l'on a peu de modes, bien séparés, les méthodes utilisées sont basées sur les éléments finis.

En réécrivant le problème dynamique dans le cas d'une sollicitation harmonique de pulsation ω , il vient (comme nous venons de le présenter avant) :

$$(-\omega^2 \mathbf{M} + i\omega \mathbf{C} + \mathbf{K}) \mathbf{q} = \mathbf{F} \quad (18.49)$$

En utilisant l'hypothèse de Basile² sur l'amortissement, il est possible de calculer les premiers modes propres associés aux plus petites fréquences propre du système. La solution approchée est alors projetée sur les sous-espaces propres associés. On résout alors un système diagonale de petite taille.

Cas des moyennes fréquences

En moyennes fréquences, l'approche précédente (par éléments finis) nécessiterait d'utiliser une grande quantité de polynômes à cause du caractère très oscillant, ou à augmenter le degré d'interpolation. De plus, il faut prendre en compte plus de modes, qui sont de moins en moins bien séparés.

2. Si le seul amortissement entrant en jeu est un amortissement structurel (dissipation interne du matériau pour une structure homogène), il est alors licite de faire l'hypothèse d'un amortissement proportionnel, encore appelé hypothèse de Basile. Dans ce cas \mathbf{C} s'exprime comme combinaison linéaire de \mathbf{M} et \mathbf{K} , et sa projection sur les modes propres est diagonale.

Notons que l'on peut minimiser l'influence du raffinement du maillage en localisant celui-ci uniquement où la dynamique locale le demande. Pour cela, des estimateurs d'erreur *a posteriori* ont été développés pour les structures, mais également pour l'acoustique.

La dispersion numérique est moindre avec les éléments finis stabilisés : tels que les Galerkin Least Squares et Galerkin Gradient Least Squares... qui n'intervienne pas sur la forme variationnelle mais sur les matrices issues de celle-ci.

On peut également essayer de diminuer la taille de la base modale à prendre en compte en ne retenant que les modes propres qui maximisent l'opérateur d'excitabilité ; mais pour cela il faut d'abord calculer la base complète...

On peut également utiliser un autre espace de projection que les modes propres classiques : par exemple les premiers modes d'un opérateur d'énergie relatif à une bande de fréquence (Soize 98 [82]). Cette approche peut être couplée à la théorie des structures floues (Soize 86 [81]) pour prendre en compte la complexité de la structure de manière probabiliste.

On peut également sous-structurer le domaine. Dans la Component Modal Synthesis, les modes propres de chaque sous-structures servent de base pour la solution de la structure entière.

Une « deuxième » approche consiste à utiliser des éléments enrichis (voir encore une fois le chapitre 23). Ces éléments sont développés pour pouvoir prendre en compte le caractère oscillant de la solution en enrichissant les fonctions de base utilisées afin de pouvoir mieux reproduire la solution.

Les éléments finis hiérarchiques, issus des méthodes p (voir le paragraphe 13.1 : l'augmentation du degré polynomial des fonctions de forme peut se voir comme une substitution à un raffinement de maillage, mais les estimateurs d'erreur des méthodes p sont meilleurs que ceux des méthodes h), permettent une réutilisation à l'ordre $p + 1$ des matrices de masse et de raideur élémentaires issues de l'ordre p .

Les éléments finis multi-échelles (voir paragraphe 13.4) où la solution est recherchée comme somme d'une composante calculable à l'échelle grossière en espace et d'une composante non calculable associée à l'échelle fine.

La Méthode de partition de l'unité (PUM) (voir paragraphe 23.3) utilise un recouvrement du domaine initial en un ensemble de maillages, chacun étant enrichi et vérifiant la partition de l'unité. La méthode des éléments finis associée à la méthode de partition de l'unité donne naissance à deux grandes familles d'approches : la G-FEM (Generalized FEM) et la X-FEM (eXtended FEM), voir paragraphe 23.4. Si les fonctions d'enrichissement ne sont pas activées, on se retrouve avec la méthode des éléments finis classique.

la Méthode d'enrichissement discontinu (Discontinuous Enrichment Method) est une méthode de Galerkin éléments finis discontinue avec multiplicateurs de Lagrange dédiée aux application de forts gradients ou oscillations rapides.

La méthode des éléments frontières ou BEM (voir paragraphe 23.1) est encore une « troisième » approche. Seuls les bords sont maillés afin de réduire le nombre de degrés de liberté. La formulation intégrale de la frontière établit un lien entre les champs intérieurs et les quantités sur les bords. La matrice obtenue est petite, mais pleine et non symétrique.

Les méthodes sans maillage (voir paragraphe 23.2) : Dans la EFGM (Element Free Galerkin Method ou méthode de Galerkin sans maillage), on n'a plus qu'un nuage de points sans connectivité entre eux. On utilise des fonctions de formes construites selon la méthode des moindres carrés mobiles et formant une partition de l'unité, qui peuvent être polynomiales ou sinusoïdales. Néanmoins, comme elle est basée sur une discrétisation nodale, elle nécessite également un grand nombre de nœuds aux fréquences plus élevées.

Les méthodes des éléments discontinus : Pour des géométries simples (utilisation en construction navale), on utilise des solutions analytiques ou semi-analytiques sur des sous-domaines simples (poutres, plaques rectangulaires...) pour construire la structure complète. Lorsque cette méthode est applicable elle donne de bon résultats aussi bien en basses fréquences, en moyennes fréquences et en hautes fréquences.

Les méthodes de Trefftz : Elles utilisent des fonctions de base définies sur tout le domaine de la sous-structure considérée et vérifiant exactement l'équation dynamique et la loi de comportement : la solution est représentée par la superposition de ces fonctions ; mais il faut encore vérifier les conditions aux limites et de transmission. Les matrices sont de petite taille mais très mal conditionnées.

Les T-éléments lient la démarche Trefftz et la méthode des éléments finis : Trefftz au sein de chaque élément..

Pour les problèmes de vibroacoustique, la WBT (Wave Based Technique) a été développée : la structure n'est pas discrétisée comme pour les T-éléments mais est décomposée en éléments de grande taille par rapport à la dimension de la structure. Les fonctions de base particulières utilisées améliorent le conditionnement des matrices, ce qui conduit à des matrices de petite taille pleines et non symétriques. L'intégration sur les bords coûte cher en moyennes fréquences et le conditionnement de la matrice en hautes fréquences se dégrade du fait de la discrétisation des amplitudes (seules certaines directions de propagation sont prises en compte).

Cas de hautes fréquences

Comme dit précédemment, on ne représente pas la solution localement mais on ne s'intéresse qu'à des grandeurs moyennées.

La SEA (Statistical Energy Analysis) est la méthode de référence pour les hautes fréquences. La structure est découpée en sous-structures. Ensuite des regroupements de modes sont construits tels que statistiquement le niveau de chacun des groupes de modes soit semblable : la méthode repose donc sur l'hypothèse d'une forte densité modale dans la bande de fréquence étudiée. Chaque groupe de mode est associé à un degré de liberté : le problème à résoudre issu de l'équilibre énergétique est par conséquent à faible nombre de degrés de liberté. Cet équilibre se traduit par un bilan de puissance dans lequel la puissance injectée à une sous-structure par des forces extérieures, aléatoires et stationnaires sur de larges bandes de fréquences, est égale à la somme de la puissance dissipée dans cette sous-structure par amortissement et la puissance transmise à l'ensemble des sous-structures voisines avec lesquelles elle est connectée, appelée puissance de couplage. L'hypothèse forte de la SEA concerne cette puissance de couplage entre deux sous-structures qui est supposée proportionnelle à la différence de leurs énergies par mode, le facteur de proportionnalité étant le coefficient de couplage.

La SEA est parfaitement adaptée aux hautes fréquences, mais trop globale et trop imprécise pour décrire finement le comportement en moyennes fréquences.

De plus les coefficients de couplages ne sont connus explicitement *a priori* que pour des géométries très particulières et nécessitent donc en général de recourir à des expériences, ce qui fait de la SEA une méthode non prédictive.

Des stratégies de calcul de ces coefficients de couplage existent, mais pour certains régimes d'excitation la notion même de coefficient de couplage n'est plus réaliste.

La méthode de diffusion d'énergie : Elle apporte un effet local à la SEA en décrivant de manière continue les variables énergétiques. Elle a été appliquée à des cas simples et l'analogie avec la thermique n'est pas démontrée pour des sollicitations et des géométries quelconques.

L'analyse ondulatoire de l'énergie : Elle généralise la SEA en ce qu'elle considère le champ d'ondes non plus diffus mais directionnel en introduisant un champ d'ondes aléatoires propagatives dans les sous-structures et des coefficients de couplages qui varient selon les angles d'incidence aux interfaces.

Les Méthodes Énergétiques Simplifiées ou MES : Elles se proposent de pallier les insuffisances de la méthode de diffusion de l'énergie en donnant une représentation locale des phénomènes. Le bilan de puissance est écrit aussi bien à l'intérieur des sous-structures que de leurs couplages. La connaissance des coefficients de couplages *a priori* demeure un problème.

D'autres méthodes existent encore : développements asymptotiques, méthode de l'enveloppe, méthode des chemins structuraux.

La méthode des rayons : consiste à suivre les rayons vibratoires le long de leur parcours par l'étude de leur propagation, réflexion et transmission entre sous-structures par les lois de Snell-Descartes³ de l'optique géométrique jusqu'à l'amortissement des ondes. La RTM permet de connaître la direction privilégiée de transfert et de répartition spatiale de l'énergie, mais à un coût numérique très élevé. De plus, les coefficients de transmissions ne sont pas connus à priori...

18.5.3 Remarques

Dans l'approche fréquentielle, il faut déterminer les fréquences déterminant la jonction BF/MF et MF/HF.

La fréquence BF/MF agit sur le coût de calcul : elle doit être la plus grande possible, mais telle qu'à partir de cette fréquence, les modes de la structure deviennent locaux, tout en conservant la séparation des modes (en pratique entre 300 et 600 Hz).

Le choix de la fréquence MF/HF joue sur la qualité de la vitesse calculée et donc sur l'énergie cinétique qui sert pour restaurer la réponse temporelle. Cette fréquence doit être au moins égale à $1/T$ où T est la durée du choc d'entrée (si $T = 1$ ms, alors $MF/HF = 2000$ Hz mini).

la méthode des éléments finis utilisée en basses fréquences doit utiliser les n premiers modes pour la base réduite avec n tel que la fréquence de ce mode soit de $2 \times$ la fréquence BF/MF. On utilisera également la règle classique d'au moins 10 éléments linéaires par longueur d'onde pour le maillage.

La FFT requiert par ailleurs que le chargement soit périodique. Le temps correspondant à cette période, T_0 , doit être choisi judicieusement. Dans le cas d'un choc, on a donc $T_0 > T$, mais il faut également le choisir tel que la réponse transitoire de la structure s'éteigne avant la fin de cet intervalle de temps. De plus, ce temps influe sur l'échantillonnage fréquentiel du calcul de la FRF. Les pulsations pour lesquelles la FRF est calculée doivent être telles que $\omega_n = 2\pi n f_0$ avec $f_0 = 1/T_0$. Il faut également que le nombre de pas de fréquences N soit une puissance de 2 pour utiliser la FFT et son efficacité, et N soit être tel que $N/T_0 \geq 2f_{\max}$ avec f_{\max} la fréquence maximale contenue dans le signal.

Le choix judicieux de T_0 influe directement sur la reconstruction temporelle de la réponse. Pour les structures peu amorties ou pour des chargements longs, T_0 est grand, et donc la FFT coûteuse. Dans ces cas, des méthodes ont été développées : les fonctions de Green ; la Implicit Fourier Transform ; et l'amortissement artificiel.

3. Selon Huygens, Snell découvrit le premier les lois de la réfraction en 1621. Il semble par ailleurs qu'on lui doive également, avant Neper, l'écriture actuelle des nombres décimaux, en France tout au moins : *e*, *dcm*... distinguant, séparées par une virgule, les parties entière *e* et décimale (*d* = dixièmes, *c* = centièmes, *m* = millièmes,...).

Chapitre 19

L'acoustique

Résumé — En complément avec ce qui a été vu aux chapitres précédents 17 et 18 sur les problèmes non stationnaires et les ondes, nous allons nous focaliser dans ce chapitre sur l'acoustique, et plus particulièrement sur le calcul de problèmes pour lesquels les fréquences restent inférieures à quelques milliers de Hz. Nous en profiterons pour effectuer une présentation pratique de l'acoustique et des solutions qui peuvent être mises en œuvre physiquement.

Nous avons déjà présenté la problématique de l'acoustique à plusieurs reprises au long de ce document : au paragraphe 6.4.2, l'équation des ondes a été donnée dans le cas de l'acoustique sous forme d'équation différentielle (6.12) dont l'inconnue est la pression acoustique ; des remarques plus « physiques » concernant l'acoustique ont été faites au paragraphe 6.7, notamment concernant les échelles d'énergies mises en œuvre, ainsi que sur les aspects solide et aérien ; mais c'est surtout au paragraphe 8.4 qu'ont été présentées l'équation d'Helmholtz (8.59) et sa formulation faible. Quant à la formulation éléments finis, elle a été abordée au chapitre 18, où ont été décrit les systèmes matriciels à traiter selon les cas (réponse libre, périodique ou transitoire, amortie ou non).

Toutefois, nous souhaitons apporter des éléments supplémentaires sur le sujet :

- au paragraphe 19.1, nous proposerons une présentation de l'acoustique « sur le terrain » : nous exposerons brièvement l'acoustique à partir des problématiques qui se posent physiquement, et nous présenterons quelques solutions typiquement utilisées pour résoudre ces problèmes d'acoustique. Nous verrons d'ailleurs que les problèmes rencontrés en hautes fréquences sont généralement aisément résolus, et qu'il convient donc de se concentrer sur les fréquences inférieures à quelques kHz.
- au paragraphe suivant 19.2 nous reviendrons plus en détails sur la mise en œuvre pratique d'un calcul acoustique par éléments finis.

La motivation principale vient de ce que, dans le paragraphe 18.5, nous avons indiqué que la méthode des éléments finis devait se cantonner aux basses fréquences, celles-ci allant jusqu'à 600 Hz environ. Il est vrai qu'au-delà, de nombreuses autres méthodes existent... et nous avons profité de ce paragraphe pour les présenter. Néanmoins, l'augmentation rapide des capacités des ordinateurs, fait qu'il est tout à fait raisonnable aujourd'hui de traiter des cas allant jusqu'à quelques milliers de Hz, disons 3000 Hz pour fixer les idées, à l'aide de la méthode des éléments finis. Cela permet de traiter la plupart des cas pratiques qui se posent à nous, puisque nous aurons vu auparavant qu'il est souvent inutile de monter plus haut en fréquence.

- Des illustrations sur quelques cas types seront faites au paragraphe 19.3.

19.1 Introduction à l'acoustique physique

Nous nous proposons d'étudier comment une nuisance vibro-acoustique se propage depuis son émission jusqu'à sa réception. La source pourra être mécanique, aéro- ou hydrodynamique... et pourra couvrir un très large spectre de fréquences, depuis les plus basses (excitation vibratoire, i.e. inférieure à environ 500 Hz), jusqu'aux plus hautes (excitation acoustique, depuis 500 jusqu'à environ 8000 Hz). La propagation nécessitera de connaître le ou les modes de transmission (solidien, aérien) ainsi que l'environnement de transmission (champ ouvert ou fermé). La réception quant à elle sera liée à la perception, qui elle-même sera décrite de manière normative (niveau, puissance) ou sensorielle (psycho-acoustique...).

19.1.1 Émission

Face à un problème vibro-acoustique, on est amené à considérer les questions suivantes :

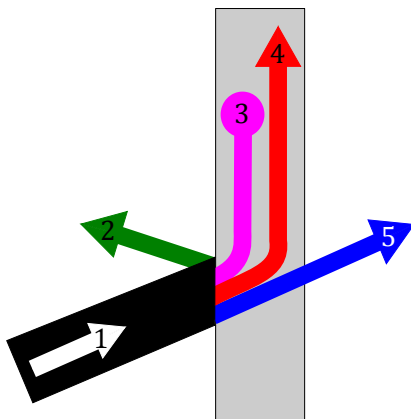
- Localisation : où est-ce que ça fait du bruit ? Il s'agit de trouver, dans un ensemble qui peut être très complexe (par exemple un véhicule, un compresseur...), où sont les sources.
- Identification/Séparation : là où ça fait du bruit, qu'est-ce qui fait du bruit exactement ? De manière plus détaillée que la localisation, il peut être nécessaire, pour les sources identifiées, de déterminer la contribution de certaines de leurs parties ou sous-ensembles : selon la modélisation souhaitée, on pourra se contenter de dire que la source vibro-acoustique dans un véhicule est le moteur, où alors souhaiter s'intéresser plus finement aux injecteurs, au carter, à la boîte de vitesse...
- Enregistrement/Caractérisation : quel bruit est émis par chaque source ? Il s'agit de réaliser des enregistrements permettant de caractériser la source vibro-acoustique considérée. Par exemple, un vibromètre laser permet de récupérer les vibrations (accélérations, vitesses), sans contact. De là, il est possible de re-synthétiser le son d'une seule pièce parmi un ensemble de pièces en fonctionnement. Les enregistrements permettent de caractériser la source en terme de spectre, puissance, directivité... et fournissent également des données utiles pour la restitution de résultats (par exemple données audio binaurales).

La caractérisation des sources peut également déboucher sur des lois phénoménologiques, qui relient les caractéristiques vibro-acoustiques à certains paramètres tels que par exemple : le rapport de boîte, le régime moteur, le régime ventilateur [33, 39, 40, 53, 56, 74, 75, 83]...

Notons enfin que lorsque l'on parle d'acoustique, la propagation finale se fait forcément dans l'air puisque c'est celle qui arrive jusqu'aux oreilles...

19.1.2 Transmission

La figure 19.1 donne la répartition schématique des puissances (ou les différents types d'ondes à considérer).



Une onde incidente (1) arrivant sur un obstacle se trouve en partie réfléchi par celui-ci (2), alors qu'une partie est directement transmise (5). Toutefois, cela ne représente pas la totalité de l'énergie incidente, car il reste encore une fraction qui est *absorbée ou dissipée par l'obstacle* (3), et une dernière qui se propage au sein de l'obstacle (4) et qui peut ressurgir plus loin dans la structure considérée.

Pour les matériaux acoustiques, la réflexion et la transmission seront déduites de mesures en tube d'impédance ou en cabine alpha, la capacité d'isolation sera caractérisée en utilisant la petite cabine, alors que

FIGURE 19.1: Répartition schématique des puissances

la capacité de rayonnement sera appréhendée par le banc RTC III.

L'isolement est une donnée globale qui traduit la capacité d'un ensemble à « faire barrière au bruit ». Il s'agit tout simplement de la différence entre ce qui attaque d'un côté et ce que l'on récupère de l'autre. Un isolement donné ne donne donc pas d'information sur la manière dont il est obtenu : réflexion, amortissement, absorption, ou plus probablement par une combinaison plus ou moins complexe (et inconnue) de ces trois phénomènes.

Modes de propagation

Comme déjà mentionné au paragraphe 6.7, il peut être nécessaire de considérer un ou plusieurs des modes de propagation suivants :

- **solidien/solidien** : la source excite mécaniquement la structure, et la vibration se propage dans la structure ;
- **aérien/aérien** : la source émet une vibration (un bruit) qui se propage dans l'air (en général). il s'agit de la propagation d'une onde acoustique de sa source jusqu'au récepteur, dans l'air ;
- **solidien/aérien** : il s'agit du cas du rayonnement. La source excite mécaniquement une structure, et celle-ci ré-émet une onde dans l'air.
- **aérien/solidien** : dans le cas de sources acoustiques très puissantes, l'excitation acoustique se propageant dans l'air peut arriver à faire vibrer une structure.

Dans les deux premiers cas, on doit résoudre un problème de propagation dans un milieu (structure ou air). La sollicitation dépend du temps, donc on peut appliquer ce qui a été vu au chapitre 17 ; mais comme elle est généralement périodique, et que l'on s'intéresse au régime stationnaire, on peut alors utiliser ce qui a été vu au chapitre 18 sur les ondes.

Dans les deux derniers cas, il s'agit d'un calcul où il faut prendre en compte le couplage fluide-structure. La source étant périodique, on utilise encore ce qui a été vu au chapitre 18, aussi bien dans la structure que dans l'air. Évidemment, ce sont les techniques du chapitre 17 qui s'appliquent si l'on s'intéresse au régime transitoire.

Dans la « réalité », tous ces cas existent bien :

- **solidien/solidien** : tout moteur, même monté sur des silenbloks qui filtrent l'excitation, génère dans les supports, puis dans toute la structure porteuse, des vibrations. On est donc bien face à la propagation de vibrations au sein de solides ;
- **aérien/aérien** : si l'on n'entre pas dans le détail de son fonctionnement mécanique, un haut-parleur est une source aérienne qui génère une onde acoustique qui se propage dans un volume d'air contenu, par exemple, dans une salle. Si l'on s'intéresse au niveau de pression acoustique en un point de la salle, on a une modélisation où seul le volume intérieur de la salle est nécessaire (et les bonnes conditions aux limites, mais nous y reviendrons) ;
- **solidien/aérien** : dans le cas de structures en treillis portant des panneaux, ce qui est le cas pour la conception de cars et bus, le moteur, situé à l'arrière, excite la structure de manière solidienne... et des vibrations se propagent dans tout le bus, où elles trouvent régulièrement, et jusqu'à l'avant, des panneaux qui vont se mettre à rayonner, i.e. à transformer cette vibration mécanique en bruit se propageant dans l'air jusqu'aux oreilles du conducteur et des passagers ;
- **aérien/solidien** : dans le cas de structures légères, par exemple pour des véhicules sans permis, le moteur excite la structure non seulement de manière solidienne, mais également de manière aérienne. Le bruit généré par le moteur est tel, au sein du compartiment moteur, que cette sollicitation aérienne est capable de faire vibrer le tablier de séparation moteur/habitacle, qui dans ce cas précis est généralement très peu isolant (typiquement en ABS thermoformé avec une épaisseur avant transformation d'environ 3 mm, donc une épaisseur, dans les endroits les plus déformés, de l'ordre du millimètre). Il suffit pour le mettre en évidence de remplacer

le moteur par un haut parleur découplé du châssis et générant une pression acoustique équivalente à celle du moteur.

Tous ces cas existent donc bien, et peuvent même se superposer. Il convient donc d'être prudent dans le choix de la modélisation la plus adaptée au problème étudié.

Fonctionnement des solutions acoustiques

Il en va de même dans l'utilisation de matériaux ou solutions acoustiques. Plusieurs phénomènes peuvent intervenir et se superposer :

- absorption : il s'agit du cas où l'énergie acoustique est dissipée dans ou par le matériau ou la solution. Pour les matériaux poreux ou fibreux, des modèles de fluides équivalents existent. Ils sont modélisés par des paramètres tels que la porosité, la tortuosité... dans le cas de résonateurs ou de foils absorbers, ce sont les caractéristiques du matériau constituant ceux-ci ainsi que leur géométrie qui déterminent leur performance d'absorption ;
- isolation : il s'agit du cas où l'on fait barrière au bruit. Les notions d'étanchéité et de masses sont prépondérantes. On commencera donc par éviter au maximum toutes les fuites acoustiques (tous les trous). Ensuite, on pourra utiliser le fait que l'isolation est liée à la masse : on gagne 6 dB par doublement de masse (en fait $20 \log_{10} 2$). Enfin, d'autres systèmes peuvent être utilisés. Parmi ceux-ci, le système « masse-ressort » constitue l'une des plus anciennes solutions pour améliorer l'isolation acoustique. Il s'agit de réaliser un système rudimentaire de double parois découplées. Lorsque la première paroi est attaquée par le bruit et les vibrations, la seconde réagit de manière découplée. On peut choisir le ressort et la masse pour « filtrer » les fréquences indésirables. Ce système présente par contre une fréquence de coupure pour laquelle l'isolation n'est pas améliorée, mais au contraire le phénomène est amplifié.
- amortissement : il s'agit du cas où l'énergie vibratoire est dissipée dans un matériau choisi pour ses propriétés d'amortissement visqueux, ou par son mode de fixation (collage, montage en contrainte...).

L'ajout d'une pièce d'insonorisation, par exemple réalisée par thermocompression de matières fibreuses, sur une structure a pour effet d'augmenter l'absorption. Toutefois, cette pièce possède une masse, même faible, qui ajoute de l'isolation. De plus, son mode de fixation peut apporter de l'amortissement à la structure étudiée. Toutefois, si la pièce est montée sur une tôle, sa masse étant vraiment faible face à celle de la tôle, il n'est pas faux de ne pas prendre en compte son apport sur l'isolation. Si de plus elle est juste maintenue mécaniquement, il est également possible de négliger son influence en terme d'amortissement.

Compléments sur l'absorption

Revenons quelques instants sur l'absorption, et plus particulièrement sur les performances d'absorption des solutions existantes, qui sont illustrées sur la figure 19.2.

On voit sur la figure 19.2 que l'on peut distinguer deux types de comportements :

- un comportement d'absorption centré sur une fréquence : il s'agit des panneaux fléchissants, des résonateurs de Helmholtz et des panneaux rigides perforés. L'avantage de ces systèmes est qu'ils permettent d'attaquer des fréquences relativement basses. Leur inconvénient est que face à une source large bande, ils ne traitent qu'une infime partie du problème ;
- et un comportement d'absorption large bande : il s'agit des matériaux poreux et fibreux. Bien que n'apportant qu'une absorption très limitée aux basses fréquences, ils permettent de bien traiter le spectre en hautes fréquences et constituent donc une solution simple et aisée à mettre en œuvre. La figure 19.3 montre le rôle de la porosité dans l'absorption, en augmentant le parcours moyen parcouru par l'onde acoustique. La dernière photo est une microscopie d'un matériau fibreux réel (on remarquera qu'il y a « beaucoup d'air » et peu de fibres finalement).

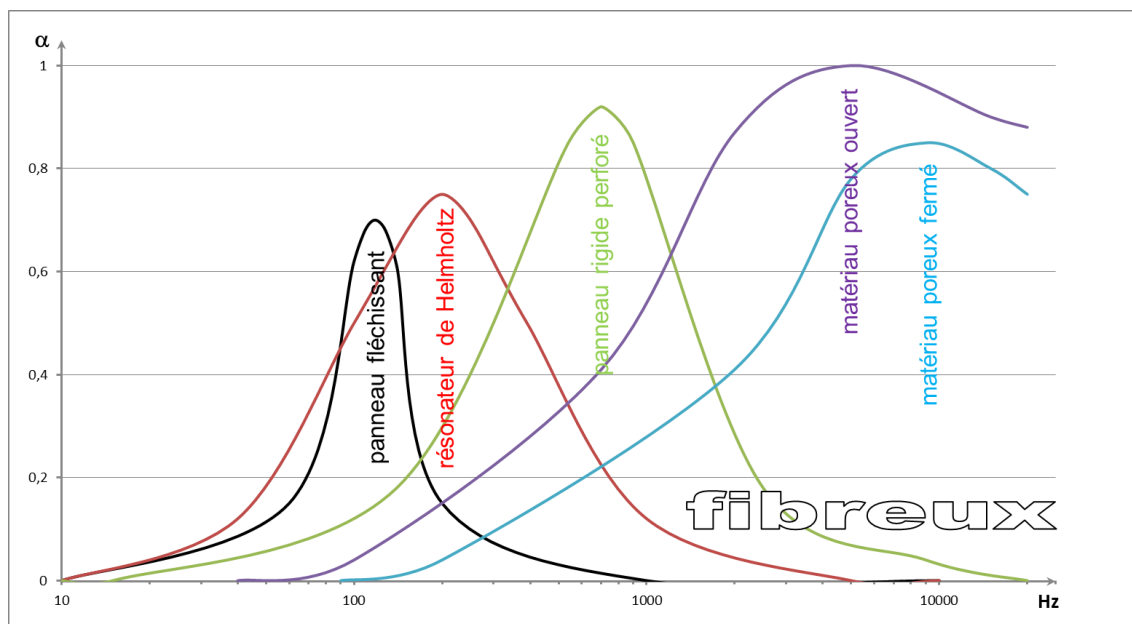


FIGURE 19.2: Coefficient d'absorption pour différents types d'absorbants

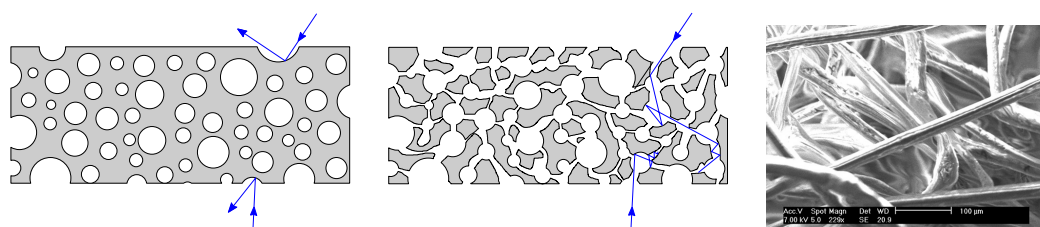


FIGURE 19.3: Rôle de la porosité

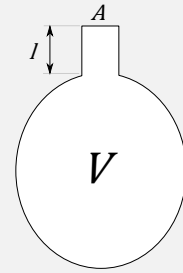
Tout cela justifie la remarque que nous faisons en introduction : la partie difficile à traiter dans les problèmes d'acoustique ne concerne que les basses fréquences où il n'existe que très peu de solutions, surtout simples de mise en œuvre. Les hautes fréquences (supérieures à 2 ou 3000 Hz) peuvent être facilement traitées (poreux, fibreux). De plus, l'oreille humaine, bien que percevant les fréquences de 20 à 20000 Hz, est surtout sensible dans la bande de fréquences allant de 1 à 4 kHz. Dès lors, bien qu'une étude acoustique doive prendre en compte la totalité du spectre (de la source et audible), du point de vue calculatoire (et du point de vue de la conception des solutions), c'est au-dessous de 3000 Hz que le travail devra se concentrer. Ainsi donc, un traitement par la méthode des éléments finis est tout à fait envisageable pour faire face à la majorité des problèmes : nous entendons par là, avec un « coût » de calcul raisonnable.

Histoire

Complément sur le résonateur de Helmholtz.

Les résonateurs de Helmholtz sont une solution très ancienne, puisqu'on utilisait déjà des vases dans les théâtres antiques pour la correction acoustique. Le nom provient d'un dispositif créé dans les années 1850 par Hermann von Helmholtz afin de déterminer la hauteur des différents tons.

Les applications des résonateurs de Helmholtz sont nombreuses. En *acoustique des salles*, ils permettent d'atténuer les fréquences médiums. Le principe du résonateur de Helmholtz est le principe de base permettant de faire des enceintes acoustiques utilisant la technique de l'*enceinte bass-reflex*. L'enceinte contenant le haut-parleur est ouverte par un tube où l'air se déplace et crée donc un résonateur auxiliaire entrant en résonance avec le haut-parleur actif principal permettant d'abaisser la fréquence de coupure basse de l'enceinte. En *aéronautique*, pour limiter le bruit des réacteurs civils, les nacelles et entrées d'air sont constituées de panneaux sandwich nid d'abeilles recouverts d'une tôle perforée aluminium ou composite. Le résonateur est ainsi constitué par la bouteille (cellule du nid d'abeilles) et le col (perçage de la tôle). En *automobile*, la résonance de Helmholtz peut être utilisée pour améliorer le remplissage en air des moteurs à combustion interne...



Résonateur de Helmholtz

Le résonateur est assimilé à une cavité fermée de volume V qui communique avec l'extérieur par l'intermédiaire d'un col de longueur l et de section A . On fait par ailleurs l'hypothèse que les dimensions du résonateur sont petites devant la longueur des ondes acoustiques considérées, que l'air se comporte comme un gaz parfait et que l'on peut négliger les effets thermiques ainsi que les effets dissipatifs. Alors la fréquence du résonateur est :

$$f_0 = \frac{c}{2\pi} \sqrt{\frac{A}{Vl}} \quad (19.1)$$

où c est la vitesse du son. Ce système se comporte comme un système masse ressort : on se retrouve face aux oscillations libres de l'inertance pneumatique $\rho A/l$ de la colonne d'air dans le col sur la raideur pneumatique $\gamma p/V$ de l'air contenu dans la cavité V . Il ne fait pas intervenir la propagation des ondes. La résonance de Helmholtz à la fréquence f_0 doit donc être distinguée des modes propres acoustiques de la cavité, qui sont solutions de l'équation d'Helmholtz, et se situent à des fréquences beaucoup plus élevées que f_0 .

Le coefficient ou facteur d'absorption acoustique est le rapport de l'intensité acoustique qui n'est pas réfléchi (absorbée, propagée et transmise) sur l'intensité acoustique incidente. On définit le coefficient d'absorption de Sabine α_s , ou valeur moyenne du coefficient d'absorption, par :

$$\alpha_s = \frac{\sum_{i=1}^n \alpha_i S_i}{S_{totale}} \quad (19.2)$$

où la surface totale S_{totale} est composée de n surfaces S_i de coefficient d'absorption α_i .

L'aire d'absorption de chaque surface S_i , quant à elle, est définie par :

$$A_i = \alpha_i S_i \quad (19.3)$$

L'aire d'absorption équivalente A est calculée par la formule de Sabine datant des années 1898 :

$$A = 4mV + \sum_{i=1}^n S_i \alpha_i = 4mV + \sum_{i=1}^N A_i \quad (19.4)$$

où m est l'amortissement du milieu (en général l'air) et V le volume de la salle (ou du volume considéré). Cette formule n'est valable que pour des valeurs de α sensiblement inférieures à 1. Pour $\alpha \geq 0,3$ on utilisera la formule d'Eyring précisée dans les années 1920, et valable pour toutes les valeurs de α :

$$A = 4mV - \sum_{i=1}^N S_i \ln(1 - \alpha_i) = \frac{S_{totale} \bar{\alpha}}{1 - \bar{\alpha}} \quad (19.5)$$

où $\bar{\alpha}$ est le coefficient moyen d'absorption qui peut être pris égale à α_s .

Le terme d'amortissement $4mV$ est souvent négligé, ce qui est particulièrement licite pour les salles de petite taille.

Diffraction et réflexion

La diffraction est un autre phénomène avec lequel il faut compter dans la propagation des ondes. Il est illustré sur plusieurs cas à la figure 19.4.

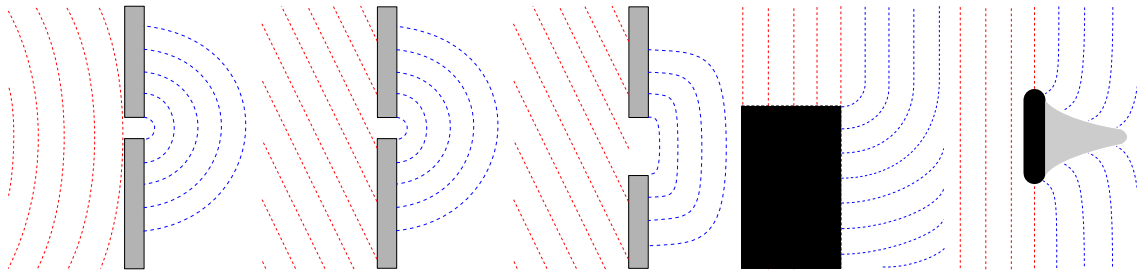


FIGURE 19.4: Phénomène de diffraction

Les figures 19.4a–19.4c représentent le phénomène de diffraction au travers d’une ouverture. L’onde après l’obstacle, i.e. l’onde diffractée, a la même fréquence, la même longueur d’onde et la même vitesse que l’onde incidente. Par contre, sa direction, son amplitude et sa forme sont modifiées (l’amplitude de l’onde diffractée est inférieure à celle de l’onde incidente).

C’est ce phénomène qui fait qu’au coin d’une rue, on entend la personne venant perpendiculairement à nous avant de la voir, comme montré sur la figure 19.4d.

Rappelons que deux ondes, diffractées ou non (mais ça a plus de sens ici en parlant d’ondes diffractées) peuvent interférer. Le phénomène analogue se produit également lorsqu’un obstacle est inséré au sein d’un fluide, comme le représente la figure 19.4e. Les ondes doivent « éviter » l’obstacle, ce qui modifie notamment leur direction sur les bords de l’obstacle. Il se crée une zone d’ombre en arrière de l’obstacle (schématisée en gris). De plus, selon la forme de l’objet, les ondes diffractées peuvent générer, plus en aval de l’obstacle, des figures d’interférence.

Le phénomène de réflexion sur un obstacle est illustré sur la figure 19.5. La figure 19.5a propose

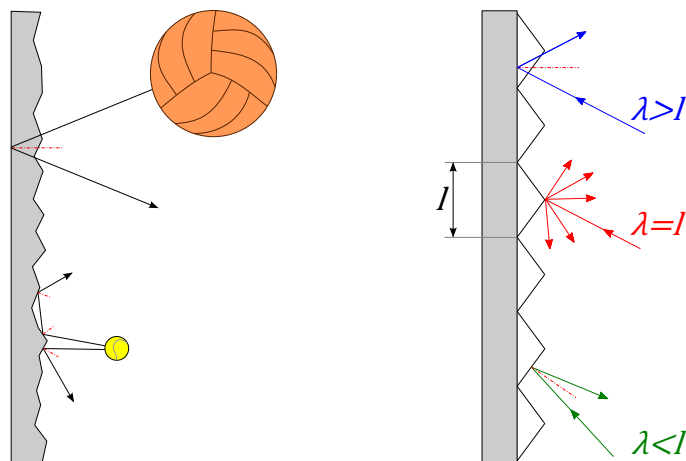


FIGURE 19.5: Phénomène de réflexion

une analogie avec des balles, de ce qui est décrit avec des ondes à la figure 19.5b, i.e. le cas de la réflexion contre une surface irrégulière.

Si la balle est grosse face à la taille des irrégularités, i.e. si la longueur d’onde est plus grande que « celle des irrégularités » (ici on a un mur « en accordéon » ayant un motif répétitif de longueur l), alors le trajet est insensible aux irrégularités de la surface : la normale prise en compte est celle de la surface sans irrégularités. Si $\lambda < l$, alors on peut encore utiliser les lois de l’optique géométrique, mais on voit qu’il est alors nécessaire de disposer d’une discrétisation suffisamment fine pour décrire ces irrégularités. Lorsque la dimension de la balle, i.e. la longueur d’onde de

l'onde incidente, s'approche de la longueur caractéristique des irrégularités, alors il convient d'être prudent : si $\lambda = l$, on a des réflexions spéculaires multiples (on parle aussi de réflexion diffuse) qu'il est souvent difficile de prendre en compte, sauf par des formulations particulières. Souvent, ce cas n'est clairement pas pris en compte.

Environnement ouvert ou fermé

La propagation d'ondes acoustiques se traite généralement dans deux cas distincts :

- celui d'un environnement ouvert, ou champ libre : l'onde peut se propager « à l'infini » ;
- ou celui d'un environnement fermé : l'onde est confinée dans un volume donné dont elle ne peut sortir.

On rappelle par ailleurs qu'une onde voit son intensité décroître avec la distance parcourue. L'intensité (vectorielle) est le produit de la pression (scalaire) par la vitesse (vecteur). Dans les conditions de champ libre, l'intensité acoustique est proportionnelle au carré de la pression acoustique. Quand la distance à la source est doublée, la pression est divisée par deux (et l'intensité acoustique est divisée par quatre). Le niveau de pression acoustique diminue alors de 6 dB.

Une question intéressante, dans le cas d'un environnement fermé, est de se demander ce qu'il se passe lorsqu'une source émet à des fréquences qui « n'ont pas la place » de se développer au sein du volume disponible, i.e. lorsque les longueurs d'ondes considérées sont supérieures aux dimensions de l'espace dans lequel émet la source. Ce cas se produit pour les basses fréquences émises par un moteur au sein de son encoffrement, ce dernier étant approximativement un cube de 60 cm de côté.

Et bien, dans ce cas, l'énergie injectée « gonfle » en quelque sorte l'encoffrement. Celui-ci est sollicité en pression... et répond évidemment selon ses modes propres. Dans de nombreux cas (encoffrement moteur, habitacle de véhicule...), il est nécessaire de prendre en compte la réponse modale du volume dans lequel on définit le problème, car celle-ci influe sur la répartition du champ acoustique, dont notamment d'éventuelles résonances.

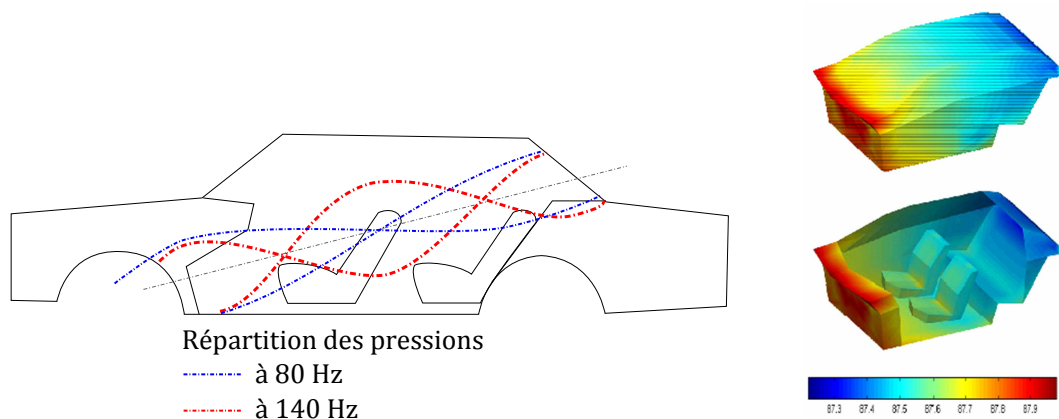


FIGURE 19.6: a) Répartition modale dans un véhicule et b) non-uniformité du champ acoustique

Le champ acoustique dans de telles situations (et dans d'autres) est non uniforme, i.e. a des valeurs très différentes d'un point à l'autre. Ces deux aspects sont illustrés à la figure 19.6.

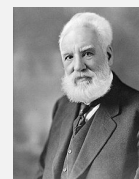
19.1.3 Réception

Histoire

Le bel (B) est utilisé dans les télécommunications, l'électronique et l'acoustique. Vers 1920, les entreprises de téléphonie utilisaient comme unité pour l'atténuation le msc, valant celle d'un mile (1,6 km) de câble standard à la fréquence de 800 Hz. Des ingénieurs des Laboratoires Bell définirent une unité de transmission indépendante du câble et de la fréquence, basée sur dix fois le logarithme décimal. Cette unité s'appela d'abord TU pour « Transmission Unit » (unité de transmission). Elle

présentait l'avantage d'être presque équivalente au msc ($1 \text{ TU} = 1,083 \text{ msc}$). Elle fut renommée décibel en 1923 ou 1924 en l'honneur du fondateur du laboratoire et pionnier des télécoms, Alexander Graham Bell.

Les Laboratoires Bell consultèrent les opérateurs téléphoniques et administrations responsables. Certaines utilisaient des logarithmes népériens, qui présentent certains avantages pour le calcul, avec une unité appelée le néper (symbole Np). Les deux unités ont coexisté, mais le néper n'a pas connu le succès du décibel. La norme ISO 80000-3, page viii, dit d'ailleurs : « L'utilisation du néper est le plus souvent limitée à des calculs théoriques sur des grandeurs de champ, où cette unité est la plus commode, alors que, dans d'autres cas, en particulier pour des grandeurs de puissance, le bel (B), ou en pratique son sous-multiple, le décibel (dB), est largement utilisé. Il convient de souligner que le fait que le néper soit choisi comme l'unité cohérente n'implique pas qu'il convienne d'éviter d'utiliser le bel. Le bel est accepté par le CIPM et l'OIML pour être utilisé avec le SI. À certains égards, cette situation est similaire au fait que l'unité degré ($^{\circ}$) est utilisée couramment à la place de l'unité SI cohérente radian (rad) pour les angles plans. »



Bell

L'usage du logarithme décimal était d'autant plus pratique qu'avant la diffusion des calculatrices électroniques, on se servait pour les calculs de tables de logarithmes décimaux. Lorsqu'on se propose de calculer l'atténuation dans une ligne de longueur l et de coefficient d'atténuation α , il faut élever $(1 - \alpha)$ à la puissance l . En pratique, on cherchait $l \log(1 - \alpha)$ dans la table avant de reconvertir le logarithme en rapport.



Fechner



Weber



Stevens

Le décibel a connu un grand succès dans le domaine de l'acoustique. Par une coïncidence fortuite, un décibel, en puissance sonore, correspond à peu près à la plus petite variation perceptible par l'oreille. Selon le philosophe et psychologue Gustav Fechner, la sensation ressentie varierait comme le logarithme de l'excitation. Une unité à progression logarithmique semblait particulièrement pertinente dans un domaine où la perception humaine était en jeu. Des recherches ultérieures

sont venues contester la loi de Weber-Fechner, comme la loi de puissance de Stevens (1951) ; mais l'usage du décibel s'était établi, même dans des cas où il complique la compréhension.

La loi de Weber-Fechner stipule que la sensation perçue, S , est proportionnelle au logarithme de l'intensité, I , du stimulus :

$$S = k \log(I) \quad (19.6)$$

alors que la loi de puissance de Stevens modélise cette relation entre ces deux grandeurs par une loi puissance :

$$S = kI^a \quad (19.7)$$

où, dans les deux cas, k est une constante. Stevens a donné des valeurs obtenues pour l'exposant a dans différents domaines dans une publication de 1957 et dans son ouvrage de 1951 *Handbook of Experimental Psychology*.

Décibels

La notion de réception sera forcément liée à une échelle de mesure (une métrique). Nous avons plusieurs fois déjà utilisé le décibel (dB). Toutefois, il convient de distinguer plusieurs échelles :

- Le décibel (dB) est une unité de grandeur adimensionnelle définie comme 10 fois le logarithme décimal du rapport entre deux puissances. Pour l'acoustique, ce rapport des puissances est défini entre la grandeur mesurée et une valeur de référence fixée par une norme.

$$n(\text{en dB}) = 10 \log_{10}(I_1/I_0) \quad (19.8)$$

où I_1 et I_0 sont deux intensités.

- Dans la formule précédente, si I_0 correspond au seuil de perception ($10^{-12} \text{ Watt/m}^2$), on obtient le décibel absolu ou dB lin ou SPL. C'est celui que l'on utilise dans les calculs et

analyses.

$$n(\text{en dB}) = 10 \log_{10}(I_1/10^{-12}) \quad (19.9)$$

Si $I_1 = 10^{-12} \text{ W/m}^2$ (seuil de perception), on obtient 0 dB. Si $I_1 = 1 \text{ W/m}^2$ (seuil de douleur), on obtient 120 dB.

- Le dBA traduit la perception par l'oreille du phénomène (protection physiologique par filtrage). C'est souvent lui qu'il faut utiliser en lien avec des normes adéquates (fonction du problème considéré) pour savoir s'il y a nuisance sonore ou non.
- Il existe d'autres pondérations : le dBC est quasiment linéaire sur plusieurs octaves et est utile pour les mesures subjectives avec de fortes pressions acoustiques. Le dBB est entre le dBA et le dBC. Le dBD est appliqué pour tenir compte de la gêne causée par les sifflements (à hautes fréquences) perçus à l'intérieur des avions modernes. Le dBG est adapté aux basses fréquences...
- Notons que, pour être plus précises, ces pondérations devraient varier avec l'âge : en effet, la perception en fonction de la fréquence et de la puissance varie avec l'âge, ainsi que l'étendue de fréquences audibles (qui diminue, phénomène appelé presbyacousie), comme on le voit sur les courbes (ou diagrammes) de Fletcher et Munson, présentées dans leur article *Loudness, its definition, measurement and calculation* paru en 1933 dans le Journal of the Acoustic Society of America, et que l'on appelle également courbes isosoniques. Fletcher et Munson travaillaient tous les deux aux laboratoires Bell.

Grandeurs acoustiques

À ce niveau, il convient de rappeler quelques définitions :

- La pression acoustique (en N/m^2) est l'amplitude des variations de la pression en un point de l'espace, par rapport à la pression atmosphérique, provoquées par le passage d'une onde sonore en ce point.
- La puissance acoustique (en W) est la quantité (ou le flux) d'énergie acoustique qu'émet une source par unité de temps.
- L'intensité acoustique (en W/m^2) est le flux d'énergie acoustique qui est transmis dans une direction donnée, à travers une unité de surface, pendant une unité de temps. Elle dépend de la puissance de la source, du milieu de propagation, de la distance à la source.

L'intensité acoustique s'exprime comme :

$$I = \frac{p^2}{Z} \quad (19.10)$$

où p est la pression acoustique et Z l'impédance (qui s'exprime en Pa.s/m ou en Ns/m^3).

Remarque. Dans le cadre de mesures :

- l'intensité acoustique est représentée par la puissance électrique (W) ;
- la pression acoustique est représentée par la tension électrique (V) (la tension électrique fournie par un microphone est proportionnelle à la pression acoustique p) ;
- l'impédance acoustique est représentée par la résistance électrique.

On peut remplacer $I = p^2/Z$ par $P = V^2/R$ et en déduire la puissance moyenne :

$$P_{\text{moy}} = \frac{V_1^2 + \dots + V_n^2}{nR} \quad (19.11)$$

On définit aussi la tension efficace ou tension RMS (Root Mean Square), qui est la tension continue qui donnerait une puissance continue égale à la puissance moyenne. On a donc $P_{\text{moy}} = V_{\text{eff}}^2/R$, soit :

$$V_{\text{eff}} = \sqrt{\frac{V_1^2 + \dots + V_n^2}{n}} \quad (19.12)$$

Notons que l'on « parle » souvent, dans le monde des acousticiens, en termes de bandes de fréquence, l'analyse en bandes fines n'étant utilisée que pour des analyses plus poussées.

L'octave étant le doublement de la fréquence $f_2 = 2f_1$, le tiers d'octave est défini par un rapport entre les fréquences de $f_2 = 2^{1/3} f_1$.

Les fréquences centrales de l'analyse par tiers d'octave normalisée sont les suivantes (en Hz) : 20, 25, 31, 40, 50, 63, 80, 100, 125, 160, 200, 250, 315, 400, 500, 630, 800, 1000, 1250, 1600, 2000, 2500, 3150, 4000, 5000, 6300, 8000, 10000, 12500.

Le bruit blanc est un son dont le spectre contient toutes les fréquences à la même amplitude. Une analyse à bande de fréquence constante donnerait donc une courbe constante égale à 1.

Le bruit rose est un bruit dont le niveau par bande d'octave est constant. Ce signal se rapproche plus de la sensibilité de l'oreille que le bruit blanc. Pour cette raison, le bruit rose est donc souvent utilisé dans l'univers audible pour calculer la réponse fréquentielle d'une chaîne de reproduction sonore.

Impédance

La notion d'impédance mérite que nous y passions quelques instants. En effet, elle nous permettra par la suite de définir des conditions aux limites pour nos modèles numériques.

L'impédance caractéristique d'un milieu (solide, liquide ou gazeux) est définie comme le rapport de la pression acoustique sur la vitesse de déplacement en milieu ouvert (i.e. en l'absence d'ondes réfléchies). C'est une propriété du matériau considéré qui est égale, dans le cas d'un espace illimité, au produit de la masse volumique du matériau par la vitesse du son dans ce même matériau.

L'impédance acoustique dans les solides est définie par :

$$Z = \frac{F}{v} = \rho c \quad (19.13)$$

où F est la force (N), v la vitesse de déplacement (m/s), ρ la densité linéique du milieu (kg/m) et c la vitesse de propagation (m/s). On remarquera que c et v sont inversement proportionnelles. Dans les solides, la vitesse de propagation est plus élevée que dans les gaz et l'énergie s'y dissipe moins rapidement : l'onde peut se propager plus loin.

L'impédance acoustique dans les gaz est définie par :

$$Z = \frac{p}{v} = \rho c \quad (19.14)$$

où cette fois p est la pression acoustique (Pa), les autres grandeurs étant les mêmes qu'auparavant.

Quelques valeurs. L'impédance de l'air vaut $Z = 413,5$ (en Pa.s/m ou N.s/m^3) à 20°C . Elle varie avec la température, tout comme la densité et la célérité du son.

La vitesse du son dans l'air vaut $c = 331,5 + 0,6\theta$ m/s, où θ est la température de l'air en $^\circ\text{C}$ (elle varie également en fonction de la pression atmosphérique et de la température).

L'impédance acoustique de l'eau est d'environ $1,5 \cdot 10^6$ Pa.s/m .

Considérons deux milieux d'impédances respectives Z_1 et Z_2 , et une onde incidente se propageant dans le milieu 1 jusqu'à arriver à sa frontière avec le milieu 2. Si $Z_1 = Z_2$, alors l'onde incidente se propage intégralement dans le milieu 2, i.e. elle est intégralement transmise. Par contre, si $Z_1 \neq Z_2$, alors l'onde incidente se scinde en une onde réfléchie et une onde transmise. L'amplitude de l'onde réfléchie est d'autant plus grande que la discontinuité entre les deux milieux est marquée. On définit alors le coefficient de réflexion par :

$$\frac{p_r}{p_i} = \frac{Z_2 - Z_1}{Z_2 + Z_1} \quad (19.15)$$

où les indices r et i désignent l'onde réfléchie et l'onde incidente respectivement.

Supposons qu'une onde se propage dans l'air ($Z_1 = 413,5$) et rencontre un solide tel que $Z_2 = 500000$. Alors le coefficient de réflexion est de 99,8%. Dans le cas où une onde passe d'un milieu plus dense à un milieu moins dense, le coefficient de réflexion devient négatif.

Notons les deux cas extrêmes suivants :

- la réflexion à une extrémité libre : dans ce cas, $Z_2 \approx 0$ et la vitesse de l'extrémité est $v = 2v_i$. On assiste à un changement de signe de v_r sans renversement de la forme de l'onde.
- la réflexion à une extrémité fixe : dans ce cas, $Z_2 \rightarrow \infty$ et la vitesse de l'extrémité est $v = 0$. Il n'y a pas de changement de signe de v_r , mais la forme de l'onde est renversée.

Normes, indices et psycho-acoustique

De très nombreuses normes et grandeurs existent dans le domaine de l'acoustique, correspondant à des problématiques différentes. Nous en listons quelques unes ci-dessous, mais cette liste est très loin d'être exhaustive :

- Le comportement acoustique d'une salle diffère selon la zone de fréquence considérée. Pour les longueurs d'ondes supérieures au double de la plus grande dimension d'une pièce : le son se comporte de manière équivalente aux changements de pression de l'air statique. Pour les longueurs d'onde comparables aux dimensions de la pièce : il y a dominance des modes propres de la salle. Il y a amplification des fréquences, ce qui « colore » le son et création d'ondes stationnaires. Pour des longueurs d'ondes inférieures à la moitié des dimensions de la salle : on est dans le cas de réflexions multiples, et on peut faire une équivalence lumineuse et utiliser les méthodes de tir de rayons.
- L'une des premières grandeurs à connaître est le bruit ambiant. Son niveau maximal est fixé par des normes en fonction de la destination du local.
- Une seconde grandeur particulièrement fondamentale pour la caractérisation de locaux, quelque soit leur destination, est le temps de réverbération ou TR. Il correspond à l'intervalle de temps nécessaire pour que la pression acoustique (d'une salle) diminue à un millième de sa valeur initiale suite à l'arrêt de la source sonore. Cela représente une diminution du niveau sonore de 60 dB. Le temps de réverbération indique la capacité qu'à la salle à réverbérer les sons.

En 1898, Sabine découvrit que le temps de réverbération T_R était proportionnel au volume de la salle V et inversement proportionnel à la surface équivalente d'absorption A :

$$T_R = \frac{kV}{A} \quad (19.16)$$

où la constante de proportionnalité k , trouvée initialement égale à 0,161, vaut $k = (24 \ln 10)/340 = 0,163$ s/m, et où l'aire d'absorption équivalente est calculée par les formules de Sabine (19.4) ou d'Eyring (19.5).

Selon la destination du local, plusieurs normes existent pour mesurer le TR. De plus, la destination du local joue également sur la valeur acceptable ou non du TR.

Le Early Decay Time (EDT) est le temps de décroissance sur les 10 premiers dB. Il est subjectivement plus important, car il se rapproche de l'impression de réverbération alors que le TR fait référence aux propriétés du local.

On utilise également le taux de décroissance spatiale DL, et plus particulièrement le taux de décroissance par doublement de distance DL_2 .

- La Clarté dépend de la distribution temporelle de l'énergie précoce, de la densité temporelle, du niveau des réflexions et de la nature du message sonore. L'intelligibilité de la parole est conditionnée par le TR, le rapport signal/bruit, le volume et la géométrie de la pièce, et la répartition des surfaces réfléchissantes et absorbantes. Plus le temps de réverbération est court, meilleure sera la compréhension de la parole, à moins que le bruit de fond ne domine. Plusieurs indicateurs seront à prendre en compte : le rapport Signal-Bruit (S/B), le Speech Transmission Index (STI), le Percentage Articulation Loss of Consonants (%-Alcons), le

RApid Speech Transmission Index (RASTI)...

$$\begin{aligned}
 \text{Clarté sur 80 ms : } C_{80} &= 10 \log_{10} \frac{\int_0^{80} p^2(t) dt}{\int_{80}^{\infty} p^2(t) dt} \\
 \text{Distinctness sur 50 ms : } D &= \frac{\int_0^{50} p^2(t) dt}{\int_0^{\infty} p^2(t) dt} \\
 \text{Temp central : } T_c &= \frac{\int_0^{\infty} t p(t) dt}{\int_0^{\infty} p(t) dt} \\
 \text{Articulation Loss Consonent : } ALC &= \frac{200 r^2 T_R^2}{VQ} \\
 \text{RApid Speech Transmission Index : } RASTI &= \begin{cases} 0 & \text{si } (S/B)_{app} \leq 15 \\ 1 & \text{si } (S/B)_{app} \geq 15 \\ \frac{(S/B)_{app} + 15}{30} & \text{si } 15 < (S/B)_{app} < 15 \end{cases} \\
 \text{avec } (S/B)_{app} &= \frac{1}{9} \sum_{i=1}^9 (S/B)_i
 \end{aligned}$$

où Q est le facteur de directivité. Des relations empiriques sont également utilisées :

$$ALC \approx 170, 5e^{-592STI}$$

$$STI \approx -0, 18 \ln(ALC) + 0, 95$$

- De là, on peut définir les rayons de discrétion et de confidentialité...
- D'autres indicateurs existent liés par exemple à la spatialisation (InterAural Cross Correlation (IACC), Lateral Efficiency (LE)...) ou à des critères de sonorisation (clarté locale, homogénéité...).
- Des indicateurs prennent également en compte la durée d'exposition au bruit, surtout évidemment dans le cadre de la législation du travail. Ainsi 80 dBA pendant 8 heures correspondent à 86 dBA pendant 2 heures, 90 dBA pendant 45 min, 95 dBA pendant 15 min, 100 dBA pendant 5 min, 107 dBA pendant 1 min ou 115 dBA pendant 28 s.
le $L_{ex,d}$ est le niveau d'exposition sonore quotidienne exprimé en dBA pour des bruits stables ou fluctuants. Le L_{pc} est le niveau de pression crête exprimé en dBC pour une exposition à des bruits impulsionnels.
- L'Organisation Internationale de Normalisation (ISO) a proposé plusieurs courbes qui correspondent toutes à un certain degré de confort acoustique (ou de gêne) : courbes d'évaluation du bruit, ou courbes de Noise Rating (NR). Grâce à ces courbes, il est possible de déterminer au moyen d'un seul chiffre le niveau de pression acoustique maximum autorisé. Les indices N40 et N80 sont souvent pris en référence.
- Dans *Éléments de physiologie et de pathologie des bruits*, Wisner définit trois seuils de bruit correspondant à un indice de gêne subjective : lorsque le bruit ambiant est en zone 1, le travail intellectuel n'est pas gêné de façon appréciable ; lorsque le bruit ambiant est en zone 2, le travail intellectuel est pénible, le travail courant n'est pas gêné de façon appréciable ; lorsque le bruit ambiant est en zone 3, le travail intellectuel est extrêmement pénible, le travail courant est difficile ; et lorsque le bruit ambiant est situé en zone 4, une exposition prolongée peut conduire à la surdité.
- La Sonie ou bruyance (loudness) est une quantification de la perception du son chez l'être humain. C'est une grandeur psycho-acoustique qui se rattache de façon complexe à la pression acoustique. Les courbes isosoniques issues des travaux de Fletcher et Munson expriment la relation entre la fréquence d'un stimulus sonore continu et la perception de sa sonie. L'ISO a donné un tracé normalisé afin de pouvoir définir le phone, unité de l'expression de la sonie.

- Il est toujours possible de construire d'autres métriques pour des problèmes particuliers. Par exemple dans [58], on trouvera une métrique adaptée à la mesure de la « beauté » des bruits périodiques.

19.2 Calculs acoustiques par éléments finis

19.2.1 Modèles simplifiés pour valeurs de référence

Dans ce petit paragraphe, nous donnons quelques formules simples. Elles permettent néanmoins d'effectuer un prédimensionnement rapide pour les salles acoustiques et/ou de disposer de quelques valeurs (au moins leur ordre de grandeur) vers lesquelles des calculs plus complexes doivent conduire.

Puissance de la source sonore et propagation

En champ libre, on a la loi de propagation pour le champ direct suivante :

$$L_p = L_w + 10 \log_{10} \left(\frac{Q}{4\pi d} \right) \quad (19.17)$$

où L_p (en dBA) est le niveau de pression sonore du champ direct, L_w le niveau de puissance de la source, d la distance à la source et Q le facteur de directivité ($Q = 1$ source omnidirectionnelle, $Q = 2$ source omnidirectionnelle posée sur le sol...).

Dans une salle (suffisamment réverbérante), on constate que le champ est diffus, et alors on a :

$$L_p = L_w + 6 - 10 \log_{10} A \quad (19.18)$$

où L_p (en dBA) est le niveau de pression sonore du champ diffus, L_w le niveau de puissance de la source et A l'aire équivalente d'absorption. On remarque que L_p est constant quelle que soit la distance à la source (puisque l'on est en hypothèse de champ diffus) et qu'il ne dépend que de A , i.e. de la capacité d'absorption disponible dans la salle. On notera que les parois, surtout si elles sont légères, présentent une certaine transparence et qu'il est bon, pour ajuster le modèle, d'augmenter l'aire d'absorption de la proportion d'énergie qui quitte la salle par transmission (défaut d'isolation).

En terme de modélisation, cela signifie qu'à proximité d'une source (jusqu'à environ 30cm), on considérera que l'on est en champ direct. Au-delà, on utilisera la relation correspondant au champ réverbéré.

Modes d'une salle

Pour les longueurs d'onde de l'ordre des dimensions de la salle, il y a prédominance des modes propres de la salle. Il faut donc pouvoir les évaluer. Pour une salle parallélépipédique de dimensions L , l et h , les fréquences de résonance sont données par :

$$f_{(i,j,k)} = \frac{c}{2} \sqrt{\left(\frac{i}{L}\right)^2 + \left(\frac{j}{l}\right)^2 + \left(\frac{k}{h}\right)^2} \quad (19.19)$$

Les modes $(i, 0, 0)$, $(0, j, 0)$ et $(0, 0, k)$ sont des modes axiaux dans les directions de L , l et h respectivement, les modes $(i, j, 0)$, $(0, j, k)$ et $(i, 0, k)$ les modes tangentiels, et les modes (i, j, k) les modes obliques. La pression acoustique se met sous la forme :

$$p(x, y, z) = \sum_i \sum_j \sum_k p_{ijk} \cos\left(\frac{i\pi}{L}x\right) \cos\left(\frac{j\pi}{l}y\right) \cos\left(\frac{k\pi}{h}z\right) \quad (19.20)$$

La formule de Maa (1939) donne le nombre $N(f)$ de fréquences propres inférieures ou égales à une fréquence donnée f :

$$N(f) = \frac{4\pi}{3} V \left(\frac{f}{c} \right)^3 + \frac{\pi}{4} S \left(\frac{f}{c} \right)^2 + \frac{L_t}{8} V \left(\frac{f}{c} \right) \quad (19.21)$$

où $V = Llh$ est le volume de la salle, $S = 2(Ll + Lh + lh)$ est la surface totale des parois de la salle, et $L_t = 4(L + l + h)$ est la longueur totale des arêtes de la salle.

En dérivant la formule de Maa par rapport à la fréquence, on peut déterminer la densité de fréquences propres à une fréquence donnée, i.e. le nombre de fréquences propres comprises dans un intervalle de 1 Hz centré sur f :

$$dN(f) = \frac{4\pi V}{c^3} f^2 + \frac{\pi S}{2c} f + \frac{L_t}{8c} \quad (19.22)$$

Pour une salle non parallélépipédique, on peut, sauf géométrie vraiment très biscornue, utiliser les relations précédentes en considérant le parallélépipède le plus proche de la géométrie réelle.

Remarques

On n'oubliera pas de prendre en compte les pertes par transmission dans le calcul des coefficients d'absorption.

Dans le cas où l'on s'intéresse à la transmission entre plusieurs salles, il faudrait effectivement ne pas oublier qu'une salle n'a pas que des sources internes, mais également des sources externes correspondant à ce qui a pu s'échapper des autres salles, ou de ce qui rayonne suite à la propagation de vibrations.

Nous avons présenté quelques formules dans le cas de l'acoustique des salles, mais d'autres formules relatives à d'autre problématiques acoustiques typiques existent : par exemple pour l'encoffrement (ou capotage) de sources acoustiques (machines...).

Méthodes de composition de source. Toutes ces méthodes « simples » permettent néanmoins de réaliser, à moindre frais, des modèles de calculs (analytiques) pouvant prendre en compte des phénomènes assez complexes. C'est ce que l'on appelle les méthodes de composition de sources.

Dans ces méthodes, on commence par identifier toutes les sources d'un problème : non seulement les sources physiques comme un moteur, mais également les sources induites comme le bruit de contact pneu/chaussée, le bruit aérodynamique, le rayonnement de surfaces... Ces sources sont fonction des données du problème telles que la vitesse du véhicule, le rapport de boîte, etc. Les sources sont positionnées dans l'espace, et chacune est prise en compte « indépendamment » des autres. L'introduction de couplages est possible, mais peut être délicate (notamment pour les phénomènes tels que l'amortissement). Ensuite, on introduit divers « facteurs de perte » entre les sources et le point d'écoute : matrices de transfert diverses dues à la distance à la source, l'absorption et l'isolation de divers éléments, l'amortissement et le filtrage... On voit bien que la mise au point d'un tel modèle peut se révéler particulièrement compliquée, même pour quelqu'un ayant une grande expérience dans le domaine de l'acoustique.

Toutefois, ces méthodes présentent l'immense intérêt de fournir les résultats instantanément (ou presque pour des méthodes semi-analytiques), et valables dans toutes les gammes de fréquences (aux hypothèses faites près, ce qui peut être restrictif). On voit également que, de par leur nature, ces méthodes sont essentiellement interpolatoires et leur utilisation extrapolatoire est plus que délicate, conduisant généralement à des résultats complètement faux. Néanmoins, elles restent pertinentes dans de nombreux cas industriels bien bordés, et si elles sont corrélées à des résultats expérimentaux, conduisent de surcroît à une interprétation aisée de ceux-ci, permettant de trouver des tendances et des liens intéressants. Nous insistons malgré tout encore une fois sur la grande expérience (de l'acoustique) nécessaire pour mettre ces modèles au point.

C'est une lapalissade, mais dans un calcul global, les phénomènes locaux (écranage...) ne sont pas pris en compte (ou alors de manière globale)... Si l'on s'intéresse au niveau moyen résultant dans un local, prendre en compte la totalité des surfaces absorbantes, qu'elles soient situées au murs, au plafond ou sous formes d'écrans, de fauteuils... fonctionne. Par contre, si l'on a besoin de descendre dans le détail local de la répartition de la pression acoustique (par exemple répartition

poste à poste), alors il convient d'utiliser des méthodes plus adéquates, dont les éléments finis, que nous allons présenter maintenant.

19.2.2 Constitution d'un modèle éléments finis

Pour réaliser notre modèle de calcul vibro-acoustique par éléments finis, nous devons disposer de sources et de conditions aux limites. Nous allons aborder ces deux points.

Sources

Si l'on ajoute des termes de sources, alors l'équation des ondes (6.12) devient :

$$\frac{1}{\rho c^2} \frac{\partial^2 p}{\partial t^2} + \nabla \cdot \left(-\frac{1}{\rho} (\nabla p - q_d) \right) = q_m \quad (19.23)$$

dont l'inconnue est toujours la pression acoustique et où q_d et q_m représentent les termes de sources pour un dipôle et un monopôle respectivement. Pour les milieux dispersifs, un terme de dissipation supplémentaire peut être introduit dans l'équation des ondes qui devient :

$$\frac{1}{\rho c^2} \frac{\partial^2 p}{\partial t^2} - d \frac{\partial p}{\partial t} + \nabla \cdot \left(-\frac{1}{\rho} (\nabla p - q_d) \right) = q_m \quad (19.24)$$

Dans la suite, on restera dans le cas $d = 0$.

Dans le cas harmonique, l'équation (19.23) devient l'équation d'Helmholtz (8.59) avec termes sources :

$$-\frac{\omega^2 p}{\rho c^2} + \nabla \cdot \left(-\frac{1}{\rho} (\nabla p - q_d) \right) = q_m \quad (19.25)$$

Remarque. Si l'on souhaite introduire le d'amortissement dans l'équation, alors les quantités c et ρ peuvent être complexes (au sens appartiennent à \mathbb{C}). Par exemple, pour un modèle d'amortissement de Rayleigh, la célérité c est une combinaison linéaire de la masse et de la raideur.

Dans les codes de calcul, la source est généralement modélisée par un monopôle ou un dipôle. Un monopôle est une source omnidirectionnelle, alors qu'un dipôle représente une source ayant deux directions opposées selon lesquelles le champ est plus fort. Un dipôle peut être vu comme deux monopôles de même puissance séparés par une petite distance et en opposition de phase. On peut également utiliser des quadripôles si besoin. On trouve également parfois des sources ponctuelles, linéaires ou ayant d'autres formes prédéfinies.

Rayonnement

En plus de sources définies ci-dessus, le phénomène de rayonnement peut être à l'origine d'une excitation acoustique. Le rayonnement est par nature un phénomène couplé fluide-structure, puisqu'il s'agit du cas où une structure, en vibrant, génère une onde acoustique. C'est un phénomène complexe, qui fait toujours l'objet de recherches. Nous allons essayer d'en dire deux mots en restant sur le mode de la vulgarisation.

Compléments sur le rayonnement. Toute structure vibrant ne génère pas de bruit, ou plus exactement, une structure soumise à un spectre ne rayonne pas l'intégralité du spectre. Certes, l'effet de l'amortissant structurel est sans aucun doute à prendre en compte, mais il n'est pas suffisant. En fait, seuls certains modes de la structure rayonnent, et ce sont ceux qui correspondent à des projections du nombre d'onde de la vibration à laquelle est soumise la structure. Il appert donc que l'utilisation d'une base modale dans le calcul numérique peut être à même de prendre en compte cette projection, tout en réduisant la taille des calculs. On voit également que la notion de vitesse de vibration, et par suite de vitesse des déformations, est d'importance dans ce phénomène. Elle peut être liée, au moins localement, à des phénomènes de compensation de pression acoustique qui peut ou non se faire pendant le temps de déformation.

Définition 79 — Fonction de Green. On appelle fonction de Green une solution d'une équation ayant au second membre un terme impulsionnel.

Si l'équation considérée est l'équation des ondes, on appelle fonction de Green en temps une solution de l'équation des ondes avec au second membre un terme source de type source ponctuelle en espace et en temps, i.e. source localisée en un point de l'espace x_0 et émettant une impulsion à un instant t_0 localisé dans le temps. On a donc pour second membre de la forme : $\delta(t - t_0)\delta(x - x_0)$ (où x est un vecteur). Elle s'écrit sous la forme :

$$G(x, t, x_0, t_0) = \frac{1}{4\pi \|x - x_0\|} \delta\left(t - t_0 - \frac{\|x - x_0\|}{c}\right) \quad (19.26)$$

Si l'équation considérée est l'équation d'Helmholtz, on appelle fonction de Green en fréquence une solution de l'équation d'Helmholtz avec au second membre un terme source ponctuelle en espace, localisé au point x_0 . Elle s'écrit sous la forme :

$$G(x, x_0) = -\frac{1}{4\pi \|x - x_0\|} e^{ik\|x - x_0\|} \quad (19.27)$$

Dans le cas du rayonnement, on s'intéresse à ce qui se passe au niveau de l'interface entre une structure vibrant et un fluide, où l'on doit avoir continuité des vitesses acoustiques et mécaniques normales.

Théorème 63 — Théorème de Kirchhoff en fréquence. Le théorème de Kirchhoff dit que le champ de pression, en régime fréquentiel, en un point d'observation x donné, est entièrement déterminé par les valeurs de la pression et de sa dérivée normale sur les bords du domaine et la donnée de la source volumique $q_m(x_0)$. Il stipule que le champ de pression au point x est donné par la relation :

$$p(x) = \int_{\Gamma=\partial\Omega} p(x_0) \frac{\partial G(x, x_0)}{\partial n_0} - G(x, x_0) \frac{\partial p(x_0)}{\partial n_0} d\Gamma + \int_{\Omega} q_m(x_0) G(x, x_0) d\Omega \quad (19.28)$$

Si l'on considère une plaque plane, dans le plan (x, y) , vibrant, alors la condition cinématique de continuité des vitesses normales à l'interface surface plane/fluide impose la condition suivante sur la vitesse des particules de fluide en $z = 0$:

$$v(x).z = v_0(x_0)$$

Et en projetant l'équation d'Euler linéarisée sur l'axe z , on peut alors relier la dérivée normale de la pression à la vitesse normale de la surface S :

$$\frac{\partial p}{\partial n} = \frac{\partial p}{\partial z} = i\omega\rho v_0(x_0) \quad (19.29)$$

En l'absence de toute autre source acoustique, le théorème de Kirchhoff ne comporte plus de terme volumique. En introduisant la fonction de Green et la valeur de la pression normale, on obtient la relation suivante, connue sous le nom d'intégrale de Rayleigh, qui permet de calculer le champ de pression acoustique rayonné par une surface plane S dont on connaît uniquement la vitesse normale :

$$p(x) = \frac{i\omega\rho}{2\pi} \int_S \frac{e^{ik\|x - x_0\|}}{\|x - x_0\|} v_0(x_0) dS \quad (19.30)$$

Remarque. Le cas du piston circulaire, que l'on trouve dans moult ouvrages, et que nous ne présenterons pas, est intéressant et serait à regarder car il se retrouve dans beaucoup d'applications telles que les haut-parleurs, les transducteurs... Le formalisme de Rayleigh est bien adapté à ce cas qui, s'il est simple d'un point de vue structurel, a un comportement acoustique complexe.

Enfin, on définit l'impédance de rayonnement Z_R comme le rapport de la pression à la surface divisée par la vitesse vibratoire. C'est un nombre complexe :

$$Z_R = \lambda_R + i\omega M_R \quad (19.31)$$

où la partie réelle λ_R agit comme un terme d'amortissement, et la partie imaginaire M_R comme un terme de masse ajoutée. Cela est à ajouter à ce qui a déjà été présenté au sujet de l'impédance au paragraphe 19.1.3.

Conditions aux limites

Disposant de sources vibro-acoustiques, il nous faut maintenant ajouter des conditions aux limites sur notre modèle. Les conditions aux limites classiques sont :

- La condition de Neumann correspond à une condition de mur rigide. Elle définit une frontière sur laquelle la composante normale de l'accélération est nulle (mur immobile parfaitement réfléchissant) :

$$-n \cdot \left(-\frac{1}{\rho} (\nabla p - q_d) \right) = 0 \quad (19.32)$$

S'il n'y a pas de dipôle (i.e. $q_d = 0$) et si la densité du fluide est constante (i.e. $\rho = \text{cte}$), alors l'équation précédente devient :

$$\frac{\partial p}{\partial n} = 0 \quad (19.33)$$

i.e. la dérivée de la pression par rapport à la normale est nulle.

Dans certains codes, on peut spécifier la valeur de l'accélération normale, i.e. le second membre de l'équation (19.32) n'est plus nul mais vaut a_n (condition de Neumann relevée), qui représente alors une source externe. Cela permet également de coupler de manière faible l'acoustique à un calcul structural : un premier calcul de structure permet de déterminer les vibrations d'une surface, et celles-ci sont introduites ensuite dans le calcul acoustique. Le couplage est donc faible car dans ce cas on fait l'hypothèse que l'acoustique en retour ne modifie pas (ou peu) la vibration de la structure. De manière inverse, on a également couplage faible si le fluide agit sur le solide et que celui-ci en retour n'agit que peu sur le fluide.

- La condition de Dirichlet correspond à une condition d'absorption parfaite. Elle est telle que la pression acoustique s'annule sur la surface :

$$p = 0 \quad (19.34)$$

En général, dans les codes, une telle condition permet de mettre la pression à zéro, ou à toute autre valeur (condition de Dirichlet relevée) définie par l'utilisateur. Le second membre n'est alors plus nul mais vaut \bar{p} . Cela permet de modéliser une source débitant une pression constante (applicable sur un nœud si besoin).

- La condition de Robin correspond à une condition d'impédance. Elle se situe entre les deux cas précédents :

$$\begin{aligned} -n \cdot \left(-\frac{1}{\rho} (\nabla p - q_d) \right) &= -\frac{i\omega p}{Z} \text{ en analyse harmonique, ou} \\ &= \frac{1}{Z} \frac{\partial p}{\partial t} \text{ en analyse temporelle} \end{aligned} \quad (19.35)$$

Lorsque $Z \rightarrow \infty$ on retrouve le cas du mur rigide, et lorsque $Z \rightarrow 0$ celui de l'absorption parfaite.

Pour les plaques perforées, on utilisera cette condition pour gérer l'absorption. Certains logiciels calculent l'impédance à notre place en fonction de paramètres physiques tels que : la viscosité dynamique, l'épaisseur et la porosité (diamètre des trous, distance entre trous, ou aire des trous et résistance de passage au fluide) et la réactance du panneau.

- L'absorption peut être gérée par l'impédance, mais certains logiciels incluent des modèles de fluides équivalent (homogénéisation), dont le plus connu est celui de Biot-Allard. Il concerne les matériaux poreux et nécessite de connaître leur porosité, résistance au passage de l'air, tortuosité, et longueurs caractéristiques visqueuse et thermique (ces deux derniers paramètres ayant une influence moins importante que les trois premiers sur l'absorption). D'autres modèles existent comme le modèle de Johnson-Champoux-Allard, celui de Johnson-Champoux-Allard-Lafarge...
- Les conditions aux limites en rayonnement permettent de définir les régions par lesquelles les ondes peuvent quitter le modèle. On trouve généralement des conditions de radiation plane, cylindrique et sphérique.
- De la même manière on trouve des conditions de rayonnement, plane ou cylindrique, permettant de modéliser une source, i.e. une onde qui arrive (et non plus une onde qui quitte le modèle).
- Lorsque le couplage faible décrit plus haut n'est pas suffisant, il faut alors intégrer le couplage fort fluide-structure au calcul : le fluide, défini par ses champs de pression et de vitesse, exerce des forces de type aérodynamique (de pression) sur la structure. Cette force aérodynamique se met sous la forme $F = \frac{1}{2} \rho v^2 SC$, où v est la vitesse, S la « surface de référence » (pour nous cette surface de référence sera la surface mouillée)¹ et C le coefficient aérodynamique. En retour, le déplacement et/ou la déformation de la structure affecte, au moins localement, le champ de l'écoulement fluide et par conséquent la direction et l'intensité des charges aérodynamiques : ce cycle des interactions entre le fluide et la structure est la caractéristique du couplage fort entre les deux milieux, i.e. chaque milieu influe significativement sur l'autre. Basiquement, on retiendra que l'analyse acoustique fournit un chargement (la pression acoustique) à l'analyse structurelle, et l'analyse structurelle fournit des accélérations à l'analyse acoustique.

Si l'on ne peut pas découpler les équations de chaque milieu (c'est parfois possible et ça vaut donc le coût d'essayer), alors on essaye généralement de procéder par la méthode dite « décalée » : chaque physique est traitée séparément, et l'on essaye ensuite de faire transiter l'information d'un code à l'autre, ce qui n'est pas toujours simple, engendre des erreurs numériques supplémentaires, et surtout nécessite de développer des algorithmes de couplage. Sinon, la méthode dite « monolithique » est théoriquement la plus optimale puisque le fluide et la structure sont traités par un même code de calcul. Toutefois, lorsque les géométries ou la physique du problème à traiter deviennent complexes ce type de méthode n'est plus envisageable puisque chaque milieu (fluide ou solide) nécessite des procédures de calcul numérique spécifiques. Notons que certains codes de calcul gèrent de manière assez simple pour l'utilisateur ce couplage entre différentes physiques.

- La continuité de la pression à l'interface entre deux milieux peut être à entrer. Il suffit de

1. Dans le cas d'étude aérodynamiques et pour les corps bien profilés, la surface de référence sera la projection de la surface dans un plan orthogonal à la direction selon laquelle on veut le résultat.

Toujours dans ce cas, on décompose la force aérodynamique selon trois axes en une force de traînée F_x parallèle à la direction moyenne de l'écoulement ; une force de dérive F_y perpendiculaire à la direction moyenne de l'écoulement, dans le plan horizontal ; et une force de portance F_z , perpendiculaire à la direction moyenne de l'écoulement, dans le plan vertical.

spécifier que le saut est nul :

$$-n \cdot \left[\left(-\frac{1}{\rho} (\nabla p - q_d) \right)_1 - \left(-\frac{1}{\rho} (\nabla p - q_d) \right)_2 \right] = 0 \quad (19.36)$$

19.2.3 Convergence

La convergence d'un calcul acoustique dépend non seulement de la taille du maillage, mais également de la taille du pas de temps.

Discrétisation en espace

Le maillage doit être en mesure de prendre en compte les longueurs d'ondes qui nous intéressent dans le calcul. Les ondes doivent pouvoir « se développer » dans le calcul, et pour cela, on utilisera la condition simple suivante :

$$h = \frac{k\lambda}{10} \quad (19.37)$$

où h est la taille de la maille, et k le degré de l'approximation utilisée dans l'élément. Ainsi $h = \lambda/10$ pour un élément linéaire, et $h = \lambda/5$ pour un élément quadratique.

Discrétisation en temps

La condition de Courant, Friedrichs et Lewy (condition CFL), énoncée dans leur article de 1928 [29], définit, au travers d'un nombre adimensionnel appelée nombre de Courant C , une condition de convergence pour résoudre certaines équations aux dérivées partielles, notamment les équations aux dérivées partielles hyperboliques, utilisant la méthode des différences finies. Pratiquement, il sert à donner le seuil dimensionnel sous lequel on observe une instabilité de calcul, erreur d'approximation dans des calculs numériques, grandissant rapidement au fur et à mesure des calculs. Si la dimension de la grille est inférieure à la distance parcourue dans l'intervalle de pas de temps par l'onde la plus rapide que permet l'équation, l'erreur grandit et envahit la solution physique.

Le nombre de courant C , pour un problème de dimension n en espace (et 1 en temps), est défini par :

$$C = \Delta t \sum_{i=1}^n \frac{u_{x_i}}{\Delta x_i} \leq C_{max}. \quad (19.38)$$

où u est la fonction inconnue du problème, Δt le pas de temps et Δx_i le pas de discrétisation de chaque variable spatiale.

La constante C_{max} prend différentes valeurs en fonction de la méthode utilisée pour résoudre l'équation discrétisée. Pour une méthode explicite, on utilisera $C_{max} = 1$, alors que pour une méthode implicite, moins sensible aux instabilités numériques, on pourra prendre une valeur plus grande pour C_{max} . On fera en sorte de prendre C_{max} le plus grand possible pour la méthode considérée, ce qui permettra par conséquent de calculer le pas de temps le plus grand possible (la taille de la discrétisation en espace étant contrainte par les fréquences visées) afin de minimiser les calculs.

19.2.4 Vers l'infini...

Lorsque l'on fait un calcul en champ ouvert... il faudrait définir le domaine jusqu'à l'infini. Généralement, on définit une zone (par exemple sphérique ou hémisphérique) entourant le problème, située à une distance suffisante, et permettant de simuler l'infini : ce qui nous intéresse est donc une

condition d'absorption parfaite. Nous avons déjà mentionné la condition de Dirichlet. Toutefois, elle n'est pas optimale.

Décrite pour la première fois en 1994 par Jean-Pierre Bérenger dans *A perfectly matched layer for the absorption of electromagnetic waves*, une zone absorbante parfaitement adaptée (en anglais Perfectly matched layer ou PML) est une zone absorbante artificielle pour les équations d'ondes, couramment utilisée pour tronquer les domaines de calcul dans les méthodes numériques de simulation de problèmes à frontières ouvertes, particulièrement dans les méthodes de différences finis en temps et d'éléments finis. La propriété essentielle d'une PML qui la distingue d'un matériau absorbant ordinaire est le fait qu'elle est conçue de telle sorte que les ondes incidentes l'atteignant depuis un matériau non PML ne se réfléchissent pas à l'interface. Cette propriété permet aux PML d'absorber fortement toutes les ondes sortant d'un domaine de calcul sans les renvoyer dans ce domaine.

Cette approche a été modifiée puis généralisée en ce que l'on appelle la stretched-coordinate PML. Les PML correspondent à une transformation dans laquelle une ou plusieurs coordonnées sont attachées à des nombres complexes ; plus techniquement, il s'agit en réalité d'une prolongation analytique de l'équation d'onde dans le domaine complexe, où les ondes propagatives (oscillantes) sont remplacées par des ondes dont l'amplitude décroît exponentiellement. Ce point de vue permet aux PML d'être adaptées aux matériaux inhomogènes comme les guides d'ondes, ainsi qu'à d'autres systèmes de coordonnées et d'équations d'ondes.

La PML amortit une certaine longueur d'onde existant dans le système, déduite de la fréquence et d'une célérité de référence des ondes. Dans le cas d'un problème uniquement fluide, cette célérité de référence sera prise égale à la vitesse du son. Dans le cas d'ondes se propageant dans un solide, on la prendra égale à la vitesse de compression du son. Pour un modèle couplé fluide-structure, on pourra prendre une valeur moyenne entre les deux précédentes comme estimation initiale. Pour une onde dont le vecteur d'onde fait un angle θ avec la direction dans laquelle on veut absorber, il faudrait prendre une valeur de $1/\cos|\theta|$.

19.2.5 ... et au-delà

Si l'on souhaite savoir ce qu'il se passe loin de la source, en dehors du domaine de simulation (sans devoir tout mailler), alors il est possible d'utiliser des techniques dites de champ lointain (Far Field).

L'idée est somme toute assez simple : disposant des informations (pression, directivité) sur la surface d'un volume donné dans lequel est décrit le problème, il est possible d'extrapoler le comportement en dehors de ce volume sous certaines hypothèses, notamment de non réflexion du champ (par exemple champ libre). La mise en œuvre, quant à elle, est loin d'être aussi simple que l'idée.

Certains logiciels de calcul offrent cette possibilité.

19.2.6 Post-Traitement

La pression $p(x, y, z)$ calculée est une pression complexe (i.e. appartenant à \mathbb{C}). Toutefois, il est aisé, à partir de celle-ci, de définir un certain nombre d'autres valeurs d'intérêt :

- $p(x, y, z, t) = \Re(p(x, y, z)e^{i\omega t})$;
- $p_{rms}(x, y, z) = \frac{1}{\sqrt{2}}|p(x, y, z)|$;
- $SPL(x, y, z) = 20 \log_{10} \left(\frac{p_{rms}(x, y, z)}{p_{ref}} \right)$;
- d'autres grandeurs liées au modèle. Souvent on définira, à partir des données et des résultats le facteur de perte. Par exemple, on peut le définir par :

$$D_{tl} = 10 \log_{10} \left(\frac{W_i}{W_t} \right) \quad (19.39)$$

où W_i est la puissance moyenne incidente et W_t la puissance moyenne transmise. Définir toutes ces quantités ne nécessite pas de refaire le calcul éléments finis. Elles ont donc un coût de calcul quasiment nul ! Il ne faut donc surtout pas se priver de définir les quantités les plus parlantes pour le problèmes considéré.

19.3 Quelques illustrations avec FREEFEM++

19.3.1 Un exemple en acoustique des salles

L'acoustique des salles d'une manière générale peut se passer de couplage. Seul le volume de la salle est à discrétiser. Les zones d'absorption situées en parois seront modélisées par leur impédance, ainsi que les pertes par transmission. Dans le cas d'un locuteur, la source sera ponctuelle.

La figure 19.7 présente une salle et sa modélisation. La salle est constitué de deux murs épais

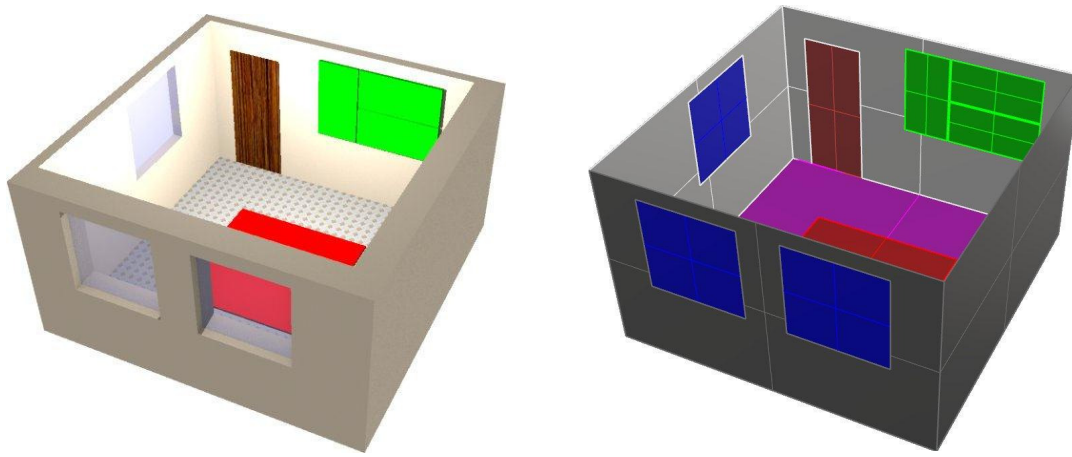


FIGURE 19.7: Un problème d'acoustique des salles

en béton et de deux murs fins en placo. Elle comporte trois fenêtres et une porte. Le sol est en carrelage, avec au centre un tapis épais. Le plafond non représenté est en béton. Au mur, à côté de la porte, se trouvent trois panneaux acoustiques.

Pour effectuer la modélisation, on ne contentera de représenter les surfaces internes de la pièce. Les murs épais (en gris foncé), le sol et le plafond seront considérés comme parfaitement réfléchissants. Pour les murs fins, on entrera une impédance correspondant à la perte par transmission. On fera de même avec les fenêtres et la porte. Les conditions d'absorption des panneaux acoustiques et de la moquettes pourront être modélisées en entrant l'impédance associée ou en utilisant un modèle d'absorption de type Biot-Allard si cela est disponible dans le logiciel.

On maillera le volume d'air intérieur défini par ces surfaces portant les conditions aux limites. On n'oubliera pas d'ajouter la source : ici un locuteur sera représenté par un monopôle, i.e. une source ponctuelle omnidirectionnelle. On lancera l'analyse en fréquences et l'on s'intéressera par exemple à la répartition spatiale de la pression acoustique à différentes fréquences choisies.

19.3.2 Un silencieux automobile

La figure 19.8 présente un silencieux d'automobile. Ce silencieux ne sera défini que par les surfaces représentées et le volume d'air qu'elles contiennent. On appliquera aux parois en rouge soit une condition de mur rigide (réflexion parfaite), soit une impédance correspondant à l'absorption due à la présence d'un matériau fibreux (laine de verre). Les surfaces grises seront également considérées comme des surfaces parfaitement réfléchissantes. Les deux sections en bleu (on n'en voit qu'une, l'autre étant à l'autre bout) correspondront à l'entrée de l'onde (onde plane) dans le silencieux, et à sa sortie.

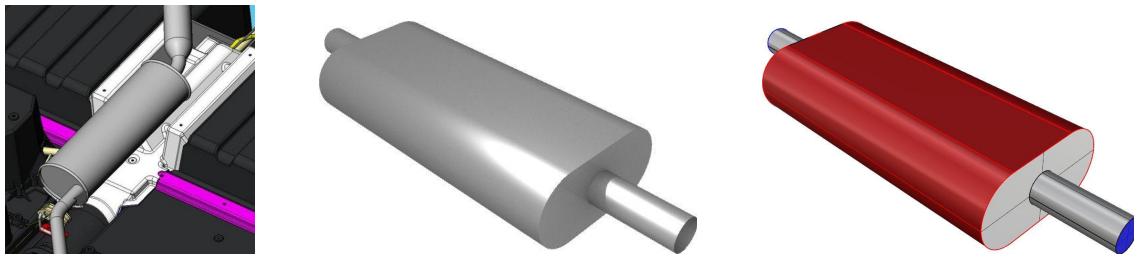


FIGURE 19.8: Modélisation d'un silencieux

On s'intéressera au facteur de perte, tel que défini par la relation (19.39). Les résultats avec et sans absorbant sont illustrés sur la figure 19.9, pour des fréquences allant de 25 à 1500 Hz par pas de 25 Hz. L'influence de l'ajout d'un absorbant est très nette sur ce graphique.

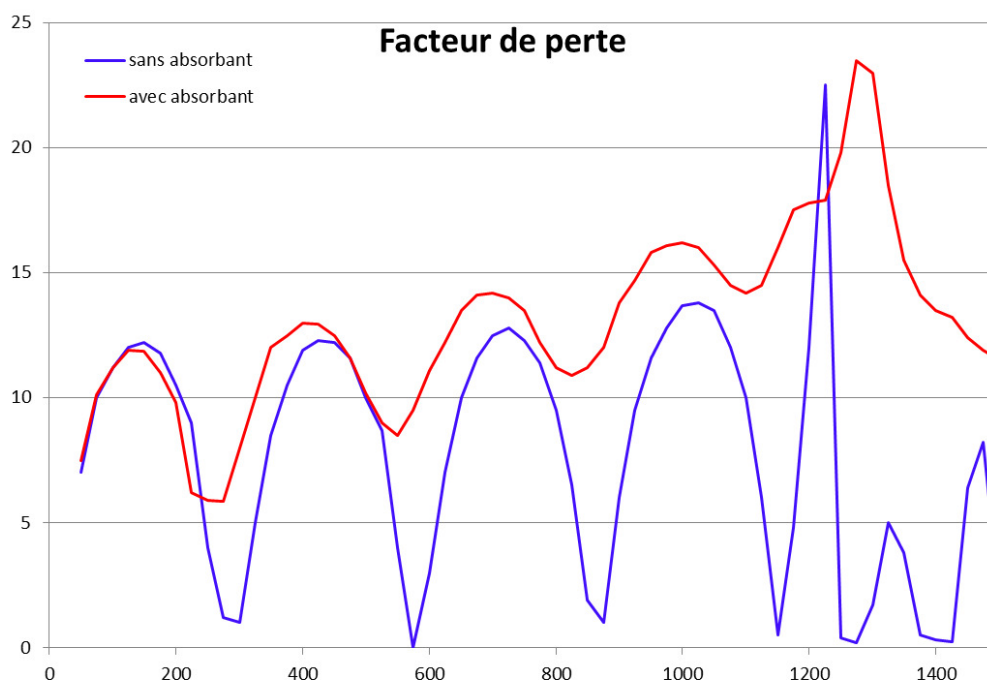


FIGURE 19.9: Facteur de perte pour le silencieux

19.3.3 Deux mots de statistiques

Nous reprenons ici une remarque faite au paragraphe 13.6.4. Considérons la figure 19.9 : elle représente la variation d'une réponse Y (ici le facteur de perte) en fonction d'une entrée X (ici la fréquence).

Supposons que nous nous intéressions maintenant à ce silencieux sollicité à une fréquence proche de 1000 Hz. On va supposer que la fréquence suit une loi normale centrée sur 1000 Hz et d'écart-type 25 Hz. Évidemment, avec les calculs menés précédemment, nous disposons de peu de points. On va même considérer disposer des points donnés au tableau 19.1. L'application de la formule (13.19) sur ces quelques points conduit à une première approximation de la moyenne du facteur de perte de 13.38 sans absorbant et 16.06 avec absorbant.

Évidemment, c'est une première approximation très grossière, qui peut être améliorée. Pour cela, il nous faudrait disposer de plus de points, aussi bien dans le domaine considéré qu'en dehors. Si l'on ne peut relancer de calcul, on peut essayer d'approximer la courbe de la réponse afin de

Fréquence	950	975	1000	1025	1050
Sans absorbant	11.6	12.8	13.7	13.8	13.5
Avec absorbant	15.8	16.1	16.2	16.0	15.3
poids 1000. p_i	2.16	9.68	15.96	9.68	2.16

Tableau 19.1: Points disponibles pour l'approximation de la moyenne

pouvoir interpoler et extrapoler des points.

On peut utiliser tout type d'interpolation (voir chapitre A) comme par exemple un polynôme de Lagrange... ici nous utiliserons une courbe de tendance polynomiale de degré 2. Dans le cas sans absorbant, celle-ci est donnée par l'équation suivante : $y = -0,0004342857x^2 + 0,8877714286x - 439,8628571429$. De là, on peut générer quelques points supplémentaires. Vu la forme de la courbe de réponse, on sait que les valeurs interpolées seront correctes, ce n'est pas le cas pour les valeurs extrapolées, surtout loin de la zone ayant servi à générer cette courbe de tendance. Par exemple on trouve 12.6 pour 1075 Hz au lieu de 12.0, 11.2 pour 1100 Hz au lieu de 10.0. Néanmoins, en fonction de la forme de la courbe, cela peut permettre une amélioration sensible de l'approximation de la moyenne de la réponse.

Pour une analyse plus complète dans le cadre stochastique, c'est au chapitre 22 que l'on se reportera. Toutes les méthodes qui y sont décrites sont applicables en fréquence.

19.3.4 Sur les conditions aux limites

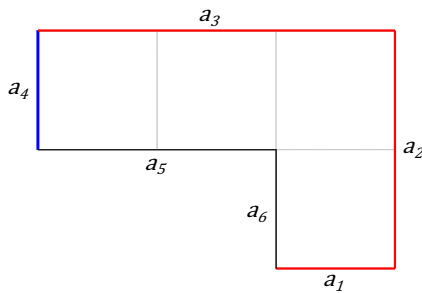


FIGURE 19.10: Acoustique et obstacle

Nous nous proposons de retrouver l'illustration 19.4d.

Pour cela, nous considérons le modèle décrit à la figure 19.10 constitué de 6 segments a_1 à a_6 . L'onde acoustique incidente arrive par le segment a_4 . L'onde est libre de se propager, et notamment elle n'est pas réfléchiée sur les côtés a_1 à a_3 , l'obstacle ne correspondant qu'aux faces a_5 et a_6 .

Pour modéliser l'onde arrivant dans le modèle, on va supposer que la surface a_4 porte une condition de Dirichlet de type $p = \bar{p}$. Nous proposons, dans un premier temps, d'imposer sur les faces a_1 à a_3 une condition de Dirichlet de type $p = 0$. Nous utiliserons

FREEFEM++ pour cet exemple.

Dans un premier temps, nous allons entrer directement la formulation faible définie à l'aide des formes bilinéaire (8.62) et linéaire (8.62) que nous rappelons :

$$a(p, q) = \int_{\Omega} \partial_{x_i} p \partial_{x_i} \tilde{q} - k^2 p \tilde{q} + \int_{\Gamma_R} i \rho c k A_n \tilde{q} \quad \text{et} \quad b(q) = - \int_{\Gamma_N} i \rho c k \bar{V}_n q \quad (19.40)$$

La description de la géométrie et le maillage sont donnés par :

```
1 real freq=200; // frequence en Hz
2 real c=343; // vitesse du son en m/s
3 func g=1; // on a p=g sur a4 => onde plane, par condition de Dirichlet
4 real k=freq/c;
5
6 border a1(t=0,1) { x= 2+t; y=0;}
7 border a2(t=0,1) { x=3; y= 2*t;}
8 border a3(t=0,1) { x= 3-3*t; y=2 ;}
9 border a4(t=0,1) { x= 0; y= 2-t ;}
10 border a5(t=0,1) { x= 2*t; y= 1 ;}
11 border a6(t=0,1) { x= 2; y=1-t ;}
12
13 mesh Th=buildmesh( a1(20) + a2(40) + a3(60) + a4(20) + a5(40) + a6(20));
```

Écrivons maintenant brutalement notre formulation faible :

```
14 fespace Vh(Th, P2);
15 Vh<complex> p, q;
16
17 solve helmholtz(p,q)= int2d(Th)(dx(p)*dx(conj(q)) + dy(p)*dy(conj(q)) - k*k*p*
    conj(q))
18 -int1d(Th,a4)(g*q) + on(a1,p=0) + on(a2,p=0)+ on(a3,p=0);
19
20 Vh pR=real(p);
21 plot (pR, wait=1, fill=1, ps="AcouDiff3.eps");
```

Le résultat est donné à la figure 19.11 pour plusieurs valeurs de la fréquence.

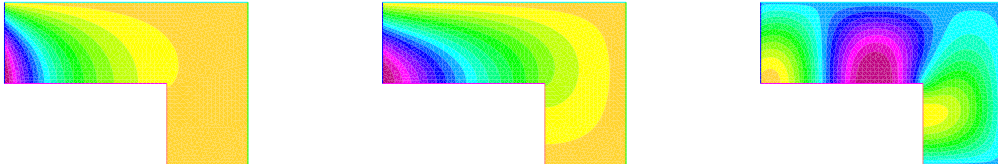


FIGURE 19.11: Acoustique au coin de la rue pour des fréquences de 100, 500 et 1000 Hz

Y a-t-il un intérêt à demander une approximation P2 ?

Y a-t-il un intérêt à demander une approximation P2 ?

Les conditions aux limites sont-elles satisfaisantes ?

On force seulement la pression à être nulle au bord, on n'a pas de condition aux limites de type « dérivée normale ».

Modifions le script précédent, et ne passons pas en complexes :

```
14 fespace Vh(Th, P1);
15 Vh p, q;
16
17 solve helmholtz(p,q)= int2d(Th)(dx(p)*dx(q) + dy(p)*dy(q) - k*k*p*q)
18 -int1d(Th,a4)(g*q) + on(a1,p=0) + on(a2,p=0)+ on(a3,p=0);
19
20 string legende="frequence = "+freq+" Hz";
21 plot (p, wait=true, cmm=legende, fill=1, value=true);
```

Y a-t-il une différence ?

Y a-t-il une différence ?

Et si on entrait une condition de Neumann ? Là, il est nécessaire d'être en complexe. L'unité complexe sous FREEFEM++ est i .

```
14 fespace Vh(Th, P2);
15 Vh<complex> p, q;
16
17 solve helmholtz(p,q)= int2d(Th)(p*q*k*k - dx(p)*dx(q) - dy(p)*dy(q))
18 -int1d(Th,a4)(g*q) -int1d(Th,a1)(1*i*c*k*q) -int1d(Th,a2)(1*i*c*k*q) -int1d(Th,a3)(1
    i*c*k*q);
19
20 string legende="frequence = "+freq+" Hz";
21 Vh pR=real(p);
22 plot (pR, wait=true, cmm=legende, fill=1, value=true);
```

Cette fois la réponse est nécessairement différente, puisque les conditions aux limites ne sont pas les mêmes... les résultats correspondants sont à la figure 19.12.

À quoi correspondent ces résultats ?

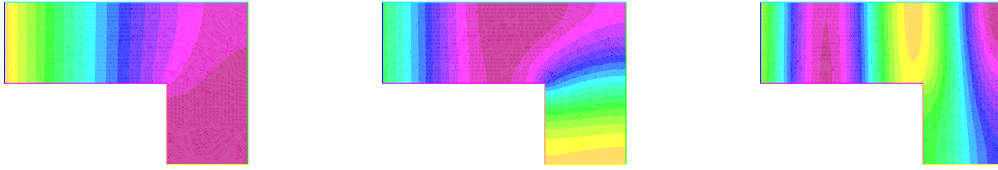


FIGURE 19.12: Avec condition de Neumann et pour des fréquences de 100, 500 et 1000 Hz

Chaque partie vibre à une vitesse telle que $\nabla \cdot \mathbf{u} = 0$.

Sauriez-vous introduire des conditions de Robin ?

Évidemment ! Il suffit de changer $\mathbf{u} \cdot \mathbf{n}$ en $\mathbf{u} \cdot \mathbf{n} + \alpha \mathbf{u}$ où α est l'impédance.

Permettraient-elles de laisser « fuir » l'onde à l'infini ?

Oui. Pour cela il faut que l'impédance de surface soit égale à celle du milieu, i.e. ici $\alpha = \rho c$. En terme de codage, cela correspond à $\mathbf{u} \cdot \mathbf{n} + \alpha \mathbf{u}$.

Est-on assez loin des conditions aux limites ?

Non. Voir figure 19.13.

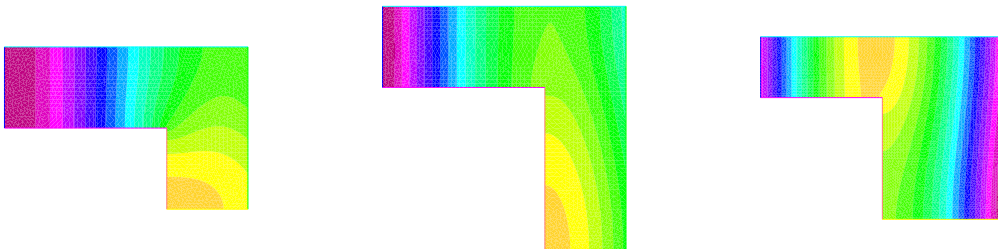


FIGURE 19.13: Distance aux conditions aux limites

19.3.5 Un peu d'amortissement

Introduisons maintenant un peu d'amortissement sur le cas précédent. L'équation des ondes avec amortissement est donnée par la relation (19.24). Il suffit donc d'ajouter à la forme bilinéaire $a(p, q)$ de l'équation (19.40) le terme :

$$+ \int_{\Omega} i d \omega p \tilde{q} \quad (19.41)$$

Attention, il ne faut pas que le facteur d'amortissement d soit trop grand... sous peine que l'onde ne se propage pas bien loin. Cela se traduit par la modification du batch FREEFEM++ :

```
14 fespace Vh(Th, P2);
15 Vh<complex> p, q;
16 real d=0.05;
17
18 solve helmholtz(p,q)= int2d(Th)(dx(p)*dx(conj(q)) + dy(p)*dy(conj(q)) - k*k*p*
    conj(q)
19 +amort*1i*freq*p*conj(q)) - int1d(Th,a4)(g*q);
20
```



```

21 Vh pR=real(p), pI=imag(p);
22
23 string legende="frequence = "+freq+" Hz et amortissement = "+d;
24 plot (pR, cmm=legende, wait=true, fill=1);
25 plot (pI, cmm=legende, wait=true, fill=1);

```

La différence entre calcul dans \mathbb{R} et calcul dans \mathbb{C} prend ici tout son sens. Jusqu'à présent, la partie imaginaire du champ de pression était nulle en tout point, ce qui n'est désormais plus le cas. La figure 19.17 présente les parties réelle (en haut) et imaginaire (en bas) pour un amortissement de 0, 1 et 2%. On remarque qu'avec seulement 1% d'amortissement, l'onde arrive péniblement

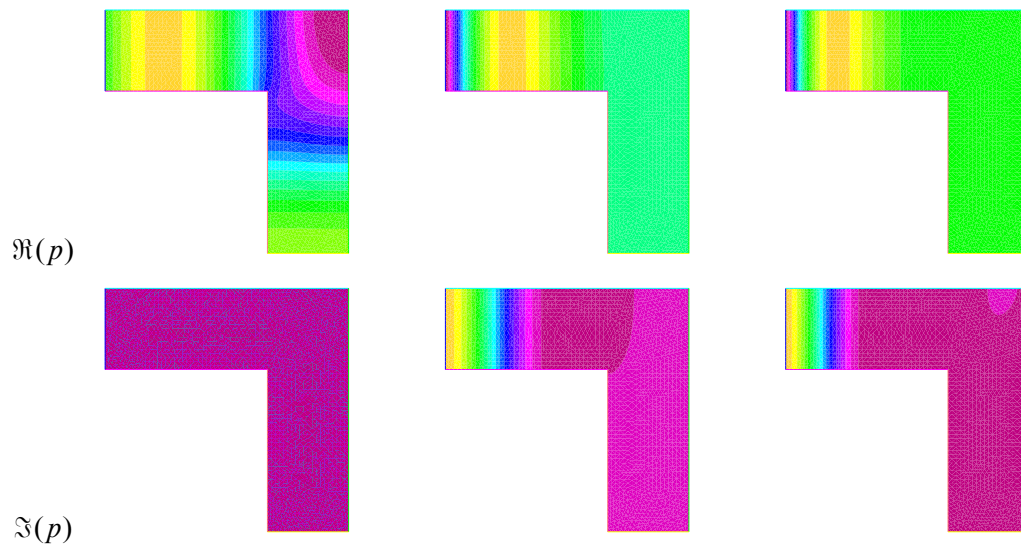


FIGURE 19.14: À 500 Hz avec un amortissement de 0, 1 et 2%

jusqu'au coude. Pour 2% d'amortissement, elle « s'évanouit » très vite, dès la partie droite.

19.3.6 Un obstacle

Pour le plaisir, introduisons un obstacle. Nous proposons le listing FREEFEM++ (abondamment commenté) suivant :

```

1 real freq=50.0; // frequence en Hz
2 real rho=1.; // densite constante
3 real c=343.0; // vitesse du son en m/s
4 func g=1.0; // on a p=g sur a4 => CL due a Green
5 real k=freq/c; // nombre d'onde
6 real amort=0.02; // 2% d'amortissement
7
8 border a1(t=0,1) { x= 3*t; y=0;}
9 border a2(t=0,1) { x=3; y= 2*t;}
10 border a3(t=0,1) { x= 3-3*t; y=2 ;}
11 border a4(t=0,1) { x= 0; y= 2-2*t ;}
12 border b1(t=0,1) { x= 1+t; y= 0.5;}
13 border b2(t=0,1) { x= 2-0.5*t; y= 0.5+t;}
14 border b3(t=0,1) { x= 1.5; y= 1.5-0.5*t;}
15 border b4(t=0,1) { x= 1.5-0.5*t; y= 1-0.5*t;}
16 mesh Th=buildmesh( a1(60) + a2(40) + a3(60) + a4(40)+ b1(-5) + b2(-5) + b3(-5) +
    b4(-5));
17
18 fespace Vh(Th, P1);
19 Vh<complex> p, q;
20
21 solve helmholtz(p,q)= int2d(Th)(-dx(p)*dx(conj(q)) - dy(p)*dy(conj(q)) +k*k*p*
    conj(q) // forme bilin non amortie
22 -amort*1i*freq*rho*p*conj(q) ) // terme d'amortissement
23 -int1d(Th,a4)(g*conj(q)) // CL paroi rayonnante (th de Green)

```

```

24 //// la CL de paroi rayonnante devrait pouvoir etre remplacee par une CL de
   Dirichlet relevee :
25 // +on (a4, p=1)// CL Dirichlet relevee = paroi rayonnante .
26 //// Puis on peut ajouter de l'absorption parfaite
27 // + on(a3, p=0)//CL Dirichlet absorption parfaite
28 //// l'impedance du milieu est Z1=rho.c.
29 //// Si l'on que l'onde quite le modele, il faut Z2=Z1 sur la CL de Robin:
30 //// La CL de Robin: +int1d (1i*rho*c*k*A2*conj(q) devient: +int1d (1i*k*conj(q)
31 -int1d(Th,a2)(1i*k*conj(q)) -int1d(Th,a1)(1i*k*conj(q))-int1d(Th,a3)(1i*k*conj(q))// CL Robin milieu infini
32 //// et les CL de Neumann s'ecrivent au besoin (il faut definir Vn)
33 // -int1d(Th,a2)(1i*rho*c*k*Vn*conj(q))-int1d(Th,a1)(1i*rho*c*k*Vn*conj(q))-
   int1d(Th,a3)(1i*rho*c*k*Vn*conj(q))// CL Neumann
34 ;
35
36 Vh pR=real(p), pI=imag(p);
37
38 string legende="frequence = "+freq+" Hz, amortissement = "+amort*100+"%";
39 plot (pR, wait=true, cmm=legende, fill=1, value=true);
40 plot (pI, wait=true, cmm=legende, fill=1, value=true);

```



FIGURE 19.15: $\Re(p)$ et $\Im(p)$ à 50 Hz avec un obstacle et 2% d'amortissement

Évidemment, des conditions aux limites peuvent (doivent) être introduites sur l'obstacle. On peut ajouter une condition de Robin d'impédance Z_n :

```

34 -int1d(Th,b3)(1i*rho*c*k/Zn*conj(q))-int1d(Th,b4)(1i*rho*c*k/Zn*conj(q))// CL
   Robin

```

ou une condition d'absorption parfaite de Dirichlet :

```

34 + on(b3, p=0)//CL Dirichlet absorption parfaite

```

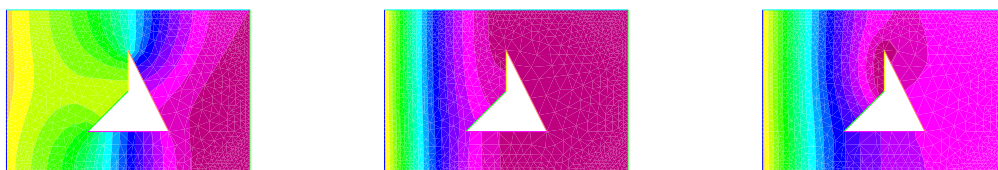


FIGURE 19.16: $\Re(p)$ pour $Z_n = 5$, $Z_n = 5000$ et pour la condition de Dirichlet $p = 0$

19.3.7 Transmission entre deux milieux

Considérons maintenant, toujours pour le plaisir, le cas d'une onde passant d'un milieu dans un autre. Nous proposons un passage d'une onde de l'air dans l'eau, modélisé en utilisant le script FREEFEM++ suivant :

```

1 real freq=50.0; // frequence en Hz
2 real rho1=1.204; // densite de l'air en kg/m3
3 real rho2=1000.0; // densite de l'eau en kg/m3

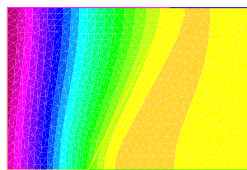
```



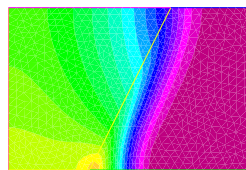
```

4 real c1=343.0; // vitesse du son dans l'air en m/s
5 real c2=1500.0; // vitesse du son dans l'eau en m/s
6 func g=1.0; // on a p=g sur a6 => onde plane, par condition de Dirichlet
7 real amort=0.0005;
8
9 border a1(t=0,1){x=t;y=0;};
10 border a2(t=1,3){x=t;y=0;};
11 border a3(t=0,2){x=3;y=t;};
12 border a4(t=3,2){x=t;y=2;};
13 border a5(t=2,0){x=t;y=2;};
14 border a6(t=2,0){x=0;y=t;};
15 border interf(t=0,1){x=1+t;y=2*t;}; // interface interne
16 mesh Th = buildmesh (a1(10)+a2(20)+a3(20)+a4(10)+a5(20)+a6(20)+interf(30));
17 //plot(Th, wait=1, cmm="maillage"); // affichage du maillage
18
19 fespace ValRho(Th, P0); // espace d'interpolation valeurs constantes par element
20 ValRho reg=region;
21 //plot(reg, fill=1, wait=1, cmm="regions", value=1); // affichage des regions
22 int nmil1=reg(0.1,0.1); // valeur prise pour la region 1 = 0
23 int nmil2=reg(2.9,0.1); // valeur prise pour la region 2 = 3
24 cout << " n milieu 1 = " << nmil1 << ", n milieu 2 = " << nmil2 << endl;
25 ValRho rho=rho1*(region==nmil1)+rho2*(region==nmil2);
26 ValRho c=c1*(region==nmil1)+c2*(region==nmil2);
27 //plot(rho, fill=1, wait=1, cmm="rho", value=1); // affichage des valeurs de rho
28 //plot(c, fill=1, wait=1, cmm="c", value=1); // affichage des valeurs de c
29 //cout << " rho(0.1,0.1)" << rho(0.1,0.1) << ", rho(2.9,0.1) = " << rho(2.9,0.1)
    << endl;
30 //cout << " rho(0.9,0.1)" << rho(0.9,0.1) << ", rho(1.9,0.1) = " << rho(1.9,0.1)
    << endl;
31 //cout << " c(0.1,0.1)" << c(0.1,0.1) << ", c(2.9,0.1) = " << c(2.9,0.1) << endl;
32 //cout << " c(0.9,0.1)" << c(0.9,0.1) << ", c(1.9,0.1) = " << c(1.9,0.1) << endl;
33
34 fespace Vh(Th, P1);
35 Vh<complex> p, q;
36
37 solve helmholtz(p,q)= int2d(Th)(-dx(p)*dx(conj(q)) - dy(p)*dy(conj(q)) +freq/c*
    freq/c*p*conj(q)) // forme bilin non amortie
38 -amort*1i*freq*rho*p*conj(q)) // terme d'amortissement
39 +int1d(Th,a6)(g*conj(q)) // CL paroi rayonnante (th de Green)
40 ;
41
42 Vh pR=real(p), pI=imag(p);
43
44 string legende="frequence = "+freq+" Hz, amortissement = "+amort*100+"%";
45 plot (pR, wait=true, cmm=legende, fill=1, value=true);
46 plot (pI, wait=true, cmm=legende, fill=1, value=true);

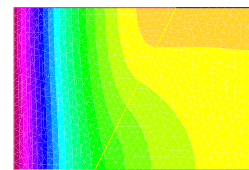
```



$\Re(p)$



$\Im(p)$



$\Re(p)$ sans amortissement

FIGURE 19.17: Passage d'une onde de l'air à l'eau avec et sans amortissement

19.3.8 Interaction fluide-structure semi-forte

Dans les cas déjà évoqués tels qu'une source acoustique puissante capable de déformer une plaque mince (tablier fin de véhicule) ou d'une vibration se propageant jusqu'à faire rayonner une surface, alors il est nécessaire de modéliser à la fois le domaine mécanique et le domaine acoustique.

Un couplage fort consiste à définir les deux volumes (fluide et solide), chacun avec sa formulation variationnelle, à coupler ces deux « physiques », par exemple par des multiplicateurs de Lagrange (le lecteur aura sans doute découvert mon inclination à utiliser lesdits multiplicateurs de Lagrange), et à résoudre le tout de manière simultanée.

Nous proposons une voie médiane, que nous appellerons couplage semi-fort, et dans lequel nous allons résoudre une série de problèmes quasiment découplés. Le cas que nous proposons de traiter est celui de l'interaction entre un domaine fluide, modélisé par l'équation d'Helmholtz en pression seule (comme dans les exemples précédents) et une structure ayant un comportement élastique linéaire isotrope. Ce calcul « semi-fort » se déroule comme suit :

1. On calcule la structure seule, soumise à ses efforts mécaniques divers (ici uniquement à son poids propre), en considérant que le fluide n'a pas encore exercé de pression. Cela nous donne une déformation de la structure, et particulièrement de la paroi d'interface avec le fluide ;
2. Cette déformation est utilisée pour modéliser le domaine fluide, dans lequel se trouve la source acoustique. On calcule la répartition de pression dans le domaine fluide, et on transforme la pression présente sur la paroi d'interface avec la structure en efforts.
3. Ces efforts supplémentaires sont introduits dans le calcul de la structure, qui se déforme à nouveau... et on boucle sur l'étape 2. Il convient de définir un critère d'arrêt. On peut considérer l'arrêt si, sur l'interface entre les deux domaines et entre deux pas successifs, la déformation n'évolue que de manière infime, ou dualement si les efforts sont quasiment constants.

Commençons par un petit rappel de mécanique, et pour cela, considérons tout d'abord le solide seul. Sa formulation variationnelle est donnée à l'équation (8.42) rappelée ici :

$$\int_{\Omega} \lambda \operatorname{div}(u) \operatorname{div}(v) + 2\mu \varepsilon(u) \cdot \varepsilon(v) = \int_{\Omega} f v + \int_{\Gamma_N} g_N v, \quad \forall v \in H_D^1(\Omega)$$

Le script FREEFEM++ ci-dessous résout le problème initial de la structure seule :

```

1 // tablier pour interaction fluide-structure semi-forte
2 // au lieu de définir les vecteurs [ux,uy], on définit par composantes
3 //
4 real E = 21.5; // Module d'Young
5 real nu = 0.29; // coefficient de Poisson
6 real gravity = -0.05;
7 int nmaille = 2; // au moins 70 pour un resultat assez lisse, 150 c'est mieux mais
  long
8 //
9 macro div(u,v) ( dx(u)+dy(v) ) // EOM
10 real sqrt2=sqrt(2.);
11 macro epsilon(u,v) [dx(u),dy(v),(dy(u)+dx(v))/sqrt2] // EOM
12 //
13 border st1(t=0,10) { x=0; y=t;};
14 border st2(t=0,-1) { x=t; y=10;};
15 border st3(t=10,0) { x=-1; y=t;};
16 border st4(t=-1,0) { x=t; y=0;};
17 mesh Th = buildmesh( st1(10*nmaille)+st2(nmaille)+st3(10*nmaille)+st4(nmaille));
18 //plot(Th);
19 //
20 real lambda = E*nu/(1+nu)/(1-2*nu); // coefficient de Lamé
21 real mu = E/2/(1+nu); // coefficient de Lamé
22 //
23 fespace Wh(Th,P1);
24 Wh ux,uy,w,s;
25 //
26 // forme faible :
27 solve StatiqueIsotrope(ux,uy,w,s)=
28   int2d(Th)( lambda*div(ux,uy)*div(w,s) +2.*mu*(epsilon(ux,uy)'*epsilon(w,s)
29     ) )
30   - int2d(Th) (gravity*s)
31   + on(st2,ux=0,uy=0) // tenu en haut
32   + on(st4,ux=0,uy=0) // tenu en bas
33 ;

```

Remarque. En complément à ce code, nous fournissons le calcul des déformations et contraintes, dont nous n'aurons pas besoin pour la suite, mais c'est cadeau...

Disposant des déformations et des contraintes, on peut trouver aisément les directions principales (il suffit de diagonaliser une matrice symétrique 2×2), puis calculer par exemple la contrainte de von Mises.

```

33 //
34 // POST-TRAITEMENT
35 //
36 // Deformee :
37 mesh Th1 = movemesh(Th, [x+ux, y+uy]);
38 //plot(Th1, wait=1);
39
40 //DEFORMATIONS
41 // Ici j'ai calcule composante par composante plutot que de generer un
   vecteur
42 // Sur le maillage non deforme: EpsXX, EpsYY, EpsXY
43 // Sur le maillage deforme: YEpsXX, YEpsYY, YEpsXY
44 fespace Xh(Th,P1);
45 Xh EpsXX=d $x$ (ux), EpsYY=d $y$ (uy), EpsXY=(d $y$ (ux)+d $x$ (uy))/sqrt2;
46 Xh SigXX=2*mu*EpsXX+lambda*(EpsXX+EpsYY);
47 Xh Sigyy=2*mu*EpsYY+lambda*(EpsXX+EpsYY);
48 Xh SygXY=2*mu*EpsXY;
49 //plot(EpsXX, fill=1, cmm="epsilon xx", wait=1);
50 fespace Yh(Th1,P1);
51 Yh YEpsXX=d $x$ (ux), YEpsYY=d $y$ (uy), YEpsXY=(d $y$ (ux)+d $x$ (uy))/sqrt2;
52 Yh YSigXX=2*mu*EpsXX+lambda*(EpsXX+EpsYY);
53 Yh YSigYY=2*mu*EpsYY+lambda*(EpsXX+EpsYY);
54 Yh YSigXY=2*mu*EpsXY;
55 //plot(YEpsXX, fill=1, cmm="epsilon xx", wait=1);
56 //plot(YSigXX, fill=1, cmm="sigma xx", wait=1);
57 //plot(YEpsYY, fill=1, cmm="epsilon yy", wait=1);
58 //plot(YSigYY, fill=1, cmm="sigma yy", wait=1);
59 //plot(YEpsXY, fill=1, cmm="epsilon xy", wait=1);
60 //plot(YSigXY, fill=1, cmm="sigma xy", wait=1);

```

Considérons maintenant le calcul du fluide seul. Le listing ressemble à ce qui a déjà été présenté précédemment. Deux points sont à considérer :

- le maillage à partir d'une donnée issue du calcul : connaissant la solution du problème structural, on dispose des déplacements ux et uy ... dont on se sert pour définir le domaine fluide :

```
border fl4(t=10,0) {x=ux(0,t); y=uy(0,t);}; // interface avec la structure
```

- le calcul des efforts dus à la pression.

Si l'on résout en vitesse et pression, alors, les efforts sont définis par :

```

sigmaxx([x+ux,y+uy]) = (2*d $x$ (ux)-p);
sigmayy([x+ux,y+uy]) = (2*d $y$ (uy)-p);
sigmaxy([x+ux,y+uy]) = (d $x$ (ux)+d $y$ (uy));

```

et on ajoute la force suivante dans la formulation variationnelle du problème de structure :

```
+ int1d(Th, interface)(sigma11*N.x*w + sigma22*N.y*s + sigma12*(N.y*w+N.x*s)
)
```

Si l'on résout en pression uniquement, on peut faire simple, et ajouter la force suivante dans la formulation variationnelle du problème de structure :

```
+ int1d(Th, interface)(pR*N.x*w + pR*N.y*s + (d $x$ (ux)+d $y$ (uy))*(N.y*w+N.x*s)
)
```

où l'on a écrit pR pour rappeler que l'on considère la partie réelle de la pression, qui peut être directement p si l'on a une solution monochromatique dans la partie fluide.

Cela conduit au script :

```

33 // On modelise le fluide par l'equation de Helmholtz
34 //
35 include "VM_StrFluid2.edp" // partie structure = statique lineaire isotrope
36 //
37 real freq=50.0; // frequence en Hz
38 real rho=1.;
39 real c=343.0; // vitesse du son en m/s

```

```

40 func g=1.0; // on a  $p=g$  sur  $a^4 \Rightarrow$  onde plane, par condition de Dirichlet
41 real k=freq/c;
42 real amort=0.0; // 2;
43 //
44 border fl1(t=0,10) {x=t; y=0;};
45 border fl2(t=0,10) {x=10; y=t;};
46 border fl3(t=10,0) {x=t; y=10;};
47 border fl4(t=10,0) {x=0; y=ux(0,t);}; // interface avec la structure
48 //
49 mesh sh = buildmesh(fl1(10*nmaille)+fl2(10*nmaille)+fl3(10*nmaille)+fl4(10*
    nmaille));
50 // plot(sh, wait=1);
51 //
52 fespace Ph(sh, P1);
53 Ph<complex> p, q;
54 //
55 problem helmholtz(p,q)= int2d(sh)(-dx(p)*dx(conj(q)) - dy(p)*dy(conj(q)) +k*k*p*
    conj(q) // forme bilin non amortie
56 -amort*1i*freq*rho*p*conj(q) ) // terme d'amortissement
57 -int1d(sh,fl2)(g*conj(q)) // CL paroi rayonnante (th de Green)
58 ;
59 solve helmholtz(p,q)
60 //
61 Ph pR=real(p), pI=imag(p);
62 plot (pR, wait=true, cmm="Re(p)", fill=1, value=true);
63 // plot (pI, wait=true, cmm="Im(p)", fill=1, value=true);

```

Il ne reste plus qu'à faire boucler le calcul. Ici nous n'avons pas introduit de critère de convergence et avons simplement effectué une boucle de trois itérations :

```

64 //
65 // boucle
66 //
67 int i;
68 for(i=0; i<3; i++) // boucle plutot qu'un while avec condition de sortie basee sur
    convergence
69 {
70 // Probleme avec efforts transmis:
71 problem StatiqueIsotropeCouple(ux,uy,w,s, init=i)=
72     int2d(Th)( lambda*div(ux,uy)*div(w,s) +2.*mu*(epsilon(ux,uy)''' // (pour
        impression LaTeX)
73     *epsilon(w,s)) )
74 - int2d(Th) (gravity*s)
75 + on(st2,ux=0,uy=0) // tenu en haut
76 + on(st4,ux=0,uy=0) // tenu en bas
77 - int1d(Th,interface)(p*N.x*w + p*N.y*s + (dx(ux)+dy(uy))*(N.y*w+N.x*s) )
78 ;
79 Th1 = movemesh(Th, [x+uu, y+vv]);
80 // plot(Th1);
81 mesh sh = buildmesh(fl1(10*nmaille)+fl2(10*nmaille)+fl3(10*nmaille)+fl4(10*
    nmaille));
82 // plot(sh);
83 solve helmholtz(p,q)
84 }
85 plot(Th1);

```

Chapitre 20

Les non linéarités

Résumé — Jusqu'à présent, nous avons évité, autant que faire se peut, d'aborder les problèmes de non linéarité...

Beaucoup de cas de non-linéarité s'imposent par la nature du problème à traiter : grands déplacements, loi de comportement choisie, contact... Ils sont donc facilement identifiables, et face à des tels cas, l'utilisateur sera par conséquent précautionneux et ne se laissera donc pas surprendre.

Toutefois, dans le cas de la dynamique, la non-linéarité existe de manière implicite, même si tout le reste est « linéaire » par ailleurs. Cela peut constituer un écueil si l'on n'en est pas conscient.

Plusieurs types de non-linéarités peuvent être considérés. En mécanique des structures, on distinguera :

les non-linéarités géométriques qui se manifestent dans les problèmes des grands déplacements, des grandes rotations et/ou de grandes déformations. La notion de « grands » déplacements signifie tout simplement que l'hypothèse des petites perturbations n'est plus vérifiable. Or celle-ci stipule que géométries déformée et initiale doivent rester relativement proches. La notion de grandes déformations, déjà mentionnée dans ce document, fait que la linéarité des relations entre déplacements et déformations n'est plus conservée.

les non-linéarités matérielles dues à la loi de comportement du solide (ou plus généralement à la loi de comportement dans le milieu Ω). Le plus souvent, cette loi peut s'exprimer sous la forme d'équations différentielles non-linéaires du premier ordre.

Nous avons déjà évoqué ce phénomène à plusieurs endroits dans ce document, et le chapitre 15 sur l'homogénéisation est une illustration du cas où l'on peut substituer un milieu homogénéisé simple à un milieu compliqué. Toutefois, nous irons un peu plus loin dans ce chapitre et présenterons les principales lois de comportement rencontrées en mécanique.

les non-linéarités liées à l'évolution des conditions aux limites. Ce type de non-linéarité apparaît en particulier dans les problèmes de contact et de frottement entre solides. Ces phénomènes sont décrits par des inéquations et des opérations de projection.

les non-linéarités liées aux instabilités du comportement qui se présentent dans l'analyse des problèmes dynamiques.

20.1 Tenseurs, décomposition des tenseurs

Le déviateur est un opérateur matriciel utilisé en mécanique des milieux continus, plus précisément en plasticité.

Soit \mathcal{M} une matrice (ou tenseur d'ordre 2) de dimension n . Le déviateur de \mathcal{M} , noté $\text{dev}\mathcal{M}$, vaut :

$$\text{dev}\mathcal{M} = \mathcal{M} - \frac{\text{tr}(\mathcal{M})}{n} \mathbf{I}_n \quad (20.1)$$

où $\text{tr}(\mathcal{M})$ est la trace de la matrice, i.e. la somme de ses termes diagonaux.

Le déviateur est un tenseur de trace nulle.

20.1.1 Tenseur des contraintes

Histoire

Le tenseur des contraintes, ou tenseur de Cauchy, n'est pas forcément introduit par la loi de Hooke généralisée qui le lie au tenseur des déformations ($\sigma = H\varepsilon$).

D'ailleurs, lorsque Cauchy l'introduit vers 1822, il le fait pour représenter les efforts intérieurs mis en jeu entre les portions déformées du milieu, via l'équilibre des efforts pour toute coupure dans un matériau (i.e. définition sous forme de forces surfaciques).

En tout point M il existe une infinité de facettes d'orientation différentes. Le théorème de Cauchy permet de définir l'état de contrainte sur une facette d'orientation quelconque à partir de la connaissance de l'état de contrainte selon trois directions différentes. L'énoncé de ce théorème est le suivant :

Théorème 64 — Théorème de Cauchy. [un des nombreux -] : Les composantes du vecteur contrainte en un point M sur une facette de normale \mathbf{n} dépendent linéairement des composantes de cette normale. Les coefficients linéaires sont les composantes du tenseur des contraintes.

Ce théorème conduit à formuler la contrainte s'exerçant sur une facette d'orientation quelconque comme :

$$\mathbf{T}(\mathbf{M}, \mathbf{n}) = \sum_{j=1}^3 n_j \mathbf{T}(\mathbf{M}, \mathbf{e}_j) \quad (20.2)$$

et, comme dans ce repère orthonormé ($\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$) chacune des trois contraintes de base a trois composantes, on a :

$$\mathbf{T}(\mathbf{M}, \mathbf{e}_j) = \sum_{i=1}^3 \sigma_{ij} \mathbf{e}_i \quad (20.3)$$

soit au final :

$$\mathbf{T}(\mathbf{M}, \mathbf{n}) = \sum_{i,j=1}^3 \sigma_{ij} n_j \mathbf{e}_i \quad (20.4)$$

Les coefficients linéaires σ_{ij} apparaissent donc comme les éléments d'un tenseur de rang 2 : il s'agit du tenseur des contraintes de Cauchy.

De manière encore plus explicite, on écrit :

$$\mathbf{T}(\mathbf{M}, \mathbf{n}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} = \sigma \mathbf{n} \quad (20.5)$$

D'un point de vue pratique, chacun des éléments σ_{ij} du tenseur des contraintes de Cauchy rend compte d'une contribution clairement identifiable : le premier indice i est l'indice de projection (direction selon laquelle s'exerce la contribution) ; le second indice j repère l'orientation de la surface sur laquelle s'exerce la contribution. Par exemple σ_{12} correspond à la composante suivant \mathbf{e}_1 de la contrainte qui s'exerce sur la facette de normale \mathbf{e}_2 .

En exploitant la condition d'équilibre appliquée au moment résultant, il est possible de démontrer que, en statique, le tenseur des contraintes est nécessairement symétrique. *D'ailleurs, en se servant de cette symétrie, on introduit la notation de Voigt ou notation de l'ingénieur. On pose $\sigma_1 = \sigma_{11}, \sigma_2 = \sigma_{22}, \sigma_3 = \sigma_{33}$, puis $\sigma_4 = \sigma_{23}, \sigma_5 = \sigma_{13}, \sigma_6 = \sigma_{12}$, et l'on peut présenter le tenseur des contraintes sous forme de vecteur : $\langle \sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6 \rangle$. Cela facilite l'écriture*

de la loi de Hooke généralisée et montre en même temps que l'espace des contraintes est un espace vectoriel à six dimensions.

Il existe (au moins) une base orthonormée dans laquelle le tenseur des contraintes est diagonal. Pour trouver ce repère, il faut résoudre le problème aux valeurs propres $\det(\sigma - \lambda I) = 0$. Les trois racines de ce polynôme de degré 3 sont les valeurs propres encore appelées contraintes principales σ_I , σ_{II} et σ_{III} . Les vecteurs propres associés sont les directions principales qui forment le repère principal. Dans la mesure où le tenseur des contraintes est symétrique, il existe bien trois valeurs propres réelles, et les vecteurs propres associés sont orthogonaux.

Par convention, on ordonnera toujours les contraintes principales de sorte que $\sigma_I > \sigma_{II} > \sigma_{III}$. Ces contraintes principales permettent de définir les invariants du tenseur des contraintes de Cauchy :

$$I_1 = \sigma_1 + \sigma_2 + \sigma_3 = \text{tr}(\sigma) \quad (20.6)$$

$$I_2 = \sigma_1\sigma_2 + \sigma_2\sigma_3 + \sigma_1\sigma_3 = \sigma_{11}\sigma_{22} + \sigma_{22}\sigma_{33} + \sigma_{33}\sigma_{11} - \sigma_{12}^2 - \sigma_{23}^2 - \sigma_{13}^2 \quad (20.7)$$

$$I_3 = \sigma_1\sigma_2\sigma_3 = \det(\sigma) \quad (20.8)$$

Comme pour les déformations, il est souvent utile (bien que la signification physique nous en échappe encore) de décomposer le tenseur des contraintes en partie sphérique et déviateur :

$$\sigma_{ij} = \frac{1}{3}\sigma_1\delta_{ij} + s_{ij} \quad (20.9)$$

ou encore :

$$\sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{12} & s_{22} & s_{23} \\ s_{13} & s_{23} & s_{33} \end{bmatrix} + \begin{bmatrix} p & 0 & 0 \\ 0 & p & 0 \\ 0 & 0 & p \end{bmatrix} \quad (20.10)$$

où $p = \frac{1}{3}\sigma_1\delta_{ij}$ est la partie sphérique (et $\sigma_1 = \sigma_{kk}$) et s_{ij} la partie déviatorique. La partie sphérique correspond à une pression isostatique, i.e. un vecteur contrainte normal à la facette pour toute direction (généralisation de la notion de pression hydrostatique dans les liquides). Physiquement, le déviateur du tenseur des contraintes correspond donc aux contributions des contraintes autres que surfaciques (cas $n = 3$) ou linéiques (cas $n = 2$). Le déviateur est un tenseur de trace nulle, i.e. $s_{ii} = 0$. Le déviateur a les mêmes directions principales que le tenseur des contraintes, on a alors dans le repère principal :

$$s_{ij} = \begin{bmatrix} s_1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & s_3 \end{bmatrix} \quad (20.11)$$

On peut définir les invariants du déviateur des contraintes de Cauchy :

$$\begin{cases} J_1 = s_1 + s_2 + s_3 = \text{tr}(s_{ij}) = 0 \\ J_2 = -s_1s_2 - s_2s_3 - s_1s_3 \\ J_3 = s_1s_2s_3 = \det(s_{ij}) \end{cases} \quad (20.12)$$

Ces invariants sont utiles pour définir la contrainte de comparaison, ou contrainte effective $\sigma_e = f(\sigma_{ij})$: cette valeur est ensuite comparée à la limite élastique pour savoir si l'on est dans le domaine élastique ou plastique. Le second invariant du déviateur des contraintes est la contrainte de von Mises. On appelle triaxialité des contraintes η le rapport entre la contrainte isostatique et la contrainte équivalente de von Mises :

$$\eta = \frac{p}{\sigma_{evm}} = \frac{\sigma_{ii}}{3\sigma_{evm}} \quad (20.13)$$

Ce paramètre est important dans l'étude de l'endommagement et de la mécanique de la rupture. Notons qu'il caractérise certains cas simples de sollicitation tels que le cisaillement pur ($\eta = 0$), ou la traction uniaxiale ($\eta = 1/3$). Lorsque l'on parle de contraintes, on se réfère toujours au tenseur de Cauchy. Toutefois, plusieurs autres mesures des contraintes ont été développées :

- Les tenseurs des contraintes de Piola-Kirchhoff : ils permettent d'exprimer les contraintes par rapport à une configuration de référence (alors que le tenseur des contraintes de Cauchy les exprime relativement à la configuration actuelle). Pour des déformations infinitésimales, les tenseurs de Piola-Kirchhoff et de Cauchy sont identiques.
On voit alors que, si le domaine étudié venait à varier, ces tenseurs seraient bien appropriés (voir paragraphe 20.2 par exemple)
- Le premier tenseur des contraintes de Piola-Kirchhoff P relie les forces dans la configuration actuelle au domaine (aires) dans une configuration de référence. Pour l'exprimer, nous aurons donc besoin du tenseur gradient de déformation F et de son jacobien J (voir ci-dessous).

$$P = J\sigma F^T \quad \text{de coordonnées} \quad P_{ij} = J\sigma_{ik} \frac{\partial u_j}{\partial x_k} \quad (20.14)$$

- Le second tenseur des contraintes de Piola-Kirchhoff S , de manière duale, relie les forces de la configuration de référence avec le domaine actuel. Avec les mêmes notations qu'au dessus, on a :

$$S = JF^{-1}\sigma F^{-T} \quad \text{de coordonnées} \quad S_{ij} = J \frac{\partial u_i}{\partial x_k} \frac{\partial u_j}{\partial x_m} \sigma_{km} \quad (20.15)$$

- Le tenseur de contraintes de Kirchhoff $\tau = J\sigma$: il est utilisé dans des algorithmes numériques en plasticité des métaux (où il n'y a pas de changement de volume pendant la déformation plastique).
- ...

Pour exprimer les tenseurs de Piola-Kirchhoff nous avons eu besoin du tenseur gradient de déformation F . Ce dernier est défini comme suit :

$$F_{ij}(X, t) = \frac{\partial x_i}{\partial X_j} = \delta_{ij} + \frac{\partial u_i}{\partial X_j} \quad \text{ou} \quad F = I + \nabla u \quad (20.16)$$

où X est la position de référence à l'instant t_0 (configuration Ω_0) et x , la position courante à l'instant t (configuration actuelle Ω). Localement les formules de transport s'écrivent (voir paragraphe 12.5.2) :

- Pour un vecteur : $dx = FdX$
- Pour un volume : $dv = JdV$ avec $J = J(F) = \det(F)$ le jacobien du tenseur gradient de déformation
- Pour une surface orientée : $ds = JF^{-1}dS$

Pour caractériser les changements de forme, on introduit le tenseur $C = F^T F$, symétrique, qui est le tenseur des dilatations ou encore le tenseur de Cauchy-Green droite. Le tenseur $E = (C - I)/2$, symétrique, est le tenseur des déformations de Green-Lagrange. Il a les mêmes directions principales que le tenseur C . Ces deux tenseurs sont lagrangiens¹ et ils opèrent sur des quantités définies sur la configuration de référence Ω_0 . Nous venons d'évoquer des tenseurs de déformation... c'est qu'il est temps de changer de paragraphe.

1. la description lagrangienne consiste à suivre dans le temps les particules le long de leurs trajectoires : c'est une description intuitive de leur mouvement. En représentation lagrangienne, la position d'un point M à l'instant t qui se trouvait en M_0 à l'instant t_0 est donnée par une relation du type $M = f(M_0, t)$. Cette méthode présente un inconvénient : le référentiel se déplace avec le fluide. Il est donc difficile de connaître l'état du fluide en un point donné de l'espace et du temps.

La description eulérienne décrit le champ de vitesses qui associe à chaque point un vecteur vitesse. La photographie avec un temps de pose assez court d'un écoulement muni de particules colorées permet de visualiser des éléments de ce champ de vitesses à un instant donné. Au contraire un temps de pose plus long permet de visualiser des trajectoires de la description lagrangienne. Le champ de vitesses est décrit en donnant à tout instant t le vecteur vitesse V en tout point M par une relation de type $V(M, t)$.

20.1.2 Tenseur des déformations

Le tenseur des déformations, ou tenseur de Green-Lagrange, est obtenu directement à partir des déplacements par la relation :

$$\varepsilon = \frac{1}{2} (\nabla u + (\nabla u)^T) \text{ soit pour chaque composante : } \varepsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} + \frac{\partial u_k}{\partial x_i} \frac{\partial u_k}{\partial x_j} \right) \quad (20.17)$$

Cette écriture provient directement du fait d'écrire l'accroissement de déplacement à partir du déplacement du point M en M' que l'on formule en fonction du déplacement du point M, et d'un accroissement de déplacement, caractérisant le fait que chaque point du solide est susceptible de subir un déplacement différent à l'origine de la déformation.

On peut alors distinguer deux grands types de contribution à la déformation en décomposant ce tenseur ε comme la somme d'un tenseur symétrique E et d'un tenseur antisymétrique G :

$$E = \frac{1}{2}(\varepsilon + \varepsilon^T) \quad \text{et} \quad G = \frac{1}{2}(\varepsilon - \varepsilon^T) \quad (20.18)$$

On obtient alors le tenseur des déformations pures :

$$E_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \quad (20.19)$$

et le tenseur des rotations pures :

$$G_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} - \frac{\partial u_j}{\partial x_i} \right) \quad (20.20)$$

(on note souvent γ_{ij} au lieu de G_{ij}). Lorsque l'on parle de tenseur des déformations, on fait souvent référence au tenseur linéarisé des déformations, obtenu en négligeant les termes d'ordre 2 du tenseur de Green-Lagrange, ou encore tenseur des déformations dans le cas des petits déplacements :

$$\varepsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \text{ que l'on note } \varepsilon_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i}) \quad (20.21)$$

En mécanique des milieux continus, le tenseur des déformations pour les petites déformations (ou tenseur de Green) est la partie symétrique de la matrice jacobienne du vecteur déplacement de chaque point du solide. Si l'on décompose le tenseur des déformations en une somme d'une partie déviatorique (ou déviateur) et d'une partie sphérique :

$$\varepsilon_{ij} = \frac{1}{3} \varepsilon_I \delta_{ij} + e_{ij} \quad (20.22)$$

avec $\varepsilon_I = \varepsilon_{kk}$ et $e_{ii} = 0$; on a une signification physique claire de chacun des termes. On vérifie en effet facilement que ε_I caractérise la variation de volume :

$$\frac{\Delta V}{V} = \varepsilon_I = \text{tr } \varepsilon \quad (20.23)$$

Il suffit, pour s'en convaincre, de partir d'un cube unité ($V = 1$) dans les axes principaux, et de calculer le volume du parallélépipède rectangle déformé :

$$\begin{aligned} 1 + \Delta V &= (1 + \varepsilon_1)(1 + \varepsilon_2)(1 + \varepsilon_3) \\ &= 1 + (\varepsilon_1 + \varepsilon_2 + \varepsilon_3) + O(\varepsilon^2) \\ &= 1 + \varepsilon_I \end{aligned} \quad (20.24)$$

Ainsi la partie sphérique représente une dilatation (si elle est positive, une contraction sinon) uniforme dans toutes les directions, tandis que le déviateur correspond à une déformation isochore (sans variation de volume). Il existe une base orthonormée dans laquelle le tenseur des déformations est diagonal :

$$\varepsilon = \begin{bmatrix} \varepsilon_I & 0 & 0 \\ 0 & \varepsilon_{II} & 0 \\ 0 & 0 & \varepsilon_{III} \end{bmatrix} \quad (20.25)$$

Les directions propres sont appelées directions principales de déformation, et les déformations ε_I , ε_{II} et ε_{III} les déformations principales. Les déformations principales sont les valeurs propres du tenseur, et les directions propres, ses vecteurs propres. Les valeurs propres λ vérifient l'équation $\det(\varepsilon - \lambda I) = 0$. La trace étant invariante par changement de base, on a $\varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33} = \varepsilon_I + \varepsilon_{II} + \varepsilon_{III}$ et et ainsi en petites déformations, la variation relative de volume vaut :

$$\frac{\Delta V}{V_0} = \varepsilon_I + \varepsilon_{II} + \varepsilon_{III} \quad (20.26)$$

Contrairement aux contraintes principales, la notion de déformation principale est assez peu utilisée pour le calcul. Elle permet par contre d'exprimer de manière simple l'énergie élastique, et est utile pour dépouiller les résultats d'extensométrie. Par ailleurs, les directions principales sont les mêmes pour le tenseur des déformations et pour le tenseur des contraintes.

En vue de prendre en compte les cas les plus généraux, le tenseur des déformations ε se décompose en quatre parties :

- une partie élastique : directement proportionnelle à la variation du tenseur des contraintes (contraintes actuelles moins tenseur des contraintes initiales, généralement nul, mais pas forcément) ;
- une partie de dilatation thermique : directement proportionnel à la variation de température (température actuelle moins température initiale). Cette partie s'écrit à l'aide d'un tenseur α , dépendant éventuellement de la température également, et qui est sphérique dans le cas des matériaux isotropes ;
- une partie plastique ;
- une partie viscoélastique.

Chaque mécanisme responsable du comportement inélastique est caractérisé par un certain nombre de variables, appelées variables d'écrouissage, caractéristique de l'état du matériau à un instant donné ainsi que de l'influence de chargement thermomécanique passé.

Les lois d'écrouissage définissent l'évolution du domaine élastique. Elles complètent le modèle pour le cas d'un matériau dont la résistance à la déformation évolue avec celle-ci. Sans écrouissage, le domaine d'élasticité est défini uniquement en fonction de l'état de contrainte.

20.2 Non linéarité géométrique

Nous avons exposé en introduction qu'il s'agit du cas où l'hypothèse des petites perturbations n'est plus vérifiée. Cela se traduit par le fait que le domaine considéré Ω varie et doit donc être introduit comme une inconnue dans le problème.

Si l'on se place dans le cas de grands déplacements, mais en conservant l'hypothèse de petites déformations, on est amené à effectuer la formulation variationnelle sur un domaine inconnu Ω_0 par l'introduction des tenseurs non linéaires de Piola-Kirchhoff (contraintes) et de Green-Lagrange (déformations) tels que :

$$\int_{\Omega} \delta \varepsilon : \sigma = \int_{\Omega_0} \delta \varepsilon_{GL} : \sigma_{PK} \quad (20.27)$$

On obtient la formulation variationnelle :

$$\int_{\Omega_0} \delta \varepsilon_{GL} : \sigma_{PK} - \int_{\Gamma_0} \delta u \cdot f_{\Gamma_0} - \int_{\Omega_0} \delta u \cdot F_{\Omega_0} \quad (20.28)$$

avec $\Gamma_0 = \delta \Omega_0$

Au niveau discret, le système non linéaire nécessite alors de recourir à une technique de linéarisation, comme la méthode de Newton-Raphson, permettant de manière itérative, d'obtenir une solution convergée. La méthode de Newton-Raphson est présentée en annexe au chapitre D.

20.3 Non linéarité matérielle

Jusqu'à présent, nous nous sommes placés, de manière implicite, dans le cadre d'un comportement élastique linéaire. On parle encore de loi de Hooke. Cela a été illustré au paragraphe 6.6.5.

Histoire

Lorsque l'on ne considère que le cas de la traction/compression, alors on a proportionnalité entre contrainte dans cette direction de chargement σ_{11} et déformation dans cette direction ε_{11} via le module d'Young E : $\sigma_{11} = E \varepsilon_{11}$.

La même loi se retrouve pour une sollicitation en cisaillement, et la contrainte de cisaillement τ_{12} est proportionnelle à l'angle de déformation relative γ_{12} via le module de Coulomb G : $\tau_{12} = G \varepsilon_{12}$.

Lorsque l'on synthétise tout cela pour toutes les directions, on parle alors de loi de Hooke généralisée, que l'on note sous la forme $\sigma = H \varepsilon$ (i.e. $\sigma_{ij} = H_{ijkl} \varepsilon_{kl}$) : il y a proportionnalité entre les tenseurs des contraintes et des déformations (On note également souvent C ou D au lieu de H).

Nous rappelons que, pour un matériau isotrope, tous les coefficients H_{ijkl} sont définis à l'aide du module d'Young E et du coefficient de Poisson ν , ou de manière équivalente par les coefficients de Lamé λ et μ .



Hooke

Young

Coulomb

Poisson

Lamé

$$\sigma = 2\mu \varepsilon + \lambda \text{tr}(\varepsilon) \mathbf{I} \quad \text{i.e.} \quad \sigma_{ij} = 2\mu \varepsilon_{ij} + \lambda \varepsilon_{kk} \delta_{ij} \quad (20.29)$$

avec :

$$\lambda = \frac{E\nu}{(1-2\nu)(1+\nu)} \quad \text{et} \quad \mu = \frac{E}{2(1+\nu)} \quad (20.30)$$

ou :

$$E = \frac{\mu(3\lambda + 2\mu)}{\lambda + \mu} \quad \text{et} \quad \nu = \frac{\lambda}{2(\lambda + \mu)} \quad (20.31)$$

Dans ce paragraphe, nous proposons d'exposer brièvement quelques lois de comportement qui vont au-delà de la simple élasticité linéaire.

20.3.1 Modèles rhéologiques

L'allure qualitative de la réponse des matériaux à quelques essais simples (traction, compression, écoulement, fluage, relaxation, triaxial, flexion, torsion...) permet de les ranger dans des classes bien définies. Ces comportements « de base », qui peuvent être représentés par des systèmes mécaniques élémentaires, sont l'élasticité, la plasticité et la viscosité :

- *Le ressort* symbolise l'élasticité linéaire parfaite, pour laquelle la déformation est entièrement réversible lors d'une décharge, et où il existe une relation biunivoque (i.e. une bijection) entre les paramètres de charge et de déformation.

- *L'amortisseur* schématise la viscosité, linéaire ou non. La viscosité est dite pure s'il existe une relation biunivoque entre la charge et la vitesse de chargement. Si cette relation est linéaire, le modèle correspond à la loi de Newton.
- *Le patin* symbolise l'apparition de déformations permanentes lorsque la charge est suffisante. Si le seuil d'apparition de la déformation permanente n'évolue pas avec le chargement, le comportement est dit plastique parfait. Si, de plus, la déformation avant écoulement est négligée, le modèle est rigide-parfaitement plastique.

Ces éléments peuvent être combinés entre eux pour former des modèles rhéologiques. La réponse de ces systèmes peut être jugée dans trois plans différents, qui permettent d'illustrer le comportement lors d'essais de type :

- *Écrouissage*, ou augmentation monotone de la charge ou de la déformation (plan $\varepsilon - \sigma$) ;
- *Fluage*, ou maintien de la charge (plan $t - \varepsilon$) ;
- *Relaxation*, ou maintien de la déformation (plan $t - \sigma$).

Les réponses de modèles classiques selon ces trois plans précédents sont présentées ci-dessous :

- *modèle du solide élastique* : $\sigma = H\varepsilon$, loi de Hooke ;
- *modèle du solide viscoélastique* comportant un ressort et un amortisseur en parallèle : $\sigma = \eta\dot{\varepsilon} + H\varepsilon$, modèle de Voigt ;
- *modèle du solide élastique-parfaitement plastique*, constitué par un ressort linéaire et un patin en série : modèle de Saint-Venant.
Lorsque le module E tend vers l'infini, le modèle devient rigide-parfaitement plastique.
- *modèle du solide élastique-plastique écrouissable*, qui donne une courbe de traction linéaire par morceaux : modèle de Saint-Venant généralisé ;
- *modèle du solide élastique-parfaitement viscoplastique*, formé par un amortisseur non linéaire : modèle de Norton-Hoff.
- *modèle du solide élastique-parfaitement viscoplastique*, qui comporte un ressort linéaire en série avec un amortisseur et un patin situés en parallèle : modèle de Bingham-Norton ;
Lorsque le seuil du patin tend vers zéro, et que l'amortisseur est choisi linéaire, ce dernier modèle dégénère en un *modèle de fluide visqueux*, comportant un ressort et un amortisseur en série : $\dot{\varepsilon} = \dot{\sigma}/E + \sigma/\eta$, modèle de Maxwell.
- *modèle du solide élastique-viscoplastique écrouissable*, qui représente le schéma le plus complexe.

Excepté le cas de l'élasticité (déjà traité), chacun des modèles présentés ci-dessus s'exprime sous forme *différentielle*, si bien que la réponse actuelle dépend de la sollicitation actuelle et de son histoire (propriété d'hérédité).

Il y a deux manières de prendre en compte cette histoire, *la première* consiste à la décrire par une dépendance fonctionnelle entre les variables ; *la seconde* fait l'hypothèse qu'il est possible de représenter l'effet de l'histoire dans des variables internes, qui « concentrent » les informations importantes définissant l'état du matériau. Sauf quelques cas exceptionnels comme celui de la viscoélasticité linéaire, la seconde méthode de travail produit des modèles dont la modélisation numérique est plus simple. Les autres hypothèses importantes qui sont classiquement utilisées pour l'écriture de modèles de comportement sont :

- Le principe de l'état local, qui considère que le comportement en un point ne dépend que des variables définies en ce point, et non pas du voisinage ;
- Le principe de simplicité matérielle, qui suppose que seul intervient dans les équations de comportement le premier gradient de la transformation ;
- Le principe d'objectivité, qui traduit l'indépendance de la loi de comportement vis-à-vis de l'observateur, et qui implique que le temps ne peut pas intervenir explicitement dans les relations de comportement.

Mentionnons enfin quelques cas type d'utilisation des modèles mentionnés :

- Comportements viscoélastique : pour les polymères thermoplastiques au voisinage de la température de fusion, pour les verres au voisinage de la température de transition, pour les

bétons frais...

- Comportements rigides-parfaitement plastiques : pour l'étude des sols, pour l'analyse limite, pour la mise en forme des métaux...
- Comportements plastiques : pour les métaux à des températures inférieures au quart de la température de fusion, pour les sols et roches...
- Comportements viscoplastiques : pour les métaux à moyenne et haute température, pour le bois, les sols (dont le sel), pour les céramiques à très haute température...

Un acier à température ambiante peut être considéré comme élastique linéaire pour le calcul des flèches d'une structure mécanique, viscoélastique pour un problème d'amortissement de vibrations, rigide-parfaitement plastique pour un calcul de charge limite, élasto-viscoplastique pour l'étude de contraintes résiduelles,

Un polymère peut être considéré comme un solide pour un problème de choc, et comme un fluide pour l'étude de sa stabilité sur de longues durées...

20.3.2 Visoélasticité

Un comportement viscoélastique correspond à la superposition d'un comportement élastique, traduit par une relation de type $\sigma = H\varepsilon$ (loi de Hooke), et un comportement visqueux, dont le plus simple est le modèle linéaire dit de Newton, traduit par une relation de type $\sigma = \eta\dot{\varepsilon}$ (η étant la viscosité du matériau).

Un échelon de contrainte appliqué à partir d'un instant t_0 produit une déformation instantanée suivie d'une déformation différée (fluage). Si, au-delà d'un instant t_1 , la charge est ramenée à zéro, il apparaît, après une nouvelle déformation instantanée, le phénomène de recouvrance, qui tend à ramener la déformation à zéro. Si une déformation est appliquée à partir de l'instant t_0 , on obtient une contrainte instantanée puis une diminution de la contrainte à partir de cette valeur instantanée (relaxation). Si, au-delà d'un instant t_1 , la déformation est ramenée à zéro, il apparaît, après une nouvelle contrainte instantanée, le phénomène d'effacement, qui tend à ramener la contrainte à zéro. Le comportement viscoélastique se caractérise par le fait que le phénomène d'effacement est total, i.e. que la contrainte revient effectivement à zéro.

De manière générale, une loi viscoélastique s'exprime comme une correspondance entre l'histoire des déformations et des contraintes par une fonction, ce que l'on note :

$$\varepsilon(t) = \Gamma_{\tau < t}(\sigma(t)) \quad (20.32)$$

De plus, le comportement sera dit viscoélastique linéaire si le comportement vérifie le principe de superposition de Boltzman : la réponse à la somme de deux sollicitations est la somme des réponses à chaque sollicitation. Cela se traduit par :

$$\varepsilon(t) = \Gamma_{\tau < t}(\sigma_1 + \sigma_2) = \Gamma_{\tau < t}(\sigma_1) + \Gamma_{\tau < t}(\sigma_2) \quad (20.33)$$

20.3.3 Visoplasticité

Plusieurs modèles viscoplastiques existent. Nous avons mentionné, le modèle de Maxwell ou *modèle de fluide visqueux*, comportant un ressort et un amortisseur linéaire en série, et dont la loi s'écrit :

$$\dot{\varepsilon} = \frac{\dot{\sigma}}{E} + \frac{\sigma}{\eta} \quad (20.34)$$

Nous avons également mentionné le modèle de Voigt ou *modèle du solide viscoélastique*, comportant un ressort et un amortisseur en parallèle, et dont la loi s'écrit :

$$\sigma = \eta\dot{\varepsilon} + H\varepsilon \quad (20.35)$$

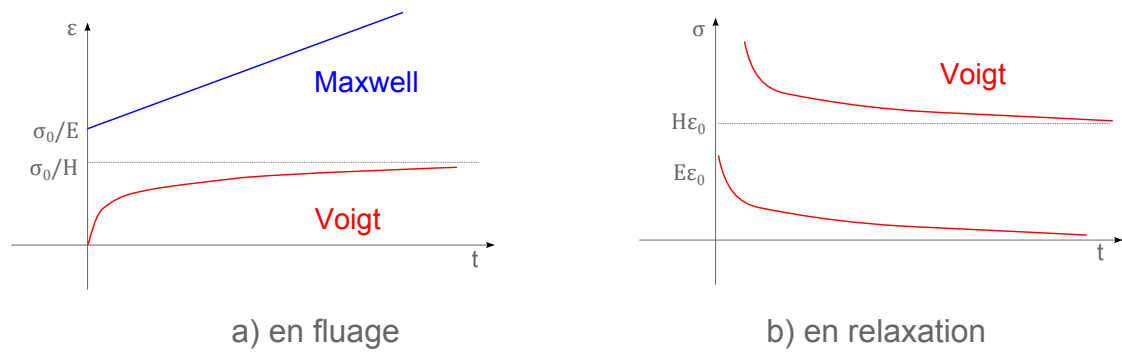


FIGURE 20.1: Modèles de Maxwell et Voigt

La figure 20.1 illustre les réponses des modèles de Maxwell et Voigt.

Le modèle de Voigt ne présente pas d'élasticité instantanée. L'application d'un saut de déformation en $t = 0$ produit une contrainte infinie. Ce modèle n'est donc pas utilisable en relaxation, sauf si la mise en charge est progressive, et sera pour cette raison associé à un ressort pour effectuer des calculs de structure (modèle de Kelvin-Voigt).

De plus, sous l'effet d'une contrainte σ_0 constante en fonction du temps, la déformation dans ce modèle de Voigt tend vers la valeur asymptotique σ_0/H : le fluage est donc limité. Si, après une mise en charge lente, la déformation est fixée à une valeur ε_0 , la contrainte asymptotique sera $H\varepsilon_0$: il n'y a donc pas disparition complète de la contrainte.

Au contraire, dans le cas du modèle de Maxwell, la vitesse de fluage est constante, et la disparition de contrainte au cours de la relaxation est totale.

20.3.4 Plasticité

Les modèles présentés jusqu'à présent étaient des modèles unidimensionnels, ou plus exactement des modèles correspondant à un chargement uniaxial. Pourtant, l'étude de ces modèles uniaxiaux (simples) met en évidence la détermination de seuils ou de limites correspondant à des modifications de comportement. Afin de pouvoir aborder l'étude des chargements multiaxiaux, il est nécessaire de se donner les moyens de définir de telles limites dans le cas tridimensionnel. C'est ce que nous allons maintenant aborder.

Considérons le cas du chargement uniaxial d'un matériau isotrope. Celui-ci fait apparaître un domaine d'élasticité au travers de deux valeurs de contrainte, l'une en traction, l'autre en compression, pour lesquelles se produit l'écoulement plastique. On a donc élasticité dans un domaine $[-\sigma_y, \sigma_y]$, puis plasticité au delà, i.e. par exemple pour une contrainte supérieure à $\sigma_y + x$, où $x = H\varepsilon_p$.

En fait, la limite du domaine de plasticité est défini par une fonction de charge f de sorte que si $f(\sigma, x) < 0$ l'état de contrainte est élastique, et si $f(\sigma, x) > 0$ l'état de contrainte est plastique.

Dans le cas général, l'ensemble des paramètres de départ A_I contiendra les contraintes et toutes les variables d'écrouissage, scalaires ou tensorielles, il faut donc définir $f(\sigma, A_I)$. On va dans un premier temps limiter la présentation à la définition du domaine d'élasticité initial, pour lequel on supposera que les variables A_I sont nulles, si bien qu'on se contentera d'écrire les restrictions des fonctions f dans l'espace des contraintes.

L'expérience montre que, pour la plupart des matériaux, le domaine d'élasticité initial est convexe (c'est en particulier vrai pour les métaux qui se déforment par glissement cristallographique). La fonction de charge doit donc elle-même être convexe en σ , ce qui implique, pour tout réel λ compris entre 0 et 1, et pour un couple (σ_1, σ_2) quelconque de la frontière :

$$f(\lambda\sigma_1 + (1 - \lambda)\sigma_2) \leq \lambda f(\sigma_1) + (1 - \lambda)f(\sigma_2) \quad (20.36)$$

Il faut également respecter les symétries matérielles. Ceci implique en particulier dans le cas d'un matériau isotrope que f soit une fonction symétrique des seules contraintes principales, ou bien encore, ce qui est équivalent, des invariants du tenseur des contraintes présentés au paragraphe 20.1.1 (il s'agit de I_1 , I_2 et I_3).

Dans les matériaux métalliques on observe généralement l'incompressibilité plastique ($\varepsilon_{ii}^p = 0$) et l'indépendance du comportement vis-à-vis de la pression isostatique. Ceci amène à considérer comme variable critique à faire figurer dans la définition du critère non plus le tenseur de contraintes lui-même, mais son déviateur s (et donc ses invariants J_1 , J_2 et J_3).

En vue de réaliser les comparaisons avec les résultats expérimentaux, il est pratique de disposer d'expressions des critères dans lesquelles les valeurs de f sont homogènes à des contraintes. On peut alors remplacer J_2 par l'invariant J (contrainte équivalente au sens de von Mises en cisaillement : $J(\sigma) = \sigma_e = \sqrt{3J_2}$), qui peut également s'exprimer en fonction des contraintes principales, ou de la contrainte appliquée dans le cas d'un état de traction simple.

Nous allons maintenant présenter quelques critères classiques de plasticité : von Mises et Tresca, ne faisant pas intervenir la pression isostatique, Drucker-Prager, Mohr-Coulomb, Hill et Tsai, faisant intervenir la pression isostatique.

Introduire la pression isostatique permet d'exprimer le fait qu'une contrainte isostatique de compression rend plus difficile la déformation plastique, ce qui conduit à une dissymétrie des critères en traction/compression.

Histoire

Un critère de plasticité, ou critère d'écoulement plastique, est un critère permettant de savoir, sous des sollicitations données, si une pièce se déforme plastiquement ou si elle reste dans le domaine élastique. De nombreux essais ont montré que l'on pouvait utiliser deux critères principaux : le critère de von Mises (critère de l'énergie de distorsion élastique) ou le critère de Tresca (critère de la contrainte de cisaillement maximal). En résistance des matériaux, on désire parfois rester dans le domaine élastique, on parle alors de critère de résistance.

La contrainte de comparaison n'est pas une contrainte réelle existant à un instant donné à l'intérieur d'un solide, mais est utilisée en mécanique pour prédire la rupture. Néanmoins, la plupart des ingénieurs l'utilisent pour déterminer si un champ de contrainte donné dans une pièce est acceptable ou non. On parle aussi de contrainte équivalente ou de contrainte effective. Elle découle des critères de plasticité.

Cette contrainte est comparée à la limite d'élasticité ou encore la contrainte de rupture obtenue par essai de traction.

Le critère dit de von Mises a été formulé initialement par Maxwell en 1865. En 1904, Huber le développa partiellement dans un article en polonais. Cependant, sa paternité est généralement attribuée à von Mises (1913). On parle aussi parfois de la théorie de Maxwell-Huber-Hencky-von Mises, ou de critère de Prandtl-Reuss, ou encore de critère de l'énergie de distorsion élastique.

La renommée de Tresca était si grande à son époque que Gustave Eiffel mit son nom en troisième position sur la Liste des soixante-douze noms de savants inscrits sur la tour Eiffel, et plus précisément sur le pilier face au Trocadéro.

Sur ce même pilier (comportant 18 noms), en plus des mathématiciens Lagrange, Laplace, Legendre et Chasles, se trouve également Navier. Celui-ci n'apparaît pas pour ses contributions aux mathématiques et à la physique^a, mais parce qu'il était considéré lui-aussi comme l'un des plus grands ingénieurs français et comme personnage public important : de 1830 à sa mort en 1836, il fut employé par le gouvernement français comme consultant afin de permettre à la France de progresser grâce aux sciences et aux technologies.

^a. Lorsque, en 1822, il modifia les équations d'Euler pour décrire un fluide en incluant la viscosité, son raisonnement mathématique était erroné, mais par chance, ou grâce à son intuition, il obtint malgré tout les bonnes équations. Le raisonnement rigoureux fut trouvé quelques années plus tard par le mathématicien irlandais Stokes



Maxwell

von Mises

Tresca

Critère de von Mises

Dans le critère de von Mises, on considère que le seuil de plasticité est lié à l'énergie élastique de cisaillement. Cela revient à négliger l'influence du troisième invariant. Si σ_y est la limite d'élasticité en traction, la fonction de charge est définie par :

$$f(\sigma) = J(\sigma) - \sigma_y \quad (20.37)$$

La direction d'écoulement est donnée par le déviateur du tenseur des contraintes.

Critère de Tresca

Le critère de Tresca, ou critère de Tresca-Guest, ou critère de la contrainte de cisaillement maximal, s'exprime en fonction des cisaillements maxima dans chaque plan principal, représentés par les quantités $(\sigma_i - \sigma_j)$ (et non plus à l'énergie élastique de cisaillement).

La spécificité du critère de Tresca est de ne retenir que le plus grand d'entre eux. Le fait de rajouter une pression à chaque terme de la diagonale ne modifie pas, comme prévu, la valeur du critère. Contrairement au critère de von Mises, l'expression du critère de Tresca ne définit en général pas une surface régulière (discontinuité de la normale, points anguleux) :

$$f(\sigma) = \max(\sigma_i - \sigma_j) - \sigma_y \quad (20.38)$$

En d'autres termes, la loi d'écoulement se définit par secteurs dans l'espace des contraintes principales.

Critère de Drucker-Prager

On trouve parfois « Drucker » au lieu de « Drucker » dans la littérature.

La fonction de charge s'écrit $f(\sigma) = (1 - \alpha)J(\sigma) + \alpha tr(\sigma)$, si bien que la normale possède une composante sphérique. La déformation plastique évaluée avec un tel critère est accompagnée d'une augmentation de volume quel que soit le chargement appliqué.

De manière générale, tout critère qui fait apparaître la pression isostatique produit un terme de changement de volume accompagnant la déformation plastique (mais pas forcément positif, comme dans le cas du modèle de Drucker-Prager qui comporte donc un défaut).

On l'écrit aussi sous une forme où il est plus facile de le voir comme une extension du critère de von Mises :

$$f(\sigma) = J(\sigma) - \frac{\sigma_y - \alpha I_1}{1 - \alpha} \quad (20.39)$$

où le coefficient α dépend du matériau et est compris entre 0 et 1/2. Pour $\alpha = 0$, on retrouve le critère de von Mises.

Critère de Mohr-Coulomb

Ce critère est apparenté à celui de Tresca, car il fait intervenir comme lui le cisaillement maximum, mais, contrairement à lui, il ajoute en même temps la contrainte « moyenne », représentée par le centre du cercle de Mohr correspondant au cisaillement maximum, soit :

$$f(\sigma) = \sigma_1 - \sigma_3 + (\sigma_1 + \sigma_3) \sin \varphi - 2C \cos \varphi \quad \text{avec} \quad \sigma_3 \leq \sigma_2 \leq \sigma_1 \quad (20.40)$$

Ce critère est sous-tendu par la notion de frottement, et suppose que le cisaillement maximal que peut subir le matériau est d'autant plus grand que la contrainte normale de compression est élevée.

Critères anisotropes

La méthode généralement utilisée pour faire apparaître de l'anisotropie consiste à faire intervenir un tenseur du quatrième ordre dans l'expression du critère, qui vient multiplier le déviateur, ou directement le tenseur des contraintes.

Une solution couramment adoptée généralise le critère de von Mises, en utilisant à la place de $J(\sigma)$ l'expression : $J_H(\sigma) = \sqrt{\sigma : H : \sigma}$.

Le critère de Hill correspond à une anisotropie particulière qui conserve trois plans de symétrie dans l'état d'érouissage du matériau.

Le critère de Tsai est obtenu à partir de celui de Hill afin de représenter la dissymétrie entre traction et compression.

20.3.5 Les élastomères

Dans ce paragraphe, nous faisons une présentation sommaire des élastomères de manière générale.

Le comportement des élastomères est fortement non linéaire. Il faut prendre en compte par exemple la précharge, la fréquence et l'amplitude d'excitation comme paramètres :

- en statique : très grandes déformations et retour à la configuration initiale sans déformation permanente. Ce sont les lois hyperélastiques ;
- en dynamique : propriétés amortissantes, dont rigidification en fréquence sous excitation harmonique. C'est la loi viscoélastique non linéaire.

Concernant leur fabrication, l'opération de vulcanisation est la suivante : on malaxe du caoutchouc brut, on ajoute du soufre et on chauffe le mélange afin d'obtenir un matériau élastique, stable dans une gamme de température beaucoup plus large que le caoutchouc naturel et résistant au fluage sous contrainte (découverte de Goodyear en 1839).

Le premier brevet sur la fabrication d'un élastomère synthétique a été déposé le 12 septembre 1909 par le chimiste allemand Fritz Hofmann.

Les élastomères sont quasi incompressibles. le module de compressibilité du caoutchouc se situe entre 1000 et 2000 MPa alors que l'ordre de grandeur de son module de cisaillement est de 1 MPa : cette différence signifie que le caoutchouc ne change quasiment pas de volume, même sous de fortes contraintes.

Les déformations de cisaillement peuvent être considérées comme linéaires : le coefficient de cisaillement est relativement indépendant du taux de cisaillement, au moins jusqu'à des niveaux de déformation modérés.

Enfin, notons également un comportement particulier, connu sous le nom d'effet Mullins (publications en 1966 et 1969) : si l'on applique un chargement cyclique sur un matériau initialement non précontraint, on observe une diminution de la raideur lors des premiers cycles. Si l'on impose ensuite une déformation cyclique jusqu'à un niveau de déformation plus élevé, on observe à nouveau une diminution de la contrainte et de l'hystérésis jusqu'à un nouvel équilibre. Ce comportement provient d'une rupture progressive des liaisons moléculaires.

Les élastomères sont connus pour leur grande élasticité. Une hystérésis est toujours présente, mais augmente avec l'ajout de charges (qui sont nécessaires pour améliorer d'autres propriétés).

Rappelons ce qu'est l'hystérésis en quelques mots : l'amortissement correspond à l'énergie dissipée au cours d'un cycle. Il est caractérisé par un angle de perte et un module complexe. Si le système est soumis à des déformations sinusoïdales cycliques : $\varepsilon(t) = \varepsilon_0 \sin(\omega t)$ et que l'on estime que la contrainte transmise répond également de façon sinusoïdale, alors cette dernière est déphasée d'un angle δ , l'angle de perte, et on a : $\sigma(t) = \sigma_0 \sin(\omega t + \delta)$. Généralement cette hypothèse de première harmonique n'est pas vérifiée et la réponse contient plusieurs harmoniques : $\sigma(t) = \sum_k \sigma_k \sin(k\omega t + \delta_k)$.

Le module complexe est défini par : $e^* = \sigma/\varepsilon$. Il se calcule comme :

$$e^* = \frac{\sigma_0}{\varepsilon_0} e^{i\delta} = \frac{\sigma_0}{\varepsilon_0} (\cos \delta + i \sin \delta) \quad (20.41)$$

ou encore :

$$e^* = E' + iE'' = E'(1 + i \tan \delta) \quad (20.42)$$

et on appelle :

- module dynamique $= |e^*| = \frac{\sigma_0}{\varepsilon_0}$;
- module de stockage $= E' = |e^*| \cos \delta$, car il mesure l'énergie emmagasinée puis restituée au cours d'un cycle ;
- module de perte $= E'' = |e^*| \sin \delta$, car il mesure l'énergie dissipée sous forme de chaleur au cours d'un cycle.

Le module d'Young complexe e^* ou le module de Coulomb complexe G^* évoluent de manière significative avec la température, la fréquence et l'amplitude de l'excitation.

Notons également que l'effet mémoire (expériences de fluage et relaxation) est présent pour ces matériaux : le niveau des contraintes à un instant dépend non seulement du niveau de sollicitation à cet instant, mais également des sollicitations auxquelles le matériau a été soumis précédemment.

Grandes déformations

En grandes déformations, il est nécessaire de bien distinguer l'état initial de l'état déformé. On utilisera alors les tenseurs adaptés (Piola-Kirchhoff), comme exposé au paragraphe 20.2.

Incompressibilité

Aux lois de comportement précédentes, on ajoute la contrainte de variation nulle de volume entre les configurations.

Pour un matériau incompressible, on a $J = 1$ (J est le jacobien du gradient de la déformation).

Hyperélasticité

Un matériau est dit élastique si le tenseur des contraintes de Cauchy à l'instant t dépend uniquement de l'état de déformation à ce même instant : la contrainte ne dépend pas du chemin suivi par la déformation, alors que le travail fourni par cette contrainte en dépend.

Un matériau élastique est dit hyperélastique si le tenseur des contraintes dérive d'une fonction d'énergie de ce matériau : cela implique que le travail fourni pour aller d'un état à un autre ne dépend pas du chemin suivi.

Si on postule l'existence d'une énergie libre Ψ , on peut la relier, pour les matériaux hyperélastiques, aux invariants des tenseurs. Cette énergie est également appelée énergie de déformation. De là on peut déduire les lois de comportement.

Approximation numérique

De nombreuses formes de l'énergie de déformation ont été proposées, s'exprimant à partir :

- des invariants ;
- des fonctions d'élongations principales ;
- des coefficients intervenant sous forme linéaire : Mooney-Rivlin et Rivlin généralisé ;
- des coefficients intervenant sous forme de puissances : Ogden (1972)
- sous forme polynomiale...

Le modèle de Rivlin généralisé, implémenté dans la plupart des codes de calcul, est donné par le développement polynomial suivant :

$$\Psi = \sum_{i,j=0}^N C_{ij} (I_1 - 3)^i (I_2 - 3)^j + \sum_{i=1}^M D_i (J - 1)^{2i} \quad (20.43)$$

ou plus simplement, si le matériau est bien incompressible :

$$\Psi = \sum_{i,j=0}^N C_{ij} (I_1 - 3)^i (I_2 - 3)^j \quad (20.44)$$

où l'énergie de déformation est développée à un ordre proportionnel à la plage de déformation souhaitée (pour $N = 3$ on a généralement une bonne corrélation avec les mesures expérimentales). $C_{00} = 0$ et les coefficients C_{ij} sont des constantes du matériau liées à sa réponse en terme de distorsion, et les coefficients D_i sont des constantes du matériau liées à sa réponse en terme de volume.

Pour les matériaux faiblement compressibles, il est commode d'introduire la décomposition des déformations. On en arrive alors à la forme :

$$\Psi(J) = \sum_{i=1}^N \frac{1}{D_i} (J - 1)^{2i} \quad (20.45)$$

le module de compressibilité étant donné par $k_0 = 2/D_1$.

Notons qu'une manière également simple de prendre en compte les élastomère a déjà été donnée : il s'agit de l'utilisation d'éléments hybrides faisant intervenir la matrice de souplesse [S] au lieu de la matrice de Hooke [H].

20.3.6 Les composites, l'anisotropie

Dans ce paragraphe, nous faisons une présentation sommaire des matériaux composites de manière générale.

Ceux-ci ont fait des apparitions en divers endroits de ce document, et quelques problèmes se rapportant à eux ont été discutés.

Définition 80 — Matériau composite. On appelle matériau composite l'association d'au moins deux matériaux non miscibles. On obtient un matériau hétérogène.

On en profite donc pour rappeler qu'un matériau peut être :

- *Homogène* : même propriétés en tout point du matériau.
- *Hétérogène* : en deux points différents, propriétés différentes.
- *Isotrope* : même propriétés dans toutes les directions.
- *Isotrope transverse* : il existe un axe de symétrie. Symétrie par rapport à une droite.
- *Orthotrope* : propriétés symétriques par rapport à deux plans orthogonaux.
- *Anisotrope* : les propriétés sont différentes selon les différentes directions.

Composants

Un matériau composite plastique correspond à l'association de deux constituants .

- Le renfort (ou armature, ou squelette) : assure la tenue mécanique (résistance à la traction et rigidité). Souvent de nature filamenteuse (des fibres organiques ou inorganiques).
- La matrice : lie les fibres renforts, répartit les efforts (résistance à la compression ou à la flexion), assure la protection chimique. Par définition, c'est un polymère ou une résine organique.

En plus de ces deux constituants de base, il faut ajouter l'interphase, qui est l'interface assurant la compatibilité renfort-matrice, et qui transmet les contraintes de l'un à l'autre sans déplacement relatif.

Des produits chimiques entrent aussi dans la composition du composite, l'interphase etc... qui peuvent jouer sur le comportement mécanique, mais n'interviennent pratiquement jamais dans le calcul de structure composite.

Remarque. On conçoit un composite en fonction du type d'application, de chargement... ce qui est différent des matériaux classiques où on adapte la conception d'une structure en fonction du matériau constitutif. On cherchera donc toujours à orienter au mieux les renforts en fonction des efforts auxquels la structure est soumise.

Avantages des matériaux composites :

- grande résistance à la fatigue ;
- faible vieillissement sous l'action de l'humidité, de la chaleur, de la corrosion (sauf alu-carbone) ;
- insensibilité aux produits chimiques « mécaniques » comme les graisses, huiles, liquides hydrauliques, peintures, solvants, pétrole ;

Mais attention aux décapants de peinture qui attaquent les résines époxydes !

Les matériaux composites structuraux

- Les monocouches représentent l'élément de base de la structure composite. Les différents types de monocouches sont caractérisés par la forme du renfort : à fibres longues (unidirectionnelles UD, réparties aléatoirement), à fibres tissées, à fibres courtes.
- Un stratifié est constitué d'un empilement de monocouches ayant chacune une orientation propre par rapport à un référentiel commun aux couches et désigné comme le référentiel du stratifié.
- Un matériau sandwich est un matériau composé de deux peaux de grande rigidité et de faible épaisseur enveloppant une âme (ou cœur) de forte épaisseur et faible résistance. L'ensemble forme une structure d'une grande légèreté. Le matériau sandwich possède une grande rigidité en flexion et est un excellent isolant thermique.

On pourra avoir des stratifiés de type :

- *Équilibré* : stratifié comportant autant de couches orientée suivant la direction $+\theta$ que de couches orientée suivant la direction $-\theta$.
- *Symétrique* : stratifié comportant des couches disposées symétriquement par rapport à un plan moyen.
- *Orthogonal* : stratifié comportant autant de couches à 0° que de couches à 90° .
- Notation « composite » : on porte, entre crochets, un nombre indiquant la valeur en degré de l'angle que fait la direction des fibres de chaque couche avec l'axe de référence. Les couches sont nommées successivement en allant de la face inférieure à la face supérieure, les couches étant séparées par le symbole « / » (ou une virgule parfois) : $[0/ + 45/ + 90/ - 45]$.

Les couches successives d'un même matériau et de même orientation sont désignées par un indice : $[0/45/45/90/ - 45/ - 45/0]$ s'écrit $[0/45_2/90/ - 45_2/0]$.

Si le stratifié est symétrique, seule la moitié est codifiée et le symbole « s » indique la symétrie : $[-45/45/ - 45/ - 45/45/ - 45]$ s'écrit $[-45/45/ - 45]_s$.

En cas de stratification hybride (différents matériaux dans un même stratifié), il faut préciser par un indice la nature de la couche.

Approximation numérique

L'ingénieur mécanicien est souvent bien au fait des diverses approximations possibles d'un matériau composite. Nous serons par conséquent assez brefs sur le sujet.

Par ailleurs le chapitre 15 sur l'homogénéisation permet, dans certains cas, de remplacer un matériau composite compliqué par un matériau homogénéisé plus simple.

La loi de Hooke, telle qu'elle a été écrite jusqu'à présent $\sigma = H\varepsilon$ ne s'oppose aucunement à l'anisotropie.

L'isotropie se traduit juste par le fait que plus de coefficients sont nuls, et que ceux qui sont non nuls ne dépendent que de deux paramètres, comme rappelé en introduction au paragraphe 20.3.

Lorsque le matériau est anisotrope, alors H peut être pleine, ce qui correspond à 21 coefficients non nuls, dépendant des paramètres : E_i : modules de tensions, G_{ij} : modules de cisaillement, ν_{ij} : coefficients de contraction, $\eta_{ij,k}$: coefficients d'influence de 1ère espèce, $\eta_{i,kl}$: coefficients d'influence de 2nde espèce, et $\nu_{ij,kl}$: coefficients de Chentsov.

L'orthotropie (orthogonal et anisotrope, i.e. 2 plans orthogonaux de symétrie) fait descendre ce nombre de coefficients non nuls à 9, dépendant des paramètres : E_1, E_2, E_3 : modules d'élasticité longitudinaux, G_{23}, G_{13}, G_{12} : modules de cisaillement, $\nu_{23}, \nu_{13}, \nu_{12}, \nu_{21}, \nu_{31}, \nu_{32}$: coefficients de Poisson.

L'isotropie transverse (1 axe de symétrie) fait encore descendre ce nombre de coefficients non nuls à 6 (on n'a plus besoin de E_2, G_{23} et G_{12} par exemple)

On peut également faire des hypothèses sur l'épaisseur des plis (faible) et la répartition des contraintes ou déformations pour obtenir des théories de plaques équivalentes. Ces plaques équivalentes pouvant ensuite être assemblées en une nouvelle plaque équivalente.

Chaque coefficient non nul peut lui-même être constant (élasticité anisotrope) ou non, permettant de prendre en compte autant de comportements que nécessaire (viscoélasticité, plasticité...). En sus, il est également nécessaire de judicieusement modéliser l'interphase, i.e. de bien décrire comment les couches peuvent ou non bouger entre elles. Par défaut, l'hypothèse utilisée dans les codes de calcul est une adhésion parfaite. Cela n'est pas forcément compatible avec le but recherché par le calcul, notamment si celui-ci vise à appréhender les modes de rupture (qui sont plus complexes pour les composites...).

Tous les codes de calcul actuels permettent de prendre en compte l'anisotropie matérielle et les matériaux composites (avec des interfaces plus ou moins sympathiques). Nous n'entrerons donc pas plus dans le détail dans ce document.

20.4 Le contact

Histoire

Les premiers calculs de Joseph Boussinesq, auteur en 1876 d'un *Essai théorique de l'équilibre des massifs pulvérulents, comparé à celui des massifs solides, sur la poussée des terres sans cohésion*, reprenant des études de Coulomb sur ce sujet, reposent sur un ensemble d'hypothèses très restrictives : 1) les corps en présence sont supposés semi-infinis (cela n'est vrai que si les zones de contact sont vraiment très petites par rapport aux autres dimensions) ; 2) au voisinage de la future zone de contact, leurs surfaces peuvent être représentées par des quadriques dont les courbures sont connues (or la rugosité, qui rend la répartition des pressions de contact très irrégulière, est généralement très éloignée du modèle théorique) ; 3) ces corps sont parfaitement élastiques, homogènes et isotropes (ce qui est très restrictif, et souvent réellement faux) ; 4) l'aire de contact est assimilée à un très petit élément plan qui ne reçoit que des efforts normaux, donc parallèles entre eux (dans beaucoup de contacts, la zone d'application des pressions est loin d'être plane et surtout, le fait de ne considérer que des charges normales suppose que l'on fasse abstraction du frottement).

En 1881, Heinrich Hertz, jeune ingénieur et docteur ès sciences de 24 ans, publie dans le célèbre *Journal de Crelle* (XVII, p. 156) sous le titre *Über die Berührung fester elastischer Körper* (Sur le contact des corps solides élastiques), un mémoire qui fera date, puisqu'il s'agit de la première théorie cohérente des contacts ponctuels.

Le fait de rester dans le domaine des déformations élastiques permet d'appliquer le principe de superposition : aux contraintes issues de 1) l'application des efforts normaux se superposent celles 2) provoquées par les efforts tangentiels résultant du frottement ou de l'adhérence, puis celles 3) dues aux contraintes résiduelles dont on favorise l'apparition par des traitements mécaniques ou thermochimiques appropriés, et enfin celles 4) qui correspondent aux autres sollicitations des pièces, tension, compression, flexion, torsion... Ces quatre groupes de contraintes peuvent être définis séparément puis combinés pour aboutir à l'état de charge complet des zones de contact.

Le calcul des contraintes supplémentaires dues au frottement a été conduit de diverses manières



Boussinesq

Coulomb

Hertz

par des chercheurs comme Liu (1950), Poritzky (1966) et quelques autres. Il est extrêmement compliqué, au point d'être pratiquement inutilisable dans les situations concrètes.

Depuis que Hertz a introduit une théorie du contact en 1881, de nombreux problèmes d'ingénieur faisant intervenir le contact ont été résolus. L'outil de calcul basé sur une approche analytique, est limité à la résolution des problèmes simples de contact : en effet la plupart des solutions analytiques supposent un contact sans frottement et des zones de contact connues, à priori, et des formes géométriques simples.

Le développement des techniques numériques de résolution a permis de traiter des problèmes de contact plus complexes. la méthode des éléments finis, en permettant la discrétisation des solides de formes quelconques et la prise en compte aisée de conditions aux limites diverses, offre un outil puissant de calcul pour étudier les problèmes de contact.

Aujourd'hui l'analyse des problèmes de contact avec frottement est très importante pour beaucoup d'applications industrielles. La modélisation des procédés industriels de mise en forme, et plus généralement des phénomènes complexes où le contact et le frottement s'ajoutent à des non-linéarités du matériau et de la géométrie, nécessite des algorithmes supplémentaires dans les logiciels généraux d'éléments finis.

Malgré la linéarité de la loi élastique, le problème de contact est intrinsèquement non-linéaire. En effet, la surface de contact et les forces de contact sont, à priori, inconnues et elles changent progressivement lorsqu'on applique le chargement externe.

Dans la littérature, de nombreuses méthodes ont été développées pour résoudre les problèmes de contact par des méthodes numériques comme la méthode des éléments finis, parmi lesquels la méthode de pénalisation, la méthode de flexibilité, la méthode de programmation mathématique, la méthode des multiplicateurs de Lagrange (voir le paragraphe 7.6 sur les multiplicateurs de Lagrange et le paragraphe 12.3 sur les diverses manières de traiter une interface, qui est un contact rigide). Une grande partie de ces articles traitent d'algorithmes numériques. Dans les codes éléments finis industriels (Ansys, Pamcrash...), les problèmes de contact avec frottement dans le contexte des grandes déformations sont presque exclusivement traités par des méthodes de pénalisation ou de régularisation. Ces méthodes présentent des inconvénients en ce qui concerne la stabilité et la précision numérique, en particulier pour tout ce qui touche à la simulation des phénomènes de frottement.

Pour pallier ces insuffisances, une méthode du Lagrangien Augmenté a été développée par Curnier et Alart (1988). Cette méthode consiste à déterminer les inconnues (déplacement et réaction) simultanément en utilisant un algorithme de Newton généralisé. Simo et Laursen ont également proposé une méthode similaire (1992). De Saxcé et Feng ont proposé une méthode bipotentielle fondée sur la théorie du Matériau Standard Implicite, dans laquelle une nouvelle formulation du lagrangien augmenté est développée.

Pour les problèmes de contact unilatéral avec frottement, la méthode bipotentielle n'utilise qu'un seul principe variationnel sur le déplacement et une seule inégalité. Ainsi, le contact unilatéral et le frottement sont couplés. Cette nouvelle approche étend également la notion de loi normale aux comportements dissipatifs non associés, en tenant compte du frottement. Cette approche variationnelle est plus simple que l'approche classique qui inclue deux principes variationnels et deux inégalités respectivement pour le contact unilatéral et le frottement. Dans la méthode bipotentielle, le problème de contact avec frottement est traité dans un système réduit par un algorithme d'Uzawa à une seule phase de prédiction-correction sur le cône de frottement. L'extension de cette méthode dans le contexte de grandes déformations a été réalisé. Pour être capable de traiter des problèmes industriels qui font intervenir le contact, il est important de disposer d'un éventail d'algorithmes afin de pouvoir moduler l'utilisation de chaque méthode selon leurs avantages et inconvénients dans chaque cas concret.

20.4.1 Lois de contact et de frottement

Considérons deux solides déformables Ω_1 et Ω_2 en contact en certains points de leurs frontières. Soient v la vitesse relative locale en un point P de Ω_1 par rapport à Ω_2 , et r la réaction que subit Ω_1 de la part de Ω_2 . Supposons défini n le vecteur unitaire normal dirigé vers Ω_1 . La projection orthogonale du point P sur la surface de Ω_2 définit un point P' , dit « point projeté de P », qui sera l'origine du repère local du contact (t_1, t_2, n) , comme illustré à la figure 20.2.

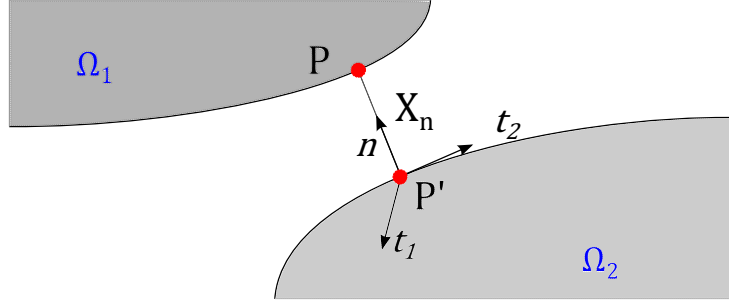


FIGURE 20.2: Contact entre deux solides déformables : notations

Les vecteurs v et r peuvent être décomposé en partie tangente v_t et r_t et partie normale v_n et r_n .

Compte tenu de la distance initiale au contact g et pour un pas de temps Δt , la distance PP' exprimée dans le repère local est alors définie par : $x_n = g + \Delta t v_n$. Par conséquent, les conditions de contact unilatéral pour chaque point en contact, peuvent être exprimées comme suit :

- Impénétrabilité : $x_n \geq 0$
- Quand la particule est en contact avec l'obstacle, elle n'est pas attirée par lui : $x_n = 0 \rightarrow r_n \geq 0$
- Quand la particule n'est pas en contact avec l'obstacle, la réaction normale est nulle : $x_n > 0 \rightarrow r_n = 0$

Ces trois conditions peuvent être condensées sous une forme complémentaire (conditions de Signorini) :

$$x_n \geq 0; \quad r_n \geq 0; \quad x_n r_n = 0 \quad (20.46)$$

ou sous la forme équivalente :

$$\forall \rho_n > 0, \quad r_n = \text{proj}_{\mathbb{R}^+}(r_n - \rho_n x_n) \quad (20.47)$$

où \mathbb{R}^+ représente un ensemble de valeurs positives.

En ce qui concerne le frottement, parmi les nombreux modèles existants, nous avons choisi le modèle de Coulomb, qui est le plus utilisé dans les problèmes de contact avec frottement sec. Dans le cas du contact-frottement isotrope, le modèle de Coulomb s'écrit :

$$\begin{cases} \|r_t\| < \mu r_n & \text{si } \|v_t\| = 0 \\ r_t = -\mu r_n \frac{v_t}{\|v_t\|} & \text{si } \|v_t\| \neq 0 \end{cases} \quad (20.48)$$

avec μ le coefficient de frottement.

On peut également utiliser la forme équivalente :

$$\forall \rho_t > 0, \quad r_t = \text{proj}_C(r_t - \rho_t u_t) \quad (20.49)$$

où C est un ensemble qui représente l'intervalle $[-\mu r_n, \mu r_n]$ dans le cas bidimensionnel, ou le disque de centre 0 et de rayon $R = \mu r_n$ dans le cas tridimensionnel. On rappelle que u_t désigne la composante tangentielle du déplacement relatif, qui est aussi le glissement.

20.4.2 Algorithme local

De manière usuelle, on définit le cône de frottement isotrope de Coulomb pour chaque point en contact :

$$K_\mu = \{(r_n, r_t) \in \mathbb{R}^3 \text{ tels que } \|r_t\| \leq \mu r_n\} \quad (20.50)$$

À l'aide de la théorie du Matériau Standard Implicite, les conditions de contact unilatéral et la loi de frottement s'expriment par une seule inégalité variationnelle :

$$\begin{aligned} &\text{Trouver } (r_n, r_t) \in K_\mu \text{ tel que :} \\ &\forall (r_n^*, r_t^*) \in K_\mu, (x_n - \mu \|v_t\|)(r_n^* - r_n) + v_t \cdot (r_t^* - r_t) \leq 0 \\ &\text{où } \exists \lambda > 0, -v_t = \lambda \frac{r_t}{\|r_t\|} \end{aligned} \quad (20.51)$$

Afin d'éviter des potentiels non différentiables qui apparaissent dans la représentation du contact, on peut utiliser la méthode du lagrangien augmenté. Cette méthode, appliquée à l'inégalité variationnelle, conduit à une équation implicite :

$$r = \text{proj}_{K_\mu}(r_n + \rho_n(x_n - \mu \|v_t\|), \quad r_t + \rho_t v_t) \quad (20.52)$$

où ρ_n et ρ_t sont des coefficients positifs qui dépendent de la matrice de flexibilité de contact.

Pour résoudre l'équation implicite, on utilise l'algorithme d'Uzawa (algorithme de gradient à pas fixe pour l'optimisation sous contraintes) qui conduit à une procédure itérative comportant les deux étapes suivantes :

— prédiction :

$$\begin{cases} \tau_n^{i+1} = \tau_n^i + \rho_n^i(x_n^i - \mu \|v_t^i\|) \\ \tau_t^{i+1} = \tau_t^i = \rho_t^i v_t^i \end{cases} \quad (20.53)$$

— correction :

$$(r_n^{i+1}, r_t^{i+1}) = \text{proj}_{K_\mu}(\tau_n^{i+1}, \tau_t^{i+1}) \quad (20.54)$$

contact avec adhérence ($\tau \in K_\mu$), non contact ($\tau \in K_\mu^*$) et contact avec glissement ($\tau \in \mathbb{R}^3 - (K_\mu \cup K_\mu^*)$), où K_μ^* est le cône dual de K_μ .

20.4.3 Algorithme global

Dans le contexte de la méthode des éléments finis, après discrétisation des solides en contact, on résout généralement un système d'équations d'équilibre au niveau global, obtenu par la méthode des résidus pondérés (exposée au paragraphe 7.1), de type :

$$\mathbf{R}^* = -\mathbf{F}_{\text{int}} + \mathbf{F}_{\text{ext}} + \mathbf{F}_{\text{reac}} = \mathbf{0} \quad (20.55)$$

où les vecteurs \mathbf{R}^* des résidus, \mathbf{F}_{int} des forces internes, \mathbf{F}_{ext} des forces extérieures et \mathbf{F}_{reac} des forces de contact et de frottement dans le repère global dépendent tous du vecteur des inconnues nodales de la structure \mathbf{q} .

Pour résoudre ce système d'équations non linéaires, on utilise une méthode itérative de type Newton-Raphson qui consiste à linéariser le système précédent en :

$$\mathbf{K}_T^i \mathbf{d}^{i+1} = -\mathbf{F}_{\text{int}}^i + \mathbf{F}_{\text{ext}}^i + \mathbf{F}_{\text{reac}}^i \quad (20.56)$$

$$\mathbf{q}^{i+1} = \mathbf{q}^i + \mathbf{d}^{i+1} \quad (20.57)$$

où \mathbf{K}_T^i est la matrice tangente de rigidité à l'itération i .

Insistons que le caractère fortement non-linéaire de ce système : en effet, celui-ci prend en compte des multiples non-linéarités mécaniques telles que les non-linéarités matérielles et les non-linéarités géométriques qui se manifestent lorsqu'apparaissent des grands déplacements ou des grandes déformations. De plus, les lois relatives au contact et au frottement sont exprimées par des inégalités, le potentiel du contact étant même non différentiable. Par conséquent, des difficultés numériques se manifestent à plusieurs niveaux au cour de :

- la résolution des équation non linéaires d'équilibre (niveau global) ;
- l'intégration des lois de comportement (niveau local) ;
- la résolution des inéquations de contact et de frottement couplées avec les équations d'équilibre (niveaux local et global).

20.5 Exemple : une toute première approche du contact avec CAST3M

Dans ce chapitre, nous présentons un petit exemple simple de contact unidirectionnel sans frottement sous CAST3M.

20.5.1 Contact sur une surface infiniment rigide

Dans ce premier exemple, nous proposons d'étudier un carré soumis sur sa face supérieure à un déplacement imposé et reposant, sur sa face inférieure, sur une surface infiniment rigide, comme illustré à la figure 20.3.

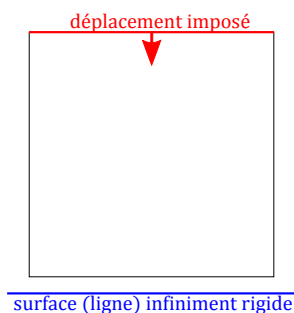


FIGURE 20.3: Problème de contact

```

1  OPTION ECHO 0 ;
2  OPTION DIME 2 ELEM QUA4 MODE PLAN CONT
3  ;
4  * Donnees
5  long1=10.;
6  nlong1=7;
7  uy0=-0.5;
8  haut1=long1;
9  nhaut1=nlong1;
10 jeu1=0.0;
11 *
12 * droite sur laquelle le carre viendra
13   buter :
14 k1 = (-1.*long1) 0.;
15 k2 = long1 0.;
16 l1 = DROI (nlong1-2) k2 k1;
17 *
18 * carre :
19 k3 = (-0.5*long1) jeu1;
20 k4 = (0.5*long1) jeu1;
21 l2 = DROI nlong1 k3 k4;
22 s1 = l2 TRAN nhaut1 (0. haut1);

```

On remarquera les choses suivantes :

- la ligne l_2 correspondant à la face inférieure du carré « repose », sans jeu, sur la ligne l_1 définissant la surface rigide sur laquelle le carré va venir buter (on aurait pu mettre un jeu initial non nul, mais ce n'est pas nécessaire) ;
- les orientations des lignes l_1 et l_2 sont inversées car les deux lignes en contact doivent se « tourner le dos » au sens des normales. Cela est obligatoire ;

- La surface de contact l_1 est modélisée en utilisant $nlong_1 - 2$ éléments pour le maillage, alors que le carré (ligne l_2) n'en a que $nlong_1$, afin que les maillages ne soient pas « compatibles » (i.e. que les nœuds ne tombent pas en face les uns des autres).

```

22 * Modele :
23 ModM1 = MODEL s1 MECANIQUE ELASTIQUE;
24 MatM1 = MATER ModM1 'YOUN' 1.E3 'NU'
    0.3 ;
25 *
26 * CL en déplacements (Rigidites):
27 * 1) il faut une condition selon UX:
28 *     on prend le point (0,0)
29 k5 = s1 POIN PROC (0. jeu1);
30 CLk5 = BLOQ k5 UX;
31 * 2) ligne de contact: rigidites

```

```

32 Dep11 Rigid1 = IMPO 12 11;
33 Rigid2 = BLOQ DEPL 11;
34 * 3) ligne sur laquelle on va imposer
    le déplacement
35 13 = s1 'COTE' 3;
36 CL13 = BLOQ 13 UY;
37 DCL13 = DEPI CL13 UY0;
38 * Totalite des conditions aux limites
39 CL0 = CLk5 ET CL13 ET Rigid1 ET Rigid2
    ;

```

Nous sommes en élasticité linéaire... forces et déplacements sont proportionnels et il n'y a aucun problème de convergence. La résolution se fait le plus simplement du monde.

```

40 * RESOLUTION
41 MR1=RIGID ModM1 MatM1;
42 dep1 = RESO (MR1 ET CL0) DCL13;

```

Puis on fait un peu d'affichage des résultats... ce qui est illustré figure 20.4.

```

43 * Post-Traitement
44 *
45 * deformee:
46 defo0 = DEFO (s1 ET 11) dep1 0. 'BLEU'
    ;
47 defo1 = DEFO (s1 ET 11) dep1 1. 'ROUG'
    ;
48 TITR 'Maillages non deforme (bleu) et
    deforme (rouge)' ;
49 TRAC (defo0 ET defo1) ;
50 *
51 * Deplacement selon Uy

```

```

52 TITR 'Champ de déplacements.' ;
53 Dep1Y1 = EXCO Dep1 UY;
54 TRAC Dep1Y1 s1;
55 *TRAC Dep1Y1 defo1;
56 *
57 * Contraintes:
58 Sig1 = SIGM ModM1 MatM1 Dep1;
59 TITR 'Champ de contraintes' ;
60 TRAC sig1 ModM1;
61 *
62 fin;

```

On pourra ensuite mettre jeu_1 a une valeur différente de zéro et refaire le calcul. On s'apercevra que cela fonctionne toujours. La visualisation de la déformée permettra de bien comprendre comment s'effectue le calcul (on rappelle que dans CAST3M, toutes les conditions aux limites sont introduites par l'intermédiaire de multiplicateurs de Lagrange : voir paragraphe 12.6.4).

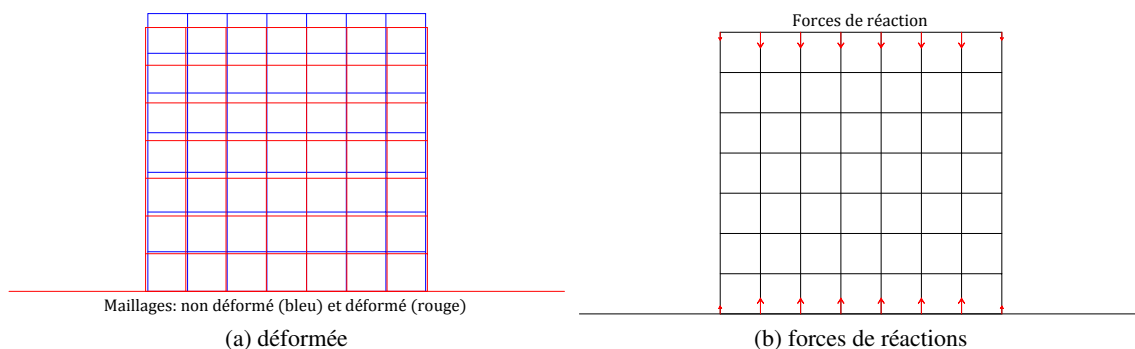


FIGURE 20.4: Premiers résultats sur le contact

20.5.2 Contact entre deux solides

On peut faire une remarque sur le calcul précédent : que se passe-t-il si la surface sur laquelle on s'appuie n'est plus infiniment rigide ?

On pourrait modéliser la partie inférieure et lui appliquer les forces de réactions qui ont été calculées précédemment. Nous n'allons évidemment pas procéder ainsi et laisser CAST3M tout calculer pour nous...

```

1  OPTION ECHO 0 ;
2  OPTION DIME 2 ELEM QUA4 MODE PLAN CONT
   ;
3  *
4  * Donnees
5  long1=10.;
6  nlong1=7;
7  uy0=-0.5;
8  haut1=long1;
9  nhaut1=nlong1;
10 jeu1=1.0;
11 *
12 * pave sur lequel le carre viendra
   buter:
13 k1 = (-1.*long1) 0.;
14 k2 = long1 0.;
15 l1 = DROI (2*nlong1-1) k2 k1;
16 s2 = l1 TRAN nhaut1 (0. (-1.0*haut1));
17 *
18 * carre
19 k3 = (-0.5*long1) jeu1;
20 k4 = (0.5*long1) jeu1;
21 l2 = DROI nlong1 k3 k4;
22 s1 = l2 TRAN nhaut1 (0. haut1);
23 trac (s1 et s2);
24 *
25 * Modele
26 ModM1 = MODEL s1 MECANIQUE ELASTIQUE;
27 MatM1 = MATER ModM1 'YOUN' 1.E3 'NU'
   0.3 ;
28 ModM2 = MODEL s2 MECANIQUE ELASTIQUE;
29 MatM2 = MATER ModM2 'YOUN' 5.E3 'NU'
   0.3 ;
30 *
31 * CL (Rigidites)
32 * 1) il faut une condition selon UX
33 k5 = s1 POIN PROC (0. jeu1);
34 CLk5 = BLOQ k5 UX;
35 * 2) ligne sur laquelle on impose le
   déplacement
36 l3 = s1 'COTE' 3;
37 CLl3 = BLOQ l3 UY ;
38 DCLl3 = DEPI CLl3 UY0;
39 * 3) encastrement sous la partie basse
40 l4 = s2 'COTE' 3;
41 CLl4 = BLOQ l4 DEPL;
42 * Totalite des conditions aux limites
43 CLO = CLk5 ET CLl3 ET CLl4;
44 *
45 * RESOLUTION
46 Dep11 Rigid1 = IMPO l2 l1;
47 MR1 = RIGID ModM1 MatM1;
48 MR2 = RIGID ModM2 MatM2;
49 dep1 = RESO (MR1 ET MR2 ET CLO ET
   Rigid1) DCLl3;
50 *
51 * Post-Traitement
52 *
53 * deformee:
54 defo0 = DEFO (s1 ET s2) dep1 0. 'BLEU'
   ;
55 defo1 = DEFO (s1 ET s2) dep1 1. 'ROUG'
   ;
56 TITR 'Maillages non deforme (bleu) et
   deforme (rouge)' ;
57 TRAC (defo0 ET defo1) ;
58 *
59 TITR 'Champ de déplacements.' ;
60 Dep1Y1 = EXCO Dep1 UY;
61 TRAC Dep1Y1 (s1 et s2);
62 *
63 * Comparaison des champs de
   contraintes:
64 Sig1 = SIGM (ModM1 ET ModM2) (MatM1 ET
   MatM2) Dep1;
65 TITR 'Champ de contraintes.' ;
66 TRAC sig1 (ModM1 ET ModM2);
67 *
68 fin;

```

Cette fois-ci, nous avons mis jeu_1 a une valeur non nulle pour bien comprendre comment se fait la résolution de ce système numérique.

On obtient les résultats illustrés à la figure 20.5 pour la déformée et les forces de réactions et à la figure 20.6 pour les déplacements et contraintes.

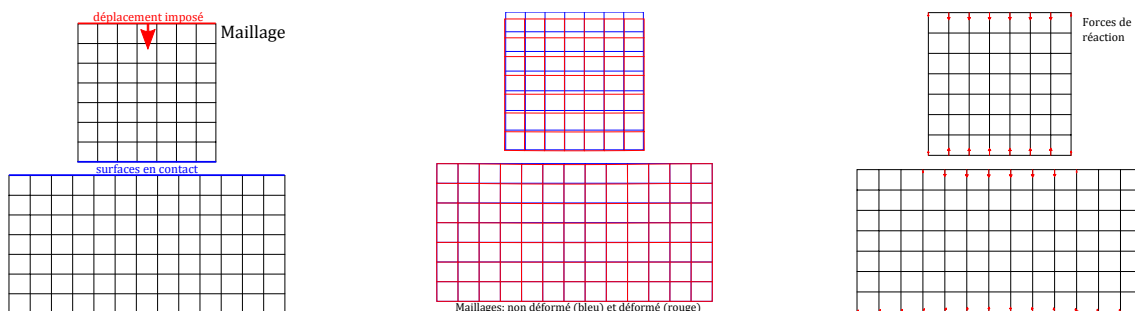


FIGURE 20.5: a) maillage, b) déformée c) forces de réaction

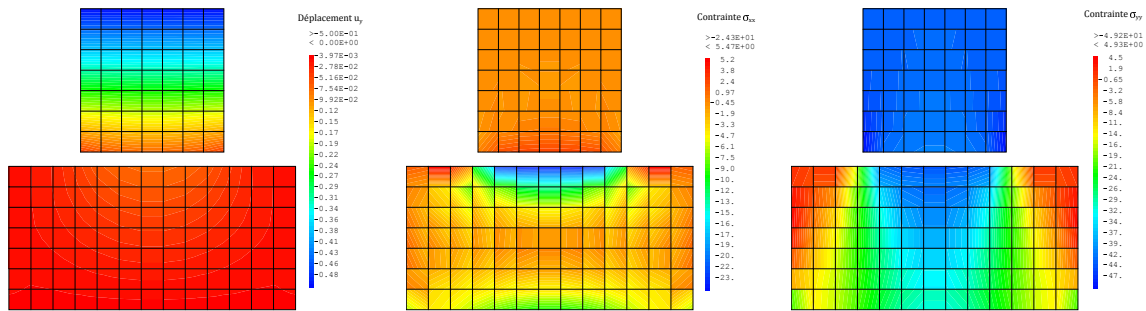


FIGURE 20.6: a) déplacement u_y b) σ_{xx} et c) σ_{yy}

20.5.3 Résolution pas à pas

Les résolutions présentées ci-dessus étaient finalement de type classique, i.e. inversion d'un système comprenant les conditions aux limites et le chargement.

Cela fonctionnait particulièrement bien car nous étions en élasticité linéaire... mais que se passerait-il si le carré considéré avait un comportement non linéaire, en particulier s'il devait être le siège de déformations permanentes ?

On se doute aisément que le modèle rudimentaire ne fonctionnerait pas. C'est pourquoi nous allons le modifier afin d'introduire une résolution pas à pas, qui permet d'appliquer un chargement (ou un déplacement imposé) de manière progressive.

Nous ne présentons pas ici de comportement non linéaire pour les matériaux, cela se fait en TD. Toutefois, nous présentons le même calcul que précédemment, mais codé pour une résolution pas à pas.

Afin de définir le chargement dans le temps, nous commençons par construire la liste de réels $Ltemp_1$ qui contient les pas de temps, à savoir 0.0 et 1.0 (2 pas de temps suffisent, puisque nous restons sur une analyse élastique linéaire dans cet exemple).

À partir de cette discrétisation du temps (vraiment sommaire pour le coup), nous définissons l'objet evo_1 de type évolution pour chacun des pas de temps. evo_1 est donc une fonction du temps, qui va nous servir à affecter le déplacement imposé : $CharU_1$ est de type chargement, nécessaire pour la résolution par PASAPAS. Avec l'option DIMP, on signifie qu'il s'agit d'un déplacement imposé. Ainsi $CharU_1$ contient l'évolution des déplacements imposés sur la condition DCL13 au cours du temps.

```

1  OPTION ECHO 0 ;
2  OPTION DIME 2 ELEM QUA4 MODE PLAN CONT
   ;
3  *
4  * Donnees
5  long1=10.;
6  nlong1=7;
7  uy0=-0.5;
8  haut1=long1;
9  nhaut1=nlong1;
10 jeu1=0.0;
11 *
12 * pave sur lequel le carre viendra
    buter :
13 k1 = (-1.*long1) 0.;
14 k2 = long1 0.;
15 l1 = DROI (2*nlong1-1) k2 k1;
16 s2 = l1 TRAN nhaut1 (0. (-1.0*haut1));
17 *
18 * carre
19 k3 = (-0.5*long1) jeu1;
20 k4 = (0.5*long1) jeu1;
21 l2 = DROI nlong1 k3 k4;

22 s1 = l2 TRAN nhaut1 (0. haut1);
23 TRAC (s1 ET s2);
24 *
25 *
26 * Modele
27 ModM1 = MODEL s1 MECANIQUE ELASTIQUE;
28 MatM1 = MATER ModM1 'YOUN' 1.E3 'NU'
    0.3 ;
29 ModM2 = MODEL s2 MECANIQUE ELASTIQUE;
30 MatM2 = MATER ModM2 'YOUN' 5.E3 'NU'
    0.3 ;
31 *ModM1 = ModM1 ET ModM2;
32 *
33 * CL (Rigidites)
34 * 1) il faut une condition selon UX,
    on prend le point (0,0)
35 k5 = s1 POIN PROC (0. jeu1);
36 CLk5 = BLOQ k5 UX;
37 * 2) ligne sur laquelle on va imposer
    le déplacement
38 l3 = s1 'COTE' 3;
39 CLl3 = BLOQ l3 UY ;
40 DCLl3 = DEPI CLl3 UY0;
41 * 3) encastrement sous la partie basse

```

```

42 l4 = s2 'COTE' 3;
43 CL14 = BLOQ L4 DEPL;
44 * 4) ligne de contact:
45 Dep1 Rigid1 = IMPO 12 11;
46 *
47 * Totalite des conditions aux limites
48 CL0 = CLk5 ET CL13 ET CL14 ET Rigid1;
49 *
50 * Chargement
51 Ltemps1 = PROG 0. 1.;
52 evo1 = EVOL 'MANU' 'TEMPS' Ltemps1 (
    PROG 0. 1.) ;
53 CharU1 = CHAR DIMP DCL13 evo1 ;
54 Char0 = CharU1 ;
55 *
56 * RESOLUTION
57 *
58 * Construction de la table PASAPAS:
59 TAB1 = TABL;
60 TAB1.'TEMPS_CALCULES' = Ltemps1;
61 TAB1.'MODELE' = (ModM1 ET ModM2)
    ;
62 TAB1.'CARACTERISTIQUES' = (MatM1 ET
    MatM2);
63 TAB1.'BLOCAGES_MECHANQUES' = CL0;
64 TAB1.'CHARGEMENT' = Char0;
65 *
66 * Resolution:
67 TAB2 = PASAPAS TAB1 ;
68 *
69 * Post-Traitement
70 *
71 dep1 = TAB2 . 'DEPLACEMENTS' . 1;
72 *
73 * deformee:
74 defo0 = DEFO (s1 ET s2) dep1 0. 'BLEU'
    ;
75 defo1 = DEFO (s1 ET s2) dep1 1. 'ROUG'
    ;
76 TITR 'Maillages non deforme (bleu) et
    deforme (rouge).';
77 TRAC (defo0 ET defo1) ;
78 *
79 TITR 'Champ de déplacements.' ;
80 Dep1Y1 = EXCO Dep1 UY;
81 TRAC Dep1Y1 (s1 et s2);
82 *
83 * Comparaison des champs de
    contraintes:
84 sig1 = TAB2 . 'CONTRAINTES' . 1 ;
85 TITR 'Champ de contraintes.' ;
86 TRAC sig1 (ModM1 ET ModM2);
87 *
88 * Visualisations des reactions:
89 reac1 = TAB2 . 'REACTIONS' . 1 ;
90 vr1 = VECT reac1 0.8E-2 'FX' 'FY' ,
    'ROUG' ;
91 TITR 'Forces de reaction.' ;
92 TRAC vr1 (s1 ET s2) ;
93 *
94 fin;

```

On obtient les mêmes résultats que ceux déjà présentés (voir figure 20.5 et figure 20.6).

Nous pouvons mentionner également que ce type de calcul permet de calculer la géométrie de pièces réalisées en thermocompression. Considérons un matelas de matière souple placé entre le plateau supérieur d'une presse et un outil comportant un logement, comme illustré à la figure 20.7a. Une fois que le plateau de presse est descendu, on obtient la pièce en forme donnée à la figure 20.7b.

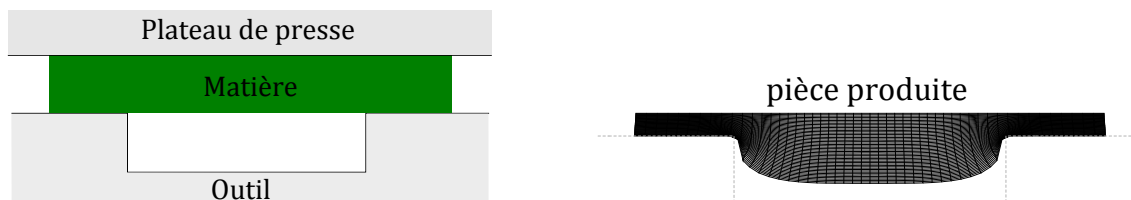


FIGURE 20.7: Thermocompression : a) problème, b) forme de la pièce après production

À partir des listings précédents, il est aisé d'obtenir les résultats de la figure 20.7b... et nous vous invitons à le faire.

Une fois ce premier calcul réalisé (et donc pleins de fierté et de confiance en nous), nous vous proposons d'utiliser un matelas de matière plus épais, vraiment plus épais. Le but est que lors du process, lorsque le plateau de la presse descend (situation réelle ou simulée), la matière vienne suffisamment remplir la cavité du moule pour qu'il y ait contact entre la pièce déformée et le fond de la cavité.

Nous allons alors constater que le contact n'est pris en compte que dans la partie supérieure du moule, mais pas sur les côtés de la cavité, ni au fond de celle-ci...

En TP, nous travaillerons à comprendre pourquoi (qu'est-ce que le code de calcul a compris de ce que nous lui avons proposé comme modélisation ?) et à réaliser un modèle correspondant à ce que nous souhaitons.

Chapitre 21

La rupture en mécanique

La mécanique de la rupture s'intéresse à la formation et à la propagation de fissures macroscopiques dans les structures, ce qui peut conduire à la séparation d'un corps en deux parties disjointes suite à une phase d'amorçage qui a vu le développement de microcavités, microfissures... mais les fissures peuvent aussi s'arrêter. Le mode de rupture peut être fragile (sans déformation plastique) ou ductile (avec déformation plastique).

La résilience est le rapport de l'énergie nécessaire pour rompre une pièce sur la section droite de matière rompue : elle caractérise l'énergie nécessaire pour produire la rupture. La résilience évolue avec la température (la température de transition caractérise le passage d'un mode de rupture à un autre : fragile ou ductile). Par ailleurs, le mode de rupture dépend de l'état de contrainte, en particulier de la *triaxialité des contraintes* (rapport du premier sur le second invariant, voir paragraphe 20.1.1) : un matériau très plastique peut développer des ruptures fragiles ; un matériau sans plasticité ne présentera que des ruptures fragiles.

En fonction du chargement et du matériau considéré :

- si le milieu est globalement plastique ou viscoplastique, on recourra à la mécanique non linéaire de la rupture, ou approche locale (description fine des contraintes et déformations en pointe de fissure à l'aide de modèles non linéaires) ;
- si la plasticité est absente ou très confinée, on utilisera la mécanique linéaire de la rupture.

21.1 Approches globale et locale

On distingue deux approches :

- Approche globale : le principal avantage est la relative simplicité d'application aux calculs éléments finis ou analytiques grâce à des *grandeurs scalaires représentant l'état de la fissure*, mais qui sont des approximations qui peuvent manquer de pertinence ;
- Approche locale : qui veut *modéliser finement le comportement réel*, mais se confronte à la singularité des contraintes en fond de fissure.

D'un point de vue numérique, les deux approches nécessitent des maillages fins en pointe de fissure afin de décrire le front de fissure et de permettre le calcul de la concentration des contraintes et déformations. Il faut ajouter également des modèles de comportement adaptés :

- l'approche globale nécessite un maillage rayonnant avec faces perpendiculaires au front de fissure ;
- alors que l'approche locale impose un maillage fin (calcul des contraintes), dont la taille est imposée ainsi que le type d'éléments.

Les fissures sont souvent trop complexes pour être parfaitement reproduites, ce qui impose des approximations du domaine et rend le choix du type d'éléments et leur taille encore plus critique.

La modélisation du comportement des matériaux est importante puisque directement liée au calcul des sollicitations en pointes de fissure, qui de plus apparaissent toujours dans des zones où

les sollicitations sont complexes (fatigue, fluage, plasticité, endommagement...).

La *modélisation de l'avancée de la fissure* conduit à de très fortes non linéarités (avec des problèmes de convergence dans les algorithmes). De plus, la modélisation de l'avancée de fissure est liée à la discrétisation faite, donc est liée à la taille des éléments.

Les mailleurs automatiques sont souvent incapables de mener à bien leur tâche car il y a une incompatibilité entre les dimensions de la structure et la finesse nécessaire pour la fissure, et parce que les approches locale et globale ont des spécificités de maillage différentes et difficiles à prendre en compte. Il est donc souvent nécessaire de développer ses propres outils de maillage. De plus beaucoup de critères mis au point en 2D ne sont pas forcément valables en 3D.

Les problèmes actuels sur lesquels portent la recherche sont :

- la *propagation*, qui reste un domaine numérique difficile à aborder, en particulier les *critères de propagation* de fissure, surtout en 3D et/ou en dynamique.
- les *mailleurs* (voir ci-dessus), et même en amont, les *critères de maillage*, surtout en 3D.

On opte également pour une approche couplée essais / calculs :

- développement des paramètres pertinents en mécanique de la rupture ;
- développement d'outils analytiques et critères associés ;
- développement / validation des critères.

Approches numériques des fissures

Fissure droite sollicitée en mode I Le trajet de propagation peut être connu *a priori*, et on trouve les méthodes de déboutonnage dans lesquelles la moitié de la structure est maillée et les nœuds situés sur la ligne de propagation sont libérés à mesure que la fissure avance. Mais il faut relâcher les nœuds progressivement en appliquant des forces nodales physiquement pertinentes.

Fissure courbe

- méthodes de remaillage : voir problème de maillage
- éléments d'interface : le trajet de la fissure est imposé par la discrétisation et il faut choisir la loi de décohesion à l'interface.

Élément avec un nœud supplémentaire au quart de ses côtés (ou ayant son nœud intermédiaire déplacé), qui permet d'intégrer exactement la singularité élastique, mais il reste nécessaire de disposer d'un outil spécifique de remaillage.

Méthodes sans maillage Elles permettent de s'affranchir des problèmes liés à la connaissance trop intime de la fissure ;

Méthodes utilisant la partition de l'unité Elles permettent de s'affranchir du maillage explicite de la fissure dont la description se fait au moyen d'éléments géométriques ou de fonctions de niveau pour le problème 3D.

Notons que les méthodes de remaillage peuvent introduire des singularités dans la discrétisation temporelle qui sont peu souvent mentionnées.

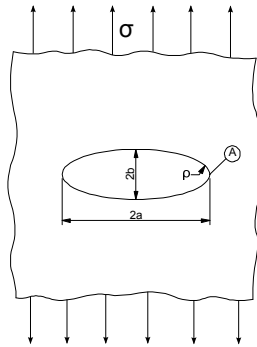
21.2 Mécanique linéaire de la rupture

21.2.1 Concentrations de contraintes des défauts

Histoire

On a souvent attribué aux défauts du matériau la cause principale de la rupture fragile. Sur la base d'une analyse des contraintes, Kirsch (1898) et Inglis (1913, cité au 8ème symposium de mécanique des roches en 1966) avaient déjà donné des solutions analytiques pour le calcul du facteur de concentration des contraintes pour des plaques infinies soumises à la traction avec respectivement un trou circulaire et un trou elliptique.

Mais le facteur de concentration des contraintes devenait infini dans le cas d'une fissure et cela signifiait que des contraintes externes très faibles suffisaient pour la rupture d'un solide fissuré, ce qui est en contradiction avec la réalité.



Inglis, donc, en 1913, a montré l'effet de concentration de contraintes.

Considérons une plaque avec un trou elliptique, de grand axe $2a$ et petit axe $2b$, dont la largeur ℓ est très supérieure aux dimensions du trou (i.e. $\ell \gg 2a$) et soumise à une contrainte nominale constante σ , comme illustré sur la figure 21.1.

La contrainte en pointe de fissure (i.e. au point A) est :

$$\sigma_A = \sigma \left(1 + \frac{2a}{b} \right) \quad (21.1)$$

FIGURE 21.1: Concentration de contraintes autour d'une fissure : cas plan

Le facteur de concentration de contraintes k_f est défini comme $k_f = \sigma_A / \sigma$:

- si $a = b$, alors $k_f = 3$;
- si $a \rightarrow \infty$, Inglis propose $\sigma_A = \sigma(1 + 2\sqrt{a/\rho})$, où $\rho = b^2/A$ est le rayon en pointe de fissure ;
- si $a \gg b$, alors $\sigma_A = 2\sigma\sqrt{\frac{a}{\rho}}$.

21.2.2 Équilibre énergétique

Histoire

La mécanique de la rupture a été inventée pendant la Première Guerre mondiale par l'ingénieur aéronautique anglais A. A. Griffith pour expliquer la rupture des matériaux fragiles.

Le travail de Griffith a été motivé par deux faits contradictoires :

1. La contrainte nécessaire pour rompre un verre courant est d'environ 100 MPa ;
2. La contrainte théorique nécessaire à la rupture de liaisons atomiques est d'environ 10 000 MPa.

Une théorie était nécessaire pour concilier ces observations contradictoires. En outre, les expérimentations sur les fibres de verre que Griffith lui-même a mené suggèrent que la contrainte de rupture augmente d'autant plus que le diamètre des fibres est petit. Par conséquent il en déduit que le paramètre de résistance uniaxiale à la rupture R_f , utilisé jusqu'alors pour prédire les modes de défaillance dans le calcul des structures, ne pourrait pas être une valeur indépendante des propriétés du matériau.

Griffith suggère que la faiblesse de la résistance à la rupture observée dans ses expériences, ainsi que la dépendance de l'intensité de cette résistance, étaient due à la présence de défauts microscopiques préexistants dans le matériau courant.

Pour vérifier l'hypothèse de défauts préexistants, Griffith a introduit une discontinuité artificielle dans ses échantillons expérimentaux. La discontinuité artificielle était une forme de fissure débouchante plus importante que les autres discontinuités supposées préexistantes dans l'échantillon.

Les expériences ont montré que le produit de la racine carrée de la longueur du défaut et de la contrainte à la rupture était à peu près constant.



Griffith

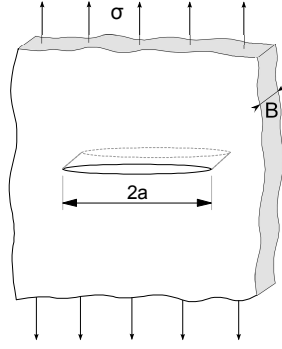


FIGURE 21.2: Concentration de contraintes autour d'une fissure : plaque mince

En 1920, Griffith considère le cas d'une plaque d'épaisseur B contenant une fissure de longueur $2a$ et dont la largeur ℓ est très supérieure aux dimensions de la fissure (i.e. $\ell \gg 2a$) et à son épaisseur (i.e. $\ell \gg B$), soumise à une contrainte nominale constante σ (voir figure 21.2). Sous l'hypothèse de contrainte plane, Griffith montre que l'équilibre énergétique est :

$$\frac{dE}{dA} = \frac{d\Pi}{dA} + \frac{dW_s}{dA} = 0 \quad (21.2)$$

où $A = 2aB$ est l'aire de la fissure, dA l'accroissement de fissure, E l'énergie totale, Π l'énergie potentielle et W_s l'énergie nécessaire pour la progression du défaut.

Griffith montre également que l'énergie potentielle Π est reliée à l'énergie potentielle de la plaque sans défaut par la relation :

$$\Pi = \Pi_0 - \left(\frac{\pi a \sigma^2}{E} \right) 2aB \quad (21.3)$$

et que W_s est donnée par :

$$W_s = 4aV\gamma_s \quad (21.4)$$

avec γ_s l'énergie de progression du défaut par unité de surface. L'équation d'équilibre énergétique de Griffith conduit à :

$$-\frac{d\Pi}{dA} = \frac{dW_s}{dA} \quad (21.5)$$

et donc :

$$\frac{\pi a \sigma^2}{E} = 2\gamma_s \quad (21.6)$$

Griffith obtient donc la contrainte globale de rupture σ_f pour les matériaux fragiles (et uniquement eux) :

$$\sigma_f = \sqrt{\frac{2E\gamma_s}{\pi a}} \quad (21.7)$$

Ces travaux seront complétés par Irwin afin de prendre en compte le cas de la rupture ductile (pour les matériaux métalliques). Ce dernier obtient la contrainte globale de rupture :

$$\sigma_f = \sqrt{\frac{2E(\gamma_s + \gamma_p)}{\pi a}} \quad (21.8)$$

où γ_p est l'énergie plastique de progression de fissure par unité de surface (i.e. en rupture fragile $W_f = \gamma_s$, et en rupture ductile $W_f = \gamma_s + \gamma_p$).

21.2.3 Taux d'énergie libre G

Histoire

L'œuvre de Griffith a été largement ignorée par la communauté des ingénieurs jusqu'au début des années 1950. Les raisons semblent être que, pour les matériaux employés dans la réalisation des structures, le niveau réel d'énergie nécessaire pour causer la rupture est de plusieurs ordres de grandeur supérieur à l'énergie de surface correspondante et que, dans les matériaux de construction il y a toujours des déformations élastiques en fond de fissure ce qui rend l'hypothèse du milieu élastique linéaire avec contraintes infinie en pointe de la fissure tout à fait irréaliste.

La théorie de Griffith concorde parfaitement avec les données expérimentales sur des matériaux fragiles tels que le verre. Pour des matériaux ductiles tels que l'acier, l'énergie de surface prédite par la théorie de Griffith est souvent irréaliste. Un groupe de travail dirigé par G. R. Irwin à l'US Naval Research Laboratory, constitué durant la Seconde Guerre mondiale, a réalisé que la plasticité doit jouer un rôle important dans la rupture des matériaux ductiles.

Dans les matériaux ductiles (et même dans des matériaux qui semblent être fragiles), une zone plastique se développe en front de fissure. L'augmentation de la dimension de la zone plastique est fonction de l'augmentation de la charge jusqu'à ce que la fissure se propage libérant les contraintes en arrière du fond de fissure. Le cycle de chargement/libération de chargement plastique aux abords du front de fissure conduit à la dissipation d'énergie comme le ferait un traitement thermique de relaxation de contrainte. Par conséquent, un terme dissipatif doit être ajouté à la relation de l'équilibre énergétique tel qu'élaborée par Griffith pour les matériaux cassants. En termes physiques, de l'énergie supplémentaire est nécessaire pour que la propagation des fissures se produise dans les matériaux ductiles si on les compare aux matériaux fragiles.



Irwin

La stratégie d'Irwin a été de partitionner l'énergie en :

1. énergie stockée en déformation élastique (effet ressort) qui se libère lors de la propagation d'une fissure et ;
2. énergie dissipée qui comprend la dissipation plastique et l'énergie de surface (et toutes les autres forces dissipatives qui peuvent être au travail).

Poursuivant ses travaux, Irwin propose en 1957 une mesure énergétique G pour caractériser la rupture :

$$G = \frac{d\Pi}{dA} \quad (21.9)$$

En négligeant l'énergie cinétique, la puissance disponible pour ouvrir une fissure de surface A est égale à la variation d'énergie potentielle totale, résultat de la variation de l'énergie élastique stockée dans la structure et de la variation d'énergie liée aux forces extérieures. Cette contribution mécanique est appelée taux de restitution d'énergie.

G est la quantité d'énergie permettant un accroissement de fissure de dA et est aussi appelé force d'expansion de fissure ou taux de restitution d'énergie.

En revenant au cas traité par Griffith, on obtient en contraintes planes :

$$G = \frac{\pi a \sigma^2}{E} \quad (21.10)$$

et ainsi, une fissure va progresser si G atteint une valeur critique G_c :

$$G_c = \frac{dW_s}{dA} = 2\gamma_s \quad (21.11)$$

où γ_s est l'énergie de progression du défaut par unité de surface. G_c est la mesure de ténacité à la rupture du matériau.

21.2.4 Facteur d'intensité de contrainte K

Histoire

Une autre réalisation importante du groupe de travail dirigé par Irwin a été de trouver une méthode de calcul de la quantité d'énergie disponible pour une fracture au niveau de la contrainte asymptotique et les champs de déplacement autour d'un front de fissure dans un solide idéalement élastique.

C'est le facteur d'intensité de contrainte.

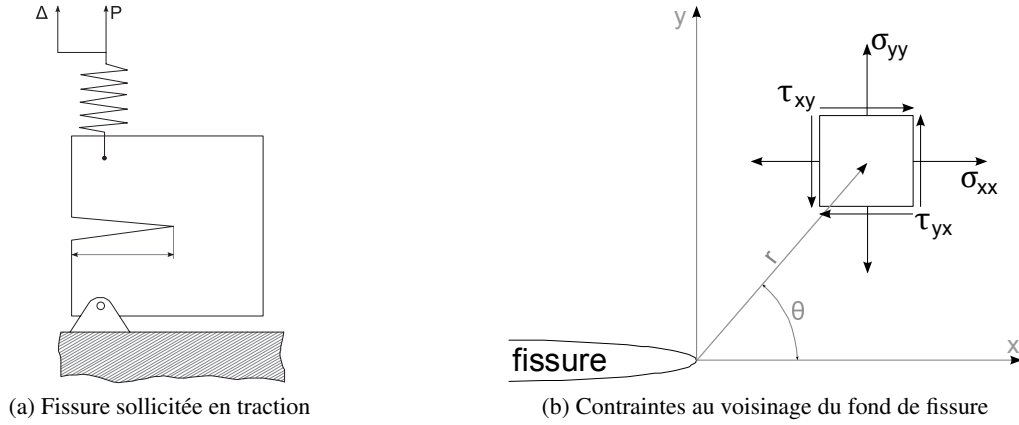


FIGURE 21.3: Calcul du facteur d'intensité de contrainte

Dans le cas de structures où une fissure se retrouve sollicitée en traction, tel que cela est illustré à la figure 21.3a, on peut montrer que :

$$\sigma_{ij} = \frac{k}{\sqrt{r}} f_{ij}(\theta) + \sum_{m=0}^{\infty} A_m r^{m/2} g_{ij}(\theta) \quad (21.12)$$

où $f_{ij}(\theta)$ est une fonction adimensionnelle, et (r, θ) les coordonnées polaires en fond de fissure. En posant :

$$K = k \sqrt{2\pi} \quad (21.13)$$

Irwin montre en 1957 qu'au voisinage du fond de fissure l'état de contraintes décrit à la figure 21.3b conduit à :

$$\lim_{r \rightarrow 0} \sigma_{ij}^{(I)} = \frac{K_I}{\sqrt{2\pi r}} f_{ij}^{(I)}(\theta) \quad (21.14)$$

$$\lim_{r \rightarrow 0} \sigma_{ij}^{(II)} = \frac{K_{II}}{\sqrt{2\pi r}} f_{ij}^{(II)}(\theta) \quad (21.15)$$

$$\lim_{r \rightarrow 0} \sigma_{ij}^{(III)} = \frac{K_{III}}{\sqrt{2\pi r}} f_{ij}^{(III)}(\theta) \quad (21.16)$$

trois modes de chargements pouvant s'appliquer à une fissure. Ces trois modes sont donnés à la figure 21.4.

Pour un mode mixte général il vient :

$$\sigma_{ij}^{(total)} = \sigma_{ij}^{(I)} + \sigma_{ij}^{(II)} + \sigma_{ij}^{(III)} \quad (21.17)$$

La détermination de K peut se faire analytiquement uniquement pour des géométries simples. Pour des cas plus complexes, il faut le faire numériquement ou expérimentalement :

- par les contraintes : par exemple K_I sera pris égale à la valeur obtenue (post-traitée) de $\sigma_{yy} \sqrt{2\pi r}$ en fond de fissure. Cette méthode simple et directe est peu précise (car on utilise les contraintes).

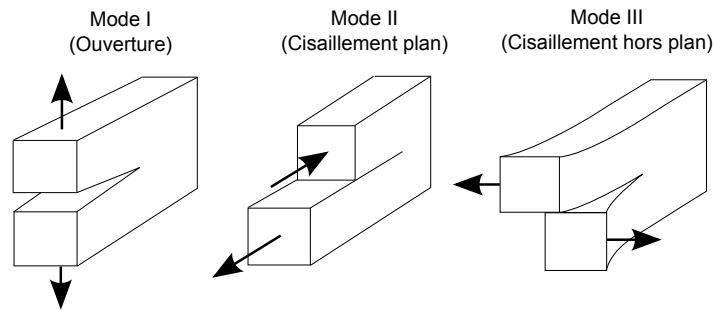


FIGURE 21.4: Modes de chargement d'une fissure

- par les déplacements : K_I est obtenu par $v \sqrt{2\pi/r}$ en fond de fissure (obtenu par extrapolation), v étant le coefficient de Poisson. Cette méthode relativement simple nécessite de raffiner le maillage autour de la fissure et sa précision n'est pas excellente (environ 5%).

Attention : il ne faut pas confondre K_I avec K_t qui est le facteur de concentration de contrainte, sans dimension, et qui caractérise le rapport entre la contrainte normale maximale et la contrainte à l'infini au voisinage d'une entaille.

$$K_t = \frac{\sigma_{22max}}{\sigma_\infty} \quad (21.18)$$

Si l'on reprend le cas de la plaque plane semi-infinie traitée par Griffith, en exprimant le champ de contraintes au voisinage du fond de fissure et en faisant tendre θ vers 0, on obtient des relations entre G et K :

$$G_I = \frac{K_I^2}{E} \quad \text{et en déformations plane :} \quad G_I = (1 - \nu^2) \frac{K_I^2}{E} \quad (21.19)$$

$$G_{II} = \frac{K_{II}^2}{E} \quad \text{et en déformations plane :} \quad G_{II} = (1 - \nu^2) \frac{K_{II}^2}{E} \quad (21.20)$$

$$G_{III} = (1 + \nu) \frac{K_{III}^2}{E} \quad (21.21)$$

21.2.5 Intégrale J

L'intégrale J (intégrale curviligne) représente un moyen de calculer le taux de restitution d'énergie de déformation ou de travail (énergie) par unité de surface de zone rompue au sein d'un matériau.

Histoire

Le concept théorique de l'intégrale J a été développé, de façon indépendante, en 1967 par Cherepanov et en 1968 par Rice. Ces travaux mettent en évidence que le contour délimitant la zone plastique aux abords du front de fissure (appelé J) est indépendant du profil (contour) de la fissure.

Par la suite, des méthodes expérimentales ont été élaborées pour permettre la mesure des propriétés de rupture de critiques à partir d'échantillons à l'échelle du laboratoire pour des matériaux dans lesquels la dimension des prélèvements est insuffisante pour garantir la validité des hypothèses de la mécanique linéaire élastique de la rupture, et d'en déduire une valeur critique de l'énergie de rupture J_{Ic} .

La quantité J_{Ic} définit le point à partir duquel se forme une zone plastique dans le matériau au moment de la propagation et pour un mode de chargement.

L'intégrale J est équivalente au taux de restitution de l'énergie de déformation d'une fissure dans un solide soumis à une charge constante. Cela est vrai, dans des conditions quasi-statiques, tant pour les matériaux linéairement élastiques que pour les échantillons expérimentés à petite échelle en passe de céder en front de fissure.



Cherepanov

Rice

Considérons un solide linéaire élastique homogène bidimensionnel Ω sur lequel agissent des forces T_k , comme illustré à la figure 21.5a. Les efforts de traction s'écrivent $T_i = \sigma_{ij}n_j$ (où n est la normale sortante à $\Gamma = \partial\Omega$). La densité d'énergie interne élastique est :

$$\omega = \int_0^{\varepsilon_{kl}} \sigma_{ij} d\varepsilon_{ij} \quad (21.22)$$

On travaille en l'absence de forces de volumes, d'où $\sigma_{ij,i} = 0$, et sous hypothèse de petites déformations, soit $\varepsilon_{ij} = (u_{i,j} + u_{j,i})/2$. Les intégrales curvilignes :

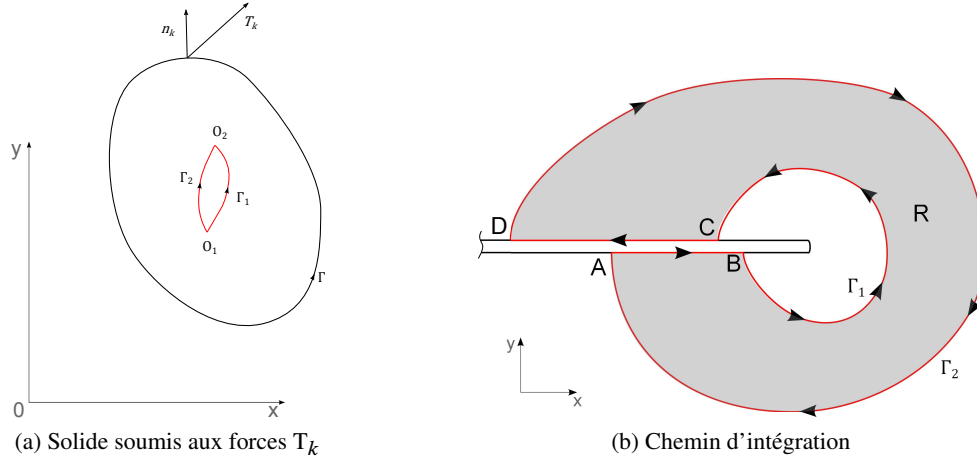


FIGURE 21.5: Calcul de l'intégrale J

$$Q_j = \int_{\Gamma} (\omega n_j - T_k u_{k,j}) d\Gamma, \quad j, k = 1, 2, 3 \quad (21.23)$$

peuvent s'écrire :

$$Q_j = \int_{\Gamma} (\omega n_j - \sigma_{lk} n_l u_{k,j}) d\Gamma = \int_{\Gamma} (\omega \delta_{jl} - \sigma_{lk} u_{k,j}) n_l d\Gamma \quad (21.24)$$

Le théorème de divergence puis quelques manipulations de l'opérande permettent d'arriver au résultat suivant : $Q_j = 0$. De là, on en tire que la densité d'énergie interne élastique :

$$\omega = \int_0^{\varepsilon_{kl}} \sigma_{ij} d\varepsilon_{ij} \quad (21.25)$$

est indépendante du chemin dans l'espace des déformations. Le même type de calcul permet de montrer que la densité d'énergie complémentaire :

$$\omega = \int_0^{\sigma_{kl}} \varepsilon_{ij} d\sigma_{ij} \quad (21.26)$$

est indépendante du chemin dans l'espace des contraintes.

On considère toujours notre solide linéaire élastique homogène 2D sur lequel agissent des forces T_k . On définit l'intégrale J comme :

$$J = Q_1 = \int_{\Gamma} (\omega n_1 - T_k u_{k,1}) d\Gamma \quad (21.27)$$

D'après ce qui précède, $J = 0$, et donc

$$J_1 = \int_{\Gamma_1} \dots = J_2 = \int_{\Gamma_2} \dots \quad (21.28)$$

Considérons maintenant une entaille (fissure) dans une pièce. Alors, on a vu que l'intégrale J est nulle le long du chemin fermé $AB\Gamma_1CD\Gamma_2A$ décrit à la figure 21.5b. Si sur AB et CD : $dy = 0$, $T_k = 0$, alors $J_{AB} = J_{CD} = 0$ et il vient : $J_{\Gamma_1} = J_{\Gamma_2}$

Rice a également démontré que la valeur de l'intégrale J représente le taux de relaxation d'énergie pour la propagation des fissures planes, ou le taux de diminution d'énergie potentielle par rapport à l'accroissement de la fissure, i.e. $J = G$.

L'intégrale J a été développée pour résoudre des difficultés rencontrées dans le calcul des contraintes aux abords d'une fissure dans un matériau linéairement élastique. Rice a montré qu'en mode de chargement constant et sans atteindre l'adaptation plastique, l'intégrale J peut aussi être utilisée pour calculer le taux de relaxation d'énergie dans un matériau plastique.

21.3 Mécanique élastoplastique de la rupture

Nous avons déjà mentionné que l'analyse linéaire élastique en mécanique de la rupture ne s'applique que pour les matériaux fragiles. En effet, dans la plupart des cas, il existe des déformations plastiques au fond de fissure.

Si l'on suppose que cette zone plastique est présente dans un rayon R autour du fond de fissure, et qu'au delà, jusqu'à un rayon D la solution singulière est dominée par K ; alors $R \ll D$ et l'on peut considérer que la solution est dominée par le facteur d'intensité de contrainte K .

Si R n'est pas petit devant D , alors il faut tenir compte de la plastification locale.

21.3.1 Détermination de la zone plastique

L'idée générale est de déterminer le lieu géométrique des points où le champ de contraintes atteint la limite élastique du matériau. Évidemment, en fonction du type de matériau, des hypothèses

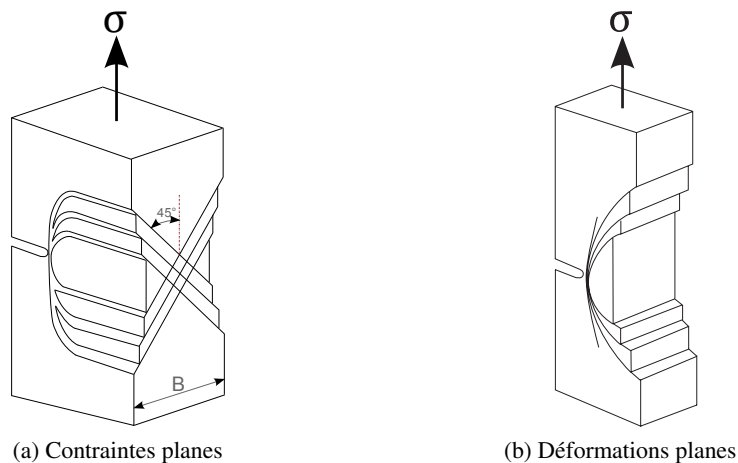


FIGURE 21.6: Zones de plastification d'une éprouvette entaillée en traction

cinématiques... cette zone peut être plus ou moins grande.

Par exemple, pour un matériau isotrope et un critère de von Mises, la zone plastique en contraintes planes est neuf fois plus grande que celle en déformations planes. Les chemins de propagation de la plasticité suivant l'épaisseur sont, dans les cas de contraintes et déformations planes, illustrés sur la figure 21.6.

21.3.2 Modèle d'Irwin

En 1960, Irwin a présenté un modèle de détermination de la zone plastique le long de l'axe de la fissure sous l'hypothèse d'un matériau élastique-plastique parfait, et en contraintes planes.

La zone plastique est donnée par :

$$r_p(0) = \frac{1}{2\pi} \left(\frac{K_I}{\sigma_Y} \right)^2 \quad (21.29)$$

σ_Y étant la limite élastique. Par conséquent, si $0 < r < r_1$, $\sigma_2 = \sigma_Y$ et lorsque $r > r_1$, $\sigma_2 = \frac{K_I}{\sqrt{2\pi r}}$. Comme $\sigma_2 = \sigma_Y$ est constant pour $r < r_1$, il y a violation d'équilibre suivant l'axe y . La réduction de σ_2 est compensée par σ_1 d'où une augmentation de r_1 . On obtient une zone plastique corrigée deux fois plus grande :

$$c = \frac{1}{\pi} \left(\frac{K_I}{\sigma_Y} \right)^2 \quad \text{en contraintes planes} \quad (21.30)$$

$$c = \frac{1}{3\pi} \left(\frac{K_I}{\sigma_Y} \right)^2 \quad \text{en déformations planes} \quad (21.31)$$

et à l'extrémité de la fissure, l'ouverture est δ :

$$\delta = \frac{4}{\pi E} \frac{K_I^2}{\sigma_Y} \quad \text{en contraintes planes} \quad (21.32)$$

$$\delta = \frac{4(1 - \nu^2)}{3\pi E} \frac{K_I^2}{\sigma_Y} \quad \text{en déformations planes} \quad (21.33)$$

21.3.3 Autres modèles

Modèle de Dugdale

Il s'agit d'un modèle valable pour les plaques très minces et constituées d'un matériau plastique parfait avec critère de Tresca, et où la plasticité est concentrée le long de l'axe de la fissure.

Critère de rupture basé sur l'ouverture de fissure

Cette approche, introduite en 1961 par Wells et Cottrell, postule qu'il y a initiation de fissures en présence de plasticité, lorsque $\delta = \delta_c$.

On peut se servir du modèle d'Irwin ou de celui de Dugdale (Burdekin et Stone)...

21.4 Modélisation numérique de la rupture

21.4.1 Par la méthode des éléments finis

Dans la formulation primale des éléments finis (dite formulation en déplacements), le calcul fournit le champ de déplacement u via les déplacements nodaux \mathbf{q} . Le champ de déformations ε est obtenu par dérivation des déplacements ; puis les contraintes sont obtenues via la loi de comportement :

$$\sigma = H\varepsilon \quad (21.34)$$

Si le matériau est élastoplastique, les contraintes seront calculées de façon incrémentale :

$$\Delta\sigma = H(\sigma, \varepsilon)\Delta\varepsilon \quad (21.35)$$

Éléments finis singuliers

Certains éléments ou configurations entraînent des singularités des déformations. Ce phénomène, que l'on cherche généralement soigneusement à éviter, est en fait idéal en mécanique de la rupture. Forcer les éléments à se comporter en $1/\sqrt{r}$ permet d'améliorer considérablement la solution même avec un maillage grossier. Cette singularité peut être obtenue en utilisant des éléments quadratiques et en déplaçant les nœuds des milieux des côtés de $1/4$, comme nous allons le voir.

Considérons l'élément de référence 2D carré quadratique à huit nœuds, souvent appelé Q8. Ses fonctions de forme N_i sont :

$$N_i = \left[(1 + \xi\xi_i)(1 + \eta\eta_i) - (1 - \xi^2)(1 + \eta\eta_i) - (1 - \eta^2)(1 + \xi\xi_i) \right] \frac{\xi_i^2 \eta_i^2}{4} \\ + (1 - \xi^2)(1 + \eta\eta_i)(1 - \xi_i^2) \frac{\eta_i^2}{2} + (1 - \eta^2)(1 + \xi\xi_i)(1 - \eta_i^2) \frac{\xi_i^2}{4} \quad (21.36)$$

Supposons maintenant que les nœuds 5 et 8 soient déplacés de $1/4$ vers le nœud 1 (qui est l'origine du repère), comme illustré à la figure 21.7. Les fonctions de forme des nœuds 1, 2 et 5, obtenues pour $\eta = -1$ sont :

$$N_1 = -\frac{1}{2}\xi(1 - \xi) \quad N_2 = \frac{1}{2}\xi(1 + \xi) \quad N_5 = (1 - \xi^2) \quad (21.37)$$

et le calcul de $x(\xi, \eta)$ donne :

$$x = -\frac{1}{2}\xi(1 - \xi)x_1 + \frac{1}{2}\xi(1 + \xi)x_2 + (1 - \xi^2)x_5 \quad (21.38)$$

soit pour $x_1 = 0$, $x_2 = L$ et $x_5 = L/4$:

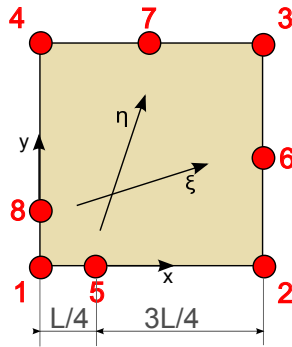


FIGURE 21.7: Position des nœuds de l'élément singulier de type Q8

$$x = \frac{1}{2}\xi(1 + \xi)L + (1 - \xi^2)\frac{L}{4} \quad \text{et par suite : } \xi = -1 + 2\sqrt{\frac{x}{L}} \quad (21.39)$$

Le terme du jacobien est $\frac{\partial x}{\partial \xi} = \frac{L}{2}(1 + \xi) = \sqrt{xL}$ qui s'annule pour $x = 0$ conduisant à une singularité des déformations.

Si l'on calcule la déformation ε_x le long du côté 1-2, on trouve :

$$\varepsilon_x = \frac{\partial u}{\partial x} = -\frac{1}{2} \left(\frac{3}{\sqrt{xL}} - \frac{4}{L} \right) u_1 + \frac{1}{2} \left(-\frac{1}{\sqrt{xL}} + \frac{4}{L} \right) u_2 + \left(\frac{2}{\sqrt{xL}} - \frac{4}{L} \right) u_5 \quad (21.40)$$

qui est bien en $1/\sqrt{x}$.

Techniques numériques de calcul de K, G et J

Il est possible d'effectuer un post-traitement des contraintes en créant un repère local centré sur la pointe de la fissure et en calculant la variable $\sigma_{22}\sqrt{2\pi r}$ en chaque point de Gauß.

On peut également post-traiter les déplacements en les pondérant en fonction de l'état de contrainte théorique censé s'appliquer sur la fissure. Le taux d'énergie libérée de Griffith peut être obtenu par la technique de progression de fissure dans laquelle on effectue deux calculs éléments finis : l'un pour une fissure de dimension a , l'autre pour une fissure de dimension $a + \Delta a$. En post-traitant l'énergie interne de toute la structure, on trouvera $G = -\Delta\Pi/\Delta a$. Si cette méthode est plus précise que les deux précédentes, elle nécessite deux calculs éléments finis.

Un calcul par dérivation de la rigidité peut aussi être effectué. Le but est encore une fois d'essayer de calculer $G = -\Delta\Pi/\Delta a$. L'énergie potentielle étant donnée, dans le cas de l'élasticité linéaire, par :

$$\Pi = \frac{1}{2} {}^T \mathbf{q} \mathbf{K} \mathbf{q} - {}^T \mathbf{q} \mathbf{F} \quad (21.41)$$

Le calcul de G conduit à :

$$G = -\frac{1}{2} {}^T \mathbf{q} \frac{\partial \mathbf{K}}{\partial a} \mathbf{q} = -\frac{1}{2} {}^T \mathbf{q} \left(\sum_{i=1}^N \frac{\partial \mathbf{k}_i}{\partial a} \right) \mathbf{q} \quad (21.42)$$

Cette méthode ne fait certes intervenir qu'un seul calcul EF, mais elle nécessite un développement de code de dérivation des matrices de rigidité.

Si l'on a bien suivi, on a vu que l'intégrale J est un outil bien adapté : ne dépendant pas du contour, cette intégrale n'est que peu sensible à la qualité du maillage. La plupart des codes actuels permettent de calculer cette intégrale J sur un chemin défini par l'utilisateur.

21.4.2 Par les méthodes sans maillage

la méthode des éléments finis souffre des limitations suivantes :

- La création de maillage pour des problèmes industriels est une tâche qui reste difficile et coûteuse ;
- En général les champs de contraintes issus des éléments finis sont discontinus, donc peu précis (mais nous avons vu comment y remédier...);
- Distorsion des éléments en grandes déformations (ou alors régénération d'éléments, ce qui est une technique compliquée) ;
- Difficulté de prédiction des chemins de propagation de fissures (ne passant pas par les nœuds), doublée de la difficulté de remailler automatiquement pour le suivi de propagation de fissures
- Difficulté de représentation de l'effritement de matériau (impacts, explosifs,...).

Une idée « simple » est d'éliminer les éléments...

Une brève description de la méthode est donnée au paragraphe 23.2.

21.4.3 Par les éléments étendus

La méthode des éléments finis étendus (X-FEM) est rapidement présentée au paragraphe 23.4.

Les éléments finis classiques ayant du mal avec les fortes discontinuités, on procède à un enrichissement des éléments, dans les zones de discontinuités. L'enrichissement peut être interne (ou intrinsèque) ou externe (ou extrinsèque). Le principe consiste à augmenter la qualité de l'approximation des fonctions en ajoutant à l'approximation déjà présente des informations sur la solution exacte (et donc sur l'approximation du problème spécifique que l'on souhaite résoudre).

21.5 Fatigue et durée de vie

Dans ce paragraphe, nous nous placerons dans le cas des matériaux composites, même si les démarches présentées sont pour beaucoup applicables à tous types de matériaux.

À l'heure actuelle, la question des critères de propagation en dynamique de la rupture restent un sujet ouvert. Bien que de nombreux essais de propagation dynamique de fissure aient été effectués ces dernières décennies, on se heurte à l'absence de méthodes numériques suffisamment fiables pour identifier les paramètres d'éventuels critères et comparer leur pertinence. Par ailleurs, de tels essais sont difficiles à mettre en œuvre.

Nous nous intéressons à des matériaux soumis à des sollicitations de faible intensité, qui individuellement ne présenteraient pas de danger, mais qui appliquées de façon cyclique conduisent à l'amorçage, puis à la propagation de fissures

Tant que les fissures restent suffisamment petites, il n'y a pas de risque de rupture. Le risque survient lorsque les microfissures (qui ne sont pas étudiées) grandissent (ou se connectent) pour former des macrofissures auxquelles on peut appliquer ce qui a été dit avant.

21.5.1 Courbe et limite de fatigue

La courbe de Wöhler ou courbe S-N (pour Stress vs Number of cycles) de beaucoup de CFRP et GFRP (CFRP = Carbon Fiber Reinforced Polymer = polymère à renfort fibre de carbone ; GFRP = Glass Fiber Reinforced Polymer = polymère à renfort fibre de verre) peut être décrite (entre 10^3 et 10^6 cycles) par une équation de la forme :

$$\frac{\sigma_a}{\sigma_u} = 1 - b \log N \quad (21.43)$$

où σ_a et σ_u sont la contrainte appliquée et la résistance ultime, N , le nombre de cycle et b , une constante.

La plupart des données disponibles en fatigue montrent une dispersion élevée pour les courbes S-N. L'analyse statistique est alors inévitable. D'après Talreja, il existe une limite de déformation en fatigue des composites, décrite comme la déformation minimale requise pour initier un mécanisme d'endommagement de faible énergie. Il suggère que pour les UD à base d'époxyde, cette limite est autour de 0.6%.

21.5.2 Cumul des dommages : principes de Miner

Les essais de fatigue en laboratoire consistent généralement à répéter une même sollicitation un grand nombre de fois. Dans ce cas, il est facile de définir un dommage : c'est le nombre de répétitions de l'événement endommageant depuis le début de l'essai.

Dans le cas général, il y a plusieurs événements endommageants, qui diffèrent les uns des autres par la grandeur des contraintes subies et par d'autres paramètres. En 1945 Miner a popularisé deux principes, initialement proposés par Palmgren en 1924, qui permettent de cumuler les dommages :

- le dommage causé par une occurrence d'un événement est mesuré par l'inverse $1/N$ du nombre N de fois qu'il faut répéter cet événement pour mener la pièce de l'état neuf jusqu'à la défaillance ;
- le dommage causé par une succession d'événements est la somme des dommages causés par chacun d'eux.

Remarquons que ces deux principes sont cohérents. Si l'on suppose qu'un même événement cause toujours le même dommage, et si l'on impose comme unité de dommage le dommage qui conduit le système jusqu'à la défaillance, alors le premier principe est une conséquence du second.

Si $1/N_A$ est le dommage de l'événement A et si n_A est le nombre d'occurrences de cet événement au cours d'un essai ou d'une utilisation du système alors le dommage total D causé par tous les

événements A, relativement à une défaillance, est défini par l'équation de Miner :

$$D = \frac{n_A}{N_A} \quad (21.44)$$

Avec cette définition, le dommage cumulé est égal à 1 au moment où la pièce se rompt. Il faut nuancer cette affirmation : des pièces apparemment identiques soumises aux mêmes sollicitations se rompent en général au bout de nombres de cycles différents. La dispersion des durées de vie peut être très importante.

On trouve par expérience que les durées de vie peuvent varier d'un facteur 5 ou 10 pour des pièces issues d'un même lot de production. Il faut donc préciser davantage la formulation de la règle de Miner : s'il faut en moyenne N_A répétitions de l'événement A pour qu'une défaillance survienne, alors le dommage causé par une occurrence de A est mesuré par $1/N_A$.

Les principes de Miner permettent de définir le dommage causé par un événement et affirment que pour cumuler les dommages, il suffit de les additionner. Ils imposent en outre le choix de l'unité de dommage. Remarquons qu'il est toujours possible de changer d'unité. Par exemple, lorsqu'il n'y a un seul événement endommageant, il est naturel de choisir comme unité de dommage le dommage causé par une occurrence de cet événement.

Les principes de Miner permettent d'établir l'égalité des endommagements produits par des événements de natures différentes. En particulier, ils permettent de préciser quand un petit nombre d'événements de grande amplitude produit un endommagement égal à un grand nombre d'événements de faible amplitude. Cela revêt une grande importance lorsqu'il faut définir un essai aussi court que possible.

21.5.3 Propagation : loi de Paris

Pour une singularité caractérisée par sa dimension a et sa forme, on étudie les courbes

$$\frac{da}{dN} = \Delta K \quad (21.45)$$

où N est le nombre de cycles et ΔK la variation du facteur d'intensité de contrainte sur un cycle.

Ces courbes, dont la partie centrale est linéaire, présentent deux asymptotes pour da/dN et ΔK faibles et pour da/dN et ΔK grands. La valeur minimale vers laquelle tend la courbe est ΔK_S , la valeur maximale ΔK_{IC} . K_{IC} est la valeur critique qui correspond à une rupture instantanée par dépassement de la valeur critique de K sous chargement monotone. K_S est la valeur en dessous de laquelle il n'y a pas de propagation de fissure, c'est un facteur d'intensité de contrainte seuil.

Concernant la partie linéaire de ces courbes dans un diagramme log-log, cela permet de les modéliser par la loi de Paris (la plus simple des lois de propagation) qui définit la vitesse de propagation par cycle comme une fonction puissance de l'amplitude du facteur d'intensité de contrainte :

$$\frac{da}{dN} = C \Delta K^m \quad (21.46)$$

où C et m sont des coefficients dépendant du matériau.

La dimension critique de la singularité a_c est liée à la caractéristique du matériau K_{IC} , la ténacité, elle entraîne la rupture fragile de la structure :

$$K_{IC} = F\sigma\sqrt{\pi a_c} \quad (21.47)$$

où σ est une contrainte effective dans une direction normale à la fissure et F un facteur de forme.

21.5.4 Prédiction de la durée de vie

Une fois un indicateur d'endommagement choisi (généralement la rigidité ou la résistance résiduelle, la première ayant l'avantage non négligeable de pouvoir être mesurée de manière non destructive), on utilise :

Théories empiriques : Les courbes S-N sont utilisées pour caractériser le comportement en fatigue du composite. Un nombre important d'essais est nécessaire pour chaque configuration spécifique ;

Théories de dégradation de résistance résiduelle : Elles sont basées sur l'hypothèse qu'un changement de la résistance résiduelle σ_r en fonction du nombre de cycles n est lié à la contrainte maximale appliquée σ_a . On a généralement une relation de la forme :

$$\frac{d\sigma_r}{dn} = -\frac{1}{\gamma} f \sigma_a^\gamma \sigma_r^{1-\gamma} \quad (21.48)$$

où γ et f sont des paramètres indépendants des contraintes mais qui peuvent dépendre de la température, de l'humidité, de la fréquence. Souvent, on suppose une distribution de Weibull¹ de la résistance des composites. La rupture par fatigue survient lorsque la résistance résiduelle du composite est atteinte par la contrainte appliquée.

Théories de perte de rigidité : Ce sont des généralisations du cas précédent.

Les plis les plus désorientés par rapport au chargement (éléments sous-critiques) commencent à s'endommager (ce qui est caractérisé par une perte de rigidité), ce qui cause une redistribution des contraintes au niveau local. La résistance des éléments critiques (les plis orientés dans le sens de la charge) est gouvernée par les équations de dégradation de résistance. Ces deux phénomènes (perte de rigidité des éléments sous-critiques et perte de résistance des éléments critiques) contribuent à la définition de la résistance résiduelle et donc à la durée de vie.

Théories d'endommagement cumulatif : Elles sont basées sur une observation expérimentale soigneuse et sur une simulation de l'accumulation de l'endommagement du sous-lamina. Il manque toutefois un critère de rupture des laminés sous chargement de fatigue tension-tension pour ces théories.

Théories d'endommagement continu : L'endommagement est pris en compte par un paramètre D sous la forme :

$$\hat{\sigma} = \frac{\sigma}{1-D} \quad (21.49)$$

1. En théorie des probabilités, la loi de Weibull, est une loi de probabilité continue. Avec deux paramètres sa densité de probabilité est :

$$f(x; k, \lambda) = (k/\lambda)(x/\lambda)^{(k-1)} e^{-(x/\lambda)^k}$$

où $k > 0$ est le paramètre de forme et $\lambda > 0$ le paramètre d'échelle de la distribution. Sa fonction de répartition est définie par $F(x; k, \lambda) = 1 - e^{-(x/\lambda)^k}$, où, ici encore, $x > 0$. Une version généralisée à trois paramètres existe.

La distribution de Weibull est souvent utilisée dans le domaine de l'analyse de la durée de vie, grâce à sa flexibilité, car elle permet de représenter au moins approximativement une infinité de lois de probabilité.

Par exemple, si le taux de panne diminue au cours du temps alors, $k < 1$. Si le taux de panne est constant dans le temps alors, $k = 1$. Si le taux de panne augmente avec le temps alors, $k > 1$. La compréhension du taux de panne peut fournir une indication au sujet de la cause des pannes :

- un taux de panne décroissant relève d'une « mortalité infantile ». Ainsi, les éléments défectueux tombent en panne rapidement, et le taux de panne diminue au cours du temps, quand les éléments fragiles sortent de la population ;
- un taux de panne constant suggère que les pannes sont liées à une cause stationnaire ;
- un taux de panne croissant suggère une « usure ou un problème de fiabilité » : les éléments ont de plus en plus de chances de tomber en panne quand le temps passe.

On dit que la courbe de taux de panne est en forme de baignoire. Les fabricants et distributeurs ont tout intérêt à bien maîtriser ces informations par type de produits afin d'adapter : 1) les durées de garantie (gratuites ou payantes), et 2) le planning d'entretien.

où σ et $\hat{\sigma}$ sont la contrainte imposée et la contrainte effective après endommagement. L'avantage de cette approche est d'éviter la prise en compte de l'endommagement microstructural parfois difficile à modéliser et mesurer.

Notons également que pour obtenir les informations nécessaires, de nombreux essais doivent être menés, surtout parce que de nombreux facteurs conditionnent la fatigue des FRP (FRP = Fibre-reinforced plastic = plastique à renfort de fibres) :

- fréquence : un échauffement du matériau avec la fréquence de sollicitation est constaté. Il faudrait modifier ses lois de comportement.
- amplitude : une sollicitation de grande amplitude suivie d'une sollicitation de faible amplitude conduit à une durée de vie inférieure au cas où l'ordre des sollicitations est inversé.
- $R = -1$ (rapport entre charge positive et négative dans le cyclage) : ce cas est très défavorable pour les FRP car leur tenue en compression est moins bonne qu'en traction (flambage des fibres et splitting...);
- la forme du signal a une influence qui serait à considérer.
- séquence d'empilement : les composites sont censés être conçus pour soutenir la charge dans le sens des fibres. Dans le cas de sollicitations complexes, les plis les moins bien orientés par rapport à la charge cèdent en premier...
- humidité;
- vieillissement naturel...
- ...

C'est un problème de structure car il y a couplage entre les modes d'endommagement à différentes échelles. Dans les cas simples, on peut utiliser un critère cumulatif de Miner ou des lois d'équivalence. Attention toutefois, l'« équivalence » entre temps, fréquence, température... ne « va pas de soi ». Il s'agit d'hypothèses, parfois restrictives, et toujours limitées à une plage d'utilisation. On les manipulera donc avec intérêt (par exemple pour la conception d'essais accélérés) et avec prudence.

21.5.5 Sur la fatigue des composites

Les remarques suivantes peuvent être faites, quant à la poursuite de travaux de recherche sur le sujet :

- Bien que la conception orientée par la rigidité pour les GFRP soit peu concernée par la contrainte de rupture, le fluage est un aspect fondamental de leur conception.
- Les effets de synergie entre le fluage et les autres types de chargement est un domaine à explorer.
- Peu de données de fatigue sont disponibles pour le domaine $10^7 - 10^8$ cycles (Les japonais ont des essais en cours).
- L'effet d'échelle sur les performances n'est pas clair. En particulier, il n'est toujours pas sûr qu'un tel effet existe.
- La représentativité des éprouvettes ou des tests accélérés restent des problèmes épineux.
- Les composites hybrides doivent être étudiés.
- La dégradation des GFRP à haute température n'est pas bien comprise. Toutefois, Tsotsis et Lee notent, suite à l'étude de deux résines particulières, que le comportement à long terme des composites soumis à des températures élevées est contrôlé par les dégradations d'oxydation et thermique.
- Les composites épais ne sont que très peu étudiés, tout comme l'influence de l'épaisseur en général. Pourtant, des mécanismes de ruptures suivant l'épaisseur peuvent sans doute apparaître de manières différentes de celles des laminés plus fins : effet de taille ou interaction de mécanismes.

Chapitre 22

Stochastique : prise en compte de la variabilité

Résumé — Certains diront que c'est enfin dans ce chapitre que nous traitons de la réalité physique... c'est un point de vue. En effet, rien n'est jamais connu de manière parfaite.

L'aléa traduit aussi bien l'impossibilité d'une description déterministe exhaustive que l'irrégularité de tout phénomène observé. Les modèles déterministes ne sont finalement que des approximations des problèmes physiques correspondants, tout comme les modèles linéaires ne sont que des approximations de comportements réels non-linéaires par nature.

Nous nous restreindrons dans le nombre de formulations afin de ne présenter que ce qui nous semble aujourd'hui le plus pertinent.

22.1 Introduction

Dans la manière traditionnelle, encore appelée approche déterministe, la conception des structures repose sur des paramètres tels que les dimensions, la résistance des matériaux et le chargement, tous caractérisés par une valeur constante, i.e. leur moyenne. Sur la base de ces constantes on utilise un modèle mathématique du comportement pour déterminer si la structure est sûre ou non. Afin d'améliorer encore la sécurité, les variables structurelles sont alors remplacées par leur pire cas. Cette philosophie de conception se révèle trop coûteuse d'un point de vue économique car on se place dans le cas où tous les paramètres sont à leur pire valeur en même temps.

Il est bien connu que, par exemple, la résistance varie d'un élément structurel à l'autre, de sorte que cette résistance ne peut être décrite par une unique valeur. De plus, il est parfois nécessaire de prendre en compte des variations temporelles. Ces mêmes variations existent également pour les dimensions et le chargement. Cela est particulièrement vrai pour les chargements naturels comme la houle, le vent et les séismes, qu'il est difficile de prendre en compte de manière déterministe. Il faut en outre garder en tête qu'une certaine incertitude existe également dans le choix des modèles mathématiques utilisés pour l'analyse de la structure.

Le but d'utiliser une approche probabiliste plutôt qu'une simple approche déterministe est d'essayer de prendre en compte les incertitudes mentionnées ci-dessus afin de réaliser une analyse plus réaliste de la sûreté de la structure.

Dans ce chapitre, nous considérerons le problème de formulation classique suivant :

$$au = f \quad (22.1)$$

Jusqu'à présent, nous nous sommes contenté du cas où a est un opérateur déterministe, f l'excitation déterministe et u la réponse déterministe. Nous allons dans ce chapitre nous intéresser au cas où f est une excitation aléatoire et a un opérateur éventuellement aléatoire. Il s'en suit que la réponse du système u est elle-aussi aléatoire.

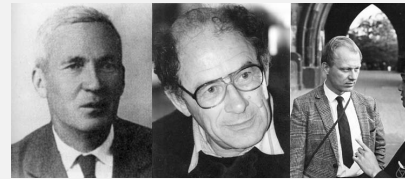
La manipulation d'équations stochastiques introduit deux difficultés :

- premièrement, les propriétés aléatoires du système doivent être modélisées « correctement » comme variables ou processus aléatoires, avec une distribution de probabilité réaliste ;
 - deuxièmement, il faut être capable de résoudre le système différentiel ainsi obtenu, et la réponse obtenue doit pouvoir être décrite par ses moments statistiques.
- Enfin, la relation entre éléments finis et probabilité recouvre deux aspects :
- d'une part le calcul des moments statistiques de la réponse autour de sa moyenne (essentiellement l'écart-type) ;
 - d'autre part les méthodes de fiabilité, par lesquelles on cherche à calculer une probabilité de défaillance associée à un critère dont les arguments dépendent du résultat d'un calcul par éléments finis.

Histoire

Les séries de fonctions, apparues à la fin du XVII^e siècle, et particulièrement les séries de Taylor (voir historique du paragraphe A.1.1), sont aujourd'hui un outil indispensable, permettant notamment d'approcher une fonction de manière facilement exploitable. C'est pourquoi la manière la plus naturelle d'appréhender le traitement de l'équation (22.1) a été de procéder à un développement en série.

Dès le début du XVIII^e siècle, la théorie des perturbations a été utilisée par les astronomes pour les besoins de la mécanique céleste : en effet, les équations différentielles décrivant un système de n corps en interaction gravitationnelle n'a pas de solution exacte générale pour $n \geq 3$. Cet aspect de la théorie des perturbations a été synthétisé à la fin du XIX^e siècle dans les ouvrages classiques de Laplace, Tisserand et Poincaré, avant de connaître de nouveaux développements dans la seconde moitié du XX^e siècle avec l'avènement en 1954 de la « théorie KAM » (théorème de mécanique classique i.e. non relativiste et non quantique, mettant en défaut l'hypothèse ergodique de Boltzmann), du nom de ses trois concepteurs : Kolmogorov, Arnold et Moser. La méthode a par ailleurs été abondamment utilisée au XX^e siècle pour les besoins de la physique quantique, d'abord en mécanique quantique non relativiste, puis en théorie quantique des champs.



Kolmogorov Arnold Moser

C'est encore cette même méthode de perturbation qui est la plus ancienne et la plus utilisée en ingénierie pour analyser les systèmes aléatoires (pas seulement aléatoires, voir la présentation de la méthode de réanalyse au paragraphe 13.6 qui est une méthode de perturbation). Bien que les méthodes mathématiques sous-jacentes soient très simples, cela ne signifie nullement que leur validité aille de soi : on fait un développement de Taylor de chaque quantité aléatoire autour de sa moyenne, les termes causant l'instabilité de la solution approchée étant d'ordres supérieurs. Toutefois, dans la pratique, il n'est pas possible d'aller au delà de l'ordre un ou deux. Cela réduit la portée des applications de la méthode aux cas des petits aléas, i.e. de petites fluctuations autour d'une valeur moyenne. Sans efforts inconsidérés, ces méthodes ne fournissent pas de statistiques d'ordres élevés.

Il a donc bien évidemment été essayé d'améliorer cette méthode afin d'obtenir des statistiques d'ordres plus élevés. Une méthode, peu connue et peu utilisée, pour atteindre ce but est la hierarchy closure approximation. Il s'agit d'exprimer les moments d'ordres élevés en fonctions de moments d'ordres moins élevés. Si l'on considère que l'opérateur a de l'équation (22.1) peut être décomposé en une partie déterministe \bar{a} et une partie aléatoire \tilde{a} , cette équation se met sous la forme :

$$(\bar{a} + \tilde{a})u = f \quad (22.2)$$

qui se résout en :

$$u = \bar{a}^{-1} f - \bar{a}^{-1} \tilde{a} u \quad (22.3)$$

En appliquant à cette dernière équation l'opérateur \tilde{a} , puis en réinjectant le résultat dans l'équation précédente, il vient :

$$u = \bar{a}^{-1} f - \bar{a}^{-1} \tilde{a} \bar{a}^{-1} f + \bar{a}^{-1} \tilde{a} \bar{a}^{-1} \tilde{a} u \quad (22.4)$$

et on peut poursuivre ainsi, si on le souhaite. Cela ne peut néanmoins être fait que si :

$$E[(\bar{a}^{-1}\tilde{a})^n u] = E[(\bar{a}^{-1}\tilde{a})^n] E[u] \quad (22.5)$$

où $E[.]$ est la moyenne (espérance mathématique).

Le découplage effectué par l'équation (22.5) n'a pas de base rigoureuse et est souvent justifié intuitivement par un argument d'indépendance locale. **On aboutit de toutes façons à une formulation très complexes pour les moments d'ordres élevés**, de sorte que cette méthode n'est utilisée en pratique elle aussi que pour le cas des petites fluctuations.

22.2 Représentation des processus stochastiques

Nous allons reprendre et compléter ce qui a été présenté au paragraphe 1.2. Nous avons donc déjà exposé qu'une probabilité est une mesure particulière. Reprenons cela avec un vocabulaire probabiliste.

22.2.1 Variable aléatoire

On appelle épreuve l'observation d'un phénomène aléatoire. Toutes les réalisations possibles d'une épreuve forment l'ensemble de tous les résultats d'une expérience aléatoire, qui sera noté Θ . Un événement E est un sous-ensemble de Θ contenant les réalisations $\theta \in \Theta$. La mesure de l'occurrence de E est une mesure de probabilité notée P . L'ensemble de tous les événements possibles ayant une probabilité ainsi définie est appelé une σ -algèbre associée à Θ , et est notée \mathcal{F} . L'espace de probabilité construit grâce à ces notions est noté (Θ, \mathcal{F}, P) . En d'autres termes, pour reprendre la partie I, l'espace de probabilité (Θ, \mathcal{F}, P) est construit sur la tribu \mathcal{F} et avec la mesure P .

Une variable aléatoire réelle X est une fonction $X : (\Theta, \mathcal{F}, P) \rightarrow \mathbb{R}$. Pour les variables aléatoires continues, la densité de probabilité est notée $f_X(x)$ et la fonction de répartition est notée $F_X(x)$. Un vecteur aléatoire est un vecteur dont les composantes sont des variables aléatoires.

Définition 81 — Moments d'une variables aléatoire. Les moments d'ordre n d'une variable aléatoire X , s'ils existent, sont définis par :

$$E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x) dx \quad (22.6)$$

et les moments centrés réduits, s'ils existent, sont :

$$\text{moyenne : } \mu = (X) = E[X] = \int_{-\infty}^{\infty} x f_X(x) dx \quad (22.7)$$

$$\text{variance : } \sigma^2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx \quad (22.8)$$

$$\text{coefficient d'asymétrie (skewness) : } \delta = \frac{E[(X - \mu)^3]}{\sigma^3} \quad (22.9)$$

$$\text{coefficient d'aplatissement (kurtosis) : } \kappa = \frac{E[(X - \mu)^4]}{\sigma^4} \quad (22.10)$$

La variance est égale au carré de l'écart-type.

22.2.2 Espace de probabilité

Soient maintenant deux variables aléatoires X et Y . La covariance de ces deux variables est :

$$\text{Cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f_{X,Y}(x, y) dx dy \quad (22.11)$$

où $f_{X,Y}$ est leur densité de probabilité conjointe.

Définition 82 L'espace vectoriel des variables aléatoires réelles d'écart-type fini (moment d'ordre 2), noté $L^2(\Theta, \mathcal{F}, P)$, est un espace de Hilbert.

Le produit scalaire sur cet espace de Hilbert est défini comme : $(X, Y) = \int xy dP$ où dP est la mesure de probabilité conjointe, ce qui équivaut à $(X, Y) = \int xy f_{X,Y} dx dy$. Avec ce qui précède, le produit scalaire est donc finalement défini par :

$$(X, Y) = E[XY] \quad (22.12)$$

et la norme associée est :

$$\|X\| = \sqrt{E[X^2]} \quad (22.13)$$

Théorème 65 Les polynômes d'Hermite sous leur forme probabiliste, H_n , présentés au paragraphe A.1.5, forment une base de l'espace $L^2(\Theta, \mathcal{F}, P)$.

On rappelle également qu'une variable aléatoire gaussienne centrée réduite (i.e. de moyenne nulle, écart-type égal à 1) a une densité de probabilité φ et une fonction de répartition Φ définies par :

$$\varphi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \quad \text{et} \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad (22.14)$$

Dans la suite du chapitre, nous noterons ξ une telle variable aléatoire gaussienne centrée réduite.

Nous présentons quelques propriétés des polynômes d'Hermite, à commencer par la relation d'orthogonalité :

$$\int_{-\infty}^{+\infty} H_n(x) H_m(x) \varphi(x) dx = n! \delta_{nm} \quad (22.15)$$

où δ_{nm} est le symbole de Kronecker.

$$\frac{dH_n(x)}{dx} = nH_{n-1}(x) \quad (22.16)$$

$$H_i(x) H_j(x) = \sum_{k=|i-j|}^{i+j} C_{ijk} H_k(x) \quad (22.17)$$

avec :

$$\begin{cases} C_{ijk} = 0 & \text{si } \frac{i+j+k}{2} \notin \mathbb{N} \\ C_{ijk} = \frac{i!j!}{\left(\frac{i+j-k}{2}\right)! \left(\frac{i+k-j}{2}\right)! \left(\frac{j+k-i}{2}\right)!} & \text{sinon} \end{cases} \quad (22.18)$$

22.2.3 Processus ou champ aléatoire

Nous venons de présenter les variables aléatoires. Or, ce qui nous intéresse, est de prendre en compte la variation d'une propriété (par exemple le module d'Young du matériau) continûment sur notre domaine Ω .

Définition 83 — Champ aléatoire ou processus stochastique. Un champ aléatoire scalaire $w(x, \theta)$ peut être défini comme un ensemble de variables aléatoires indexées par un paramètre continu $x \in \Omega$. le paramètre x n'est pas forcément continu, il peut être discret. Nous nous intéressons ici au cas d'un processus continu. Ainsi un processus stochastique $w(x, \theta)$ définit une fonction de deux variables x et θ et représente :

- x et θ variables : une famille de fonctions ;
- x variable et θ fixé : une fonction de x , i.e. une réalisation du champ aléatoire (i.e. une fonction de \mathbb{R}^d dans \mathbb{R} , où notre domaine Ω est un ouvert de \mathbb{R}^d) ;
- x fixé et θ variable : une variable aléatoire ;
- x et θ fixés : un nombre.

Un champ aléatoire est dit :

- vectoriel si la quantité $w(x, \theta)$ attachée au point x est un vecteur aléatoire.
- unidimensionnel si $d = 1$ et multidimensionnel sinon (cas qui nous intéresse).
- gaussien si tout vecteur $\{w(x_1), \dots, w(x_n)\}$ est un vecteur gaussien. Il est alors complètement défini par sa moyenne, sa variance et sa fonction d'autocovariance définie par :

$$C(x, x') = \text{Cov}[w(x), w(x')] \quad (22.19)$$

- stationnaire (au sens faible) si sa moyenne et sa variance sont constantes et si sa fonction d'autocorrélation ρ ne dépend que de la différence $(x - x')$. La fonction d'autocorrélation est définie par :

$$\rho(x, x') = \frac{C(x, x')}{\sigma(x)\sigma(x')} \quad (22.20)$$

22.2.4 Discrétisation de champs aléatoires

Maintenant que nous avons défini ce qu'est un processus stochastique $w(x, \theta)$, il nous faut le discrétiser. On se propose donc d'approcher $w(x, \theta)$ par $w_h(x, \chi)$, où χ est un vecteur aléatoire constitué de n variables aléatoires ξ_i , i.e. sous la forme :

$$w(x, \theta) \approx w_h(x, \chi) \quad (22.21)$$

Plusieurs méthodes sont possibles :

discrétisation par valeurs moyennes : les ξ_i sont des intégrales pondérées de $w(x, \theta)$ sur un domaine Ω_i , et dans le cas des éléments finis, sur chaque élément. Cette méthode a pour conséquence de lisser le processus stochastique ;

discrétisation par valeurs ponctuelles : les ξ_i sont sélectionnées parmi les valeurs de $w(x, \theta)$ en certains points x . Cette méthode de collocation a pour conséquence de générer des irrégularités additionnelles ;

développement en séries : le champ est représenté par une série de variables aléatoires et de fonctions spatiales déterministes. C'est uniquement ce type de méthode que nous allons présenter.

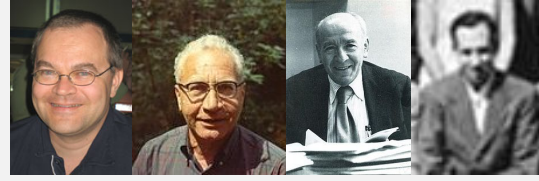
22.2.5 Développement en série de Karhunen-Loève

Histoire

Nous avons déjà mentionné dans ce chapitre que la méthode de perturbation, basée sur un développement en série de Taylor, avait des limitations.

Toutefois, le recours à un développement en série reste une idée intéressante... reste à trouver comment procéder.

Pour améliorer le développement en série d'un champ aléatoire, on peut penser à utiliser un développement à la Fourier : scinder la dépendance des variables x et θ , et utiliser une base de fonctions orthogonales. Ainsi cherchons à écrire le processus stochastique $w(x, \theta)$ sous la forme :



Karhunen

Loève

Kac

Siegert

$$w(x, \theta) = \sum_{i=0}^{\infty} \alpha_i \xi_i(\theta) f_i(x) \quad (22.22)$$

où les α_i sont des constantes à déterminer, les ξ_i des variables aléatoires et les f_i des fonctions déterministes orthogonales entre elles. C'est exactement ce que propose le développement de Karhunen-Loève, obtenu indépendamment par Karhunen en 1947, Loève en 1948, mais également en 1947 par Kac et son ami Siegert du Radiation Laboratory.

La méthode de Karhunen-Loève se propose elle-aussi de décomposer le champ aléatoire $w(x, \theta)$ en une partie déterministe (sa moyenne $\mu(x)$) et une partie aléatoire. Cette fois, la partie aléatoire sera décomposée sur la base des valeurs propres λ_i et des fonctions propres $\varphi_i(x)$ de la fonction d'autocovariance

$$C_{ww}(x, x') = \sigma(x)\sigma(x')\rho(x, x') \quad (22.23)$$

Il faut donc résoudre le problème aux valeurs propres suivant :

$$\int_{\Omega} C_{ww}(x, x') \varphi(x') d\Omega = \lambda_i \varphi_i(x), \quad \forall i \geq 1 \quad (22.24)$$

qui est une équation intégrale de Fredholm de deuxième espèce. Le noyau $C_{ww}(\cdot, \cdot)$ étant une fonction d'autocovariance, il est borné, symétrique et défini positif (Loève, 1977). l'ensemble des $\{\varphi_i\}$ forme une base complète orthogonale de l'espace des fonctions $L^2(\Omega)$. Les valeurs propres sont réelles, positives et en nombre fini.

Chaque réalisation de w peut donc être développée sur cette base sous la forme :

$$w(x, \theta) = \mu(x) + \sum_{i=1}^{\infty} \sqrt{\lambda_i} \xi_i(\theta) \varphi_i(x) \quad (22.25)$$

où les ξ_i sont les coordonnées de la réalisation du champ aléatoire par rapport à l'ensemble des fonctions déterministes φ_i .

On montre aisément que $E[\xi_m \xi_n] = \delta_{mn}$ et $E[\xi_n] = 0$. La famille $\{\xi_i\}_{i \geq 1}$ forme un ensemble orthonormal de variables aléatoires.

Comme il ne peut y avoir de valeurs propres multiples (sauf 0), il est possible de les classer en une suite décroissante. En tronquant la somme de l'équation (22.25) à l'ordre M (i.e. en ne retenant que les M plus grands valeurs propres), on obtient l'approximation de Karhunen-Loève du champ aléatoire :

$$w_h(x, \theta) = \mu(x) + \sum_{i=1}^M \sqrt{\lambda_i} \xi_i(\theta) \varphi_i(x) \quad (22.26)$$

Comme les fonction propres sont orthonormales, on obtient une expression de chacune des variables aléatoires ξ_i apparaissant dans la série de Karhunen-Loève :

$$\xi_i(\theta) = \frac{1}{\sqrt{\lambda_i}} \int_{\Omega} (w(x, \theta) - \mu(x)) \varphi_i(x) d\Omega \quad (22.27)$$

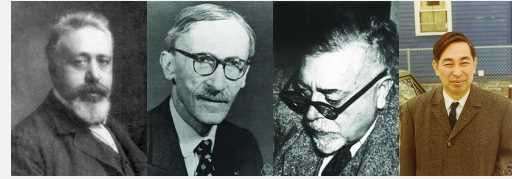
Cette dernière équation nous dit également que lorsque w est un champ gaussien, alors chaque variable aléatoire ξ_i est également gaussienne. Les $\{\xi_i\}_{i \geq 1}$ forment un ensemble de variables aléatoires gaussiennes centrées réduites indépendantes.

22.2.6 Chaos polynomial

Histoire

La théorie des fonctionnelles non-linéaires a été développée par Volterra dès 1913. Il généralisa le développement des fonctions en série de Taylor au cas des fonctionnelles.

Wiener s'intéressa au problème par ses contact avec Paul Lévy, élève de Volterra. C'est lui qui appliqua les idées de Volterra pour la première fois à l'analyse stochastique. Il appliqua sa théorie du mouvement brownien à l'intégration des fonctionnelles de Volterra et développa ce que l'on appelle désormais le chaos homogène.



Volterra

Lévy

Wiener

Itô

Poursuivant les travaux de Wiener, Cameron et Martin en 1947 proposèrent un développement de type Fourier pour les fonctionnelles non-linéaires : le développement de Fourier-Hermite. C'est à nouveau Wiener qui, à partir de ses travaux de 1923 sur les espaces différentiels et de 1938 sur les chaos homogène, appliqua le développement de Fourier-Hermite en 1958 aux problèmes impliquant des phénomènes aléatoires. Il en résultera le développement de Wiener-Hermite, qui a été appliqué depuis à une grande variété de problèmes.

Le chaos homogène de Wiener a ensuite été raffiné par Itô en 1951 pour obtenir ce que l'on appelle maintenant l'intégrale multiple de Wiener.

L'utilisation du développement de Karhunen-Loève nécessite de connaître la fonction de covariance du processus à développer. Or, si l'on souhaite appliquer un tel développement aux coefficients aléatoires de l'opérateur dans l'équation (22.1)... et bien justement on ne connaît pas la fonction de covariance ni par conséquent ses fonctions propres. Il est nécessaire de trouver un autre développement qui contourne le problème.

Le chaos polynomial est une forme particulière du chaos homogène. Il est défini comme suit.

Définition 84 — Chaos polynomial. Le chaos polynomial de dimension M et d'ordre p est défini comme l'ensemble des polynômes d'Hermite multidimensionnels en les M variables aléatoires gaussiennes centrées réduites ξ_1, \dots, ξ_M de degré inférieur ou égale à p .

Chacun de ces polynômes est complètement défini par une liste de M entiers positifs ou nuls $\alpha_1, \dots, \alpha_M$ comme suit à partir des polynômes d'Hermite H_n :

$$\Psi_{\alpha} = \prod_{i=1}^M H_{\alpha_i}(\xi_i) \quad (22.28)$$

où α est la liste considérée.

L'ensemble des chaos polynomiaux est un sous-espace linéaire de l'espace des fonctions L^2 de variables aléatoires Θ défini plus haut, et est un anneau pour la multiplication fonctionnelle.

Déterminons la dimension de la base du chaos polynomial de dimension M et d'ordre p :

Démonstration. (On pourrait aboutir au résultat plus directement, mais nous en profitons pour parler des polynômes homogènes). Considérons l'anneau des polynômes en M indéterminées $\mathbb{R}[X_1, \dots, X_M]$. Un polynôme homogène, ou forme algébrique, est un polynôme en plusieurs indéterminées dont tous les monômes

non nuls sont de même degré total. L'ensemble des polynômes homogènes de degré p dans $\mathbb{R}[X_1, \dots, X_M]$ forme un \mathbb{R} -espace vectoriel. Sa base canonique est l'ensemble des monômes $X_1^{\alpha_1} X_2^{\alpha_2} \dots X_M^{\alpha_M}$ où $\alpha_1 + \alpha_2 + \dots + \alpha_M = p$. Sa dimension est donc le nombre de p -combinaisons avec répétition de l'ensemble $\{1, 2, \dots, M\}$, i.e. $\Gamma_M^p = \binom{M+p-1}{p}$. La dimension de cette base vaut $\sum_{i=0}^p \Gamma_M^i$, avec $\Gamma_M^0 = 1$. Tous calculs faits, La dimension de la base du chaos polynomial de dimension M et d'ordre p est donc donnée par :

$$\binom{M+p}{p} \quad (22.29)$$

□

Le tableau 22.1 donne le nombre de polynômes présents dans la base du chaos polynomial considéré en fonction du nombre M de variables aléatoires et du degré maximal p choisi.

	$p = 1$	$p = 2$	$p = 3$	$p = 4$
$M = 1$	2	3	4	5
$M = 2$	3	6	10	15
$M = 3$	4	10	20	35
$M = 4$	5	15	35	70
$M = 5$	6	21	56	126
$M = 6$	7	28	84	210

Tableau 22.1: Nombre de polynômes dans la base du chaos : M est le nombre de variables aléatoires, et p le degré maximal

22.3 Éléments finis stochastiques

Nous rappelons l'équation (22.1) :

$$au = f \text{ que l'on peut écrire maintenant : } a(x, \theta)u(x, \theta) = f(x, \theta)$$

où l'opérateur a est défini sur le produit d'un espace de Hilbert de type L^2 (en fonction du problème) par l'espace de Hilbert $L^2(\Theta, \mathcal{F}, P)$. C'est un opérateur différentiel dont les coefficients ont des fluctuations aléatoires dépendant d'une ou plusieurs variables.

On fait l'hypothèse que l'opérateur a est un opérateur différentiel dont les coefficients aléatoires sont contraints à être des processus stochastiques du second ordre. Cette hypothèse n'est pas sévère en fait, car la plupart des processus physiquement mesurables sont de ce type. On pourra donc écrire chacun des coefficients $a_i(x, \theta)$ de a comme la somme d'une composante déterministe $\bar{a}_i(x) = E[a_i(x, \theta)]$ et d'une composante aléatoire $\tilde{a}_i(x, \theta)$. On obtiendra à nouveau un système sous la forme :

$$(\bar{a}(x) + \tilde{a}(x, \theta))u(x, \theta) = f(x, \theta) \quad (22.30)$$

qui se résout en :

$$u = g(\tilde{a}_i(x, \theta), f(x, \theta)) \quad (22.31)$$

avec g une fonctionnelle non-linéaire.

22.3.1 Développement en séries de von Neumann

Notre problème consiste à résoudre l'équation (22.30), i.e. à inverser l'opérateur. Lorsque un opérateur est inversible, son inverse admet un développement en série convergent. La méthode de développement en séries de von Neumann consiste à écrire la solution de l'équation (22.30) sous la forme :

$$u(\tilde{a}(\theta), x) = \sum_{i=0}^{\infty} (-1)^i (\bar{a}^{-1}(x) \tilde{a}(x, \theta))^i f(x, \theta) \quad (22.32)$$

qui ne converge que si :

$$\|\bar{a}^{-1}(x) \tilde{a}(x, \theta)\| < 1 \quad (22.33)$$

L'intérêt est que l'on ne doit inverser que $\bar{a}^{-1}(x)$. Toutefois, on ne sait pas calculer symboliquement les termes d'ordre supérieur à deux, et on se contente de simulations numériques au delà. On peut faire les mêmes remarques que pour la méthode de perturbation et il est difficile d'obtenir les moments d'ordre supérieur à deux, surtout si l'on considère la valeur moyenne du processus sur chaque élément, comme nous allons l'illustrer immédiatement en mécanique.

22.3.2 Un exemple en statique linéaire

Considérons le cas de la statique linéaire décrite par le système :

$$\mathbf{K}\mathbf{q} = \mathbf{F}$$

mais où cette fois, le matériau utilisé est supposé avoir un module d'Young représenté par un champ aléatoire gaussien. La loi de comportement, traduite par la matrice d'élasticité, s'écrit alors :

$$\mathbf{H}(x, \theta) = w(x, \theta) \mathbf{H}_0 \quad (22.34)$$

où \mathbf{H}_0 est une matrice déterministe constante.

Chaque matrice de rigidité élémentaire, donnée par $\mathbf{k}^e = \int_{\Omega_e} \mathbf{N}_u^{-1} \mathcal{L}^{-1} \mathbf{H} \mathcal{L} \mathbf{N}_u d\Omega_e$ s'écrit, en reportant le développement de Karhunen-Loève de $w(x, \theta)$ est donnée par la relation (22.25) :

$$\mathbf{k}^e(\theta) = \mathbf{k}_0^e + \sum_{i=1}^{\infty} \mathbf{k}_i^e \xi_i(\theta) \quad (22.35)$$

avec

$$\mathbf{k}_i^e = \sqrt{\lambda_i} \int_{\Omega_e} \varphi_i \mathbf{N}_u^{-1} \mathcal{L}^{-1} \mathbf{H}_0 \mathcal{L} \mathbf{N}_u d\Omega_e \quad (22.36)$$

Suite à l'assemblage, on obtient le système :

$$\left(\mathbf{K}_0 + \sum_{i=1}^{\infty} \mathbf{K}_i \xi_i \right) \mathbf{q}(\theta) = \mathbf{F} \quad (22.37)$$

où \mathbf{K}_0 est la matrice de rigidité globale calculée aux valeurs moyennes, et les matrices \mathbf{K}_i des matrices déterministes obtenues par les techniques d'assemblage classique.

On réécrit le système précédent :

$$\mathbf{K}_0 \left(\mathbf{I} + \sum_{i=1}^{\infty} \mathbf{K}_0^{-1} \mathbf{K}_i \xi_i \right) \mathbf{q}(\theta) = \mathbf{F} \quad (22.38)$$

ce qui conduit à :

$$\mathbf{q}(\theta) = \left(\mathbf{I} + \sum_{i=1}^{\infty} \mathbf{K}_0^{-1} \mathbf{K}_i \xi_i \right)^{-1} \mathbf{q}_0, \quad \text{avec} \quad \mathbf{q}_0 = \mathbf{K}_0^{-1} \mathbf{F} \quad (22.39)$$

et le développement en série de von Neumann de cette dernière équation donne :

$$\mathbf{q}(\theta) = \sum_{k \geq 0} (-1)^k \left(\sum_{i=1}^{\infty} \mathbf{K}_0^{-1} \mathbf{K}_i \xi_i \right)^k \mathbf{q}_0 \quad (22.40)$$

En tronquant les développements de Karhunen-Loève et de von Neumann, on obtient une approximation de $\mathbf{q}(\theta)$. On pourra également noter la ressemblance avec ce qui a été exposé au paragraphe 13.6 sur la méthode de réanalyse.

22.3.3 Projection sur la base du chaos polynomial

Les équations (22.32) et (22.40) montrent que la solution s'exprime comme série polynomiale de variables gaussiennes centrées réduites ξ_i . Il est possible de réarranger tous les termes de l'équation (22.32) en l'indexant par un seul indice j sous la forme :

$$u^i(\theta) = \sum_{j=0}^{\infty} u_j^i \mathcal{P}_j \quad (22.41)$$

où les \mathcal{P}_j sont les polynômes homogènes en les variables gaussiennes centrées réduites ξ_k (avec $\mathcal{P}_0 \equiv 1$). Comme les polynômes homogènes forment une base de $L^2(\Theta, \mathcal{F}, P)$, les coefficients u_j^i sont les coordonnées de $u^i(\theta)$ dans cette base. Mais les polynômes homogènes ne forment pas une base orthogonale de $L^2(\Theta, \mathcal{F}, P)$ au sens du produit scalaire défini par la relation (22.12), contrairement au chaos polynomial. La solution peut donc également s'écrire dans la base du chaos polynomial sous la forme :

$$u(\theta) = \sum_{j=0}^{\infty} u_j \Psi_j \quad (22.42)$$

Il s'agit simplement d'un changement de base, et dans ce cas, même le premier terme de la série dans la base du chaos polynomial n'est pas le même que dans la série de von Neumann.

Pour en revenir à l'exemple précédent, cela signifie que l'on peut décomposer les déplacements dans la base du chaos polynomial : $\mathbf{q}(\theta) = \sum_{i \geq 0} \mathbf{q}_i \Psi_i(\theta)$, où les \mathbf{q}_i sont des vecteurs déterministes (de dimension égale au nombre de degrés de liberté). Dans ce cas, tous les termes sont différents du développement en série de von Neumann, même le premier \mathbf{q}_0 .

22.3.4 Obtention des coefficients du système matriciel à résoudre

Nous allons poursuivre notre exemple de statique linéaire pour obtenir explicitement le système à résoudre. Pour cela, nous substituons le développement sur la base du chaos polynomial donné par la relation (22.42) dans le développement de Karhunen-Loève de l'équation (22.37). En notant $\xi_0(\theta) \equiv 1$, on obtient :

$$\left(\sum_{i=0}^{\infty} \mathbf{K}_i \xi_i(\theta) \right) \left(\sum_{i=0}^{\infty} \mathbf{q}_i \Psi_i(\theta) \right) = \mathbf{F} \quad (22.43)$$

Et il est évidemment nécessaire de tronquer les séries dans cette équation pour pouvoir résoudre numériquement.

Nous supposons donc que le développement de Karhunen-Loève est tronqué à l'ordre $M + 1$ et que la base de chaos polynomial sera considérée jusqu'à l'ordre p . Le résidu correspondant à cette troncature de l'équation (22.43) est :

$$R_{M,p} = \sum_{i=0}^M \sum_{j=0}^{p-1} \mathbf{K}_i \mathbf{q}_j \xi_i(\theta) \Psi_j(\theta) - \mathbf{F} \quad (22.44)$$

La meilleure approximation de $\mathbf{q}(\theta)$ est obtenue en minimisant le résidu au sens des moindres carrés. Cela se produit lorsque le résidu est orthogonal au sous-espace engendré par les $\{\Psi_k\}_{k=0}^{p-1}$ dans $L^2(\Theta, \mathcal{F}, P)$, ce qui s'écrit :

$$E[R_{M,p} \cdot \Psi_k] = 0, \quad \forall k \in \{0, \dots, p-1\} \quad (22.45)$$

Maintenant, posons :

$$c_{ijk} = E[\xi_i \Psi_j \Psi_k] \quad (22.46)$$

$$\mathbf{F}_k = E[\Psi_k \mathbf{F}] \quad (22.47)$$

$$\mathbf{K}_{jk} = \sum_{i=0}^M c_{ijk} \mathbf{K}_i \quad (22.48)$$

On remarquera au passage que \mathbf{F}_k est nul pour tout $k > 0$ dans le cas d'un chargement déterministe (ce qui est le cas dans notre exemple).

On obtient finalement le système matriciel suivant à résoudre :

$$\begin{bmatrix} \mathbf{K}_{0,0} & \cdots & \mathbf{K}_{0,p-1} \\ \vdots & & \vdots \\ \mathbf{K}_{p-1,0} & \cdots & \mathbf{K}_{p-1,p-1} \end{bmatrix} \begin{pmatrix} \mathbf{q}_0 \\ \vdots \\ \mathbf{q}_{p-1} \end{pmatrix} = \begin{pmatrix} \mathbf{F}_0 \\ \vdots \\ \mathbf{F}_{p-1} \end{pmatrix} \quad (22.49)$$

qui est un système de taille p (ordre du développement dans la base du chaos polynomial) \times nombre de degrés de liberté du modèle déterministe.

Les blocs diagonaux du système ci-dessus \mathbf{K}_{jj} représentent la contribution due à la valeur moyenne. Plus les fluctuations des propriétés matérielles sont faibles, plus le système est diagonal par blocs. Le système reste par ailleurs symétrique, défini-positif et à caractère bande. Il peut être résolu directement par une factorisation de Cholesky, ou par une méthode hiérarchique.

22.3.5 Synthèse et remarques

Toute variable aléatoire X de $L^2(\Theta, \mathcal{F}, P)$ de fonction de répartition $F_X(x)$ peut être transformée en une variable gaussienne centrée réduite ξ en utilisant la relation $\xi = F_X^{-1}(\Phi(\xi))$ (où Φ est la fonction de répartition gaussienne centrée réduite) et donc se décomposer sur une base orthogonale de $L^2(\Theta, \mathcal{F}, P)$ telle que celle des polynômes d'Hermite :

$$X = \sum_{i=0}^{\infty} a_i H_i(\xi)$$

Les coefficients a_i , dans cette base, s'expriment comme :

$$a_i = \frac{E[XH_i(\xi)]}{E[H_i(\xi)^2]} = \frac{E[XH_i(\xi)]}{i!}$$

De la même manière, un vecteur aléatoire \mathbf{X} de M variables aléatoires indépendantes peut être représenté par M variables aléatoires gaussiennes centrées réduites ξ_i ($i = 1 \dots M$). Enfin, toute

fonction S suffisamment régulière d'une variables ou d'un vecteur aléatoire X à valeur scalaire ou vectorielle peut être décomposée sur la base du chaos polynomial composée des polynômes d'Hermite multidimensionnels (i.e. en M variables) de variables gaussiennes centrées réduites et s'écrire sous la forme :

$$S = \sum_{i=0}^{\infty} S_i \Psi(\xi_j), \quad j = 1 \dots M$$

La méthode des éléments finis stochastiques spectraux initialement proposée par Ghanem et Spanos avait pour but de modéliser la variabilité spatiale des propriétés matérielles d'un système mécanique, puis de propager cette incertitude à travers le modèle. Le module d'Young est modélisé comme un champ gaussien $E(x)$ et un développement en série de Karhunen-Loève permet de le représenter à partir d'une famille de variables gaussiennes centrées réduites :

$$E(x) = \mu(x) + \sum_{i=0}^{\infty} \sqrt{\lambda_i} \varphi_i(x) \xi_i$$

à partir des valeurs propres λ_i et des vecteurs propres φ_i de la fonction d'autocovariance du champ aléatoire. En pratique, on n'utilisera évidemment qu'un nombre limité M de termes, correspondant aux valeurs propres les plus grandes.

La matrice de rigidité est alors elle-aussi aléatoire et on peut l'exprimer aussi bien dans la base de variables gaussiennes que dans la base du chaos polynomial :

$$\mathbf{K} = \sum_{i=0}^{\infty} \mathbf{K}_i \xi_i = \sum_{j=0}^{\infty} \mathbf{K}_j \Psi_j$$

Si le champ aléatoire considéré n'est plus gaussien (par exemple lognormal), ou si la dépendance de la rigidité en fonction de ce champ n'est plus linéaire, Ghanem montre que la matrice de rigidité peut encore être représentée sur la base du chaos polynomial. On a montré que cela est encore valable quelque soit la loi des variables aléatoires.

Dans l'exemple présenté, le chargement était considéré comme déterministe, mais il peut lui-aussi être modélisé par des variables aléatoires de loi quelconque du moment qu'elles sont décomposées sur la base du chaos polynomial, et on obtient le système discrétisé suivant :

$$\left(\sum_{i=0}^{\infty} \mathbf{K}_i \Psi_i \right) \left(\sum_{j=0}^{\infty} \mathbf{q}_j \Psi_j \right) = \left(\sum_{k=0}^{\infty} \mathbf{F}_k \Psi_k \right) \quad (22.50)$$

La résolution numérique se fait en tronquant les séries de sorte que le résidu associé soit minimal au sens de Galerkin.

Le processus stochastique porte sur des propriétés liées au domaine considéré Ω . Le plus simple des choix est donc de considérer la même discrétisation pour le modèle éléments finis que pour la discrétisation du processus stochastique. Toutefois, le maillage du domaine Ω correspond au problème traité et plus particulièrement à l'obtention de certaines quantités : pour un calcul mécanique, le maillage doit permettre d'appréhender correctement le champ de contraintes. Or ce maillage « compatible » avec les contraintes n'a strictement aucune raison d'être adapté à la discrétisation du champ stochastique ! C'est pourquoi il est souvent préférable d'utiliser deux maillages différents. La minimisation du résidu dans le choix de la troncature des séries essaye de « faire au mieux » avec les nœuds disponibles, mais cela peut se révéler insuffisant.

Dans la mise en place de telles méthodes, il convient de comparer les résultats avec des résultats de référence... or, comme nous l'avons vu, à part dans de très rares cas, il n'est pas possible d'obtenir des formulations simples et donc de disposer de résultats analytiques. Dans ce cas, la méthode Monte-Carlo est utilisée. Le nom de ces méthodes, qui fait allusion aux jeux de hasard pratiqués à Monte-Carlo, a été inventé en 1947 par Nicholas Metropolis, et publié pour la première fois

en 1949 dans un article coécrit avec Stanislaw Ulam. Pour chacune des variables aléatoires, un grand nombre de tirages aléatoires est effectué dans les distributions de probabilité déterminées précédemment, afin de trouver la probabilité d'occurrence de chacun des résultats en appliquant la loi des grands nombres. Il s'agit d'une méthode simple et brutale... mais qui permet à coup sûr d'obtenir un résultat que l'on peut considérer comme la référence... à condition d'avoir effectué un nombre suffisamment élevé de tirages !

22.4 Post-traitement : moments statistiques

Quels que soient les paramètres considérés comme aléatoires, le vecteur des déplacements nodaux et les tenseurs de déformations et de contraintes en tout point peuvent s'écrire sous la forme d'un développement sur la base du chaos polynomial. Nous considérerons disposer à présent de la réponse du système S (déplacement, déformation, contrainte) écrite dans la base du chaos polynomial :

$$S = \sum_{i=0}^{p-1} s_i \Psi_i \quad (22.51)$$

Les post-traitements vont s'effectuer sur une expression analytique et ne coûtent presque rien en terme de temps de calcul. On peut ainsi se focaliser sur une analyse en tendance centrale (les premiers moments statistiques de la réponse) ou bien effectuer des analyses de fiabilité (et donc s'intéresser aux queues de distribution). La densité de probabilité de la réponse peut également être obtenue de différentes manières.

À partir de la relation (22.51), on obtient les moments :

$$\text{Moyenne : } E[S] = s_0 \quad (22.52)$$

$$\text{Variance : } \sigma_S^2 = \sum_{i=1}^{p-1} E[\Psi_i^2] s_i^2 \quad (22.53)$$

$$\text{Coefficient d'asymétrie : } \delta_S = \frac{1}{\sigma_S^3} \sum_{i=1}^{p-1} \sum_{j=1}^{p-1} \sum_{k=1}^{p-1} E[\Psi_i \Psi_j \Psi_k] s_i s_j s_k \quad (22.54)$$

$$\text{Coefficient d'aplatissement : } \kappa_S = \frac{1}{\sigma_S^4} \sum_{i=1}^{p-1} \sum_{j=1}^{p-1} \sum_{k=1}^{p-1} \sum_{l=1}^{p-1} E[\Psi_i \Psi_j \Psi_k \Psi_l] s_i s_j s_k s_l \quad (22.55)$$

L'obtention simple de ces moments statistiques justifie pleinement l'utilisation de la base de chaos polynomial... c'est pourquoi notre présentation s'est focalisée dessus.

22.5 Fiabilité des structures

22.5.1 Définitions

La fiabilité des structures vise à calculer la probabilité de défaillance d'une structure en tenant compte de la variabilité de paramètres (géométrie, matériau, chargement). Ces paramètres sont représentés par des variables aléatoires X, et S est la réponse du système.

Pour chaque mode de défaillance de la structure, une fonction d'état limite $g(X, S(X))$ est définie dans l'espace des paramètres, à laquelle on associe l'hypersurface $\{g(X, S(X)) = 0\}$, appelée surface d'état limite, qui sépare le domaine de sûreté $D_s = \{X | g(X, S(X)) > 0\}$ du domaine de défaillance $D_f = \{X | g(X, S(X)) \leq 0\}$.

La probabilité de défaillance de la structure P_f est définie par :

$$P_f = \text{Prob}[g(X, S(X)) \leq 0] = \int_{g(X, S(X)) \leq 0} f_X(x) dx \quad (22.56)$$

À partir des résultats d'un calcul par éléments finis stochastiques, on dispose de la représentation (approchée) de $S(X)$ aussi bien sous forme de série dans la base du chaos polynomial que sous forme de série en les variables gaussiennes centrées réduites.

En appelant R la résistance et S le chargement (notation classique), une illustration est donnée à la figure 22.1.

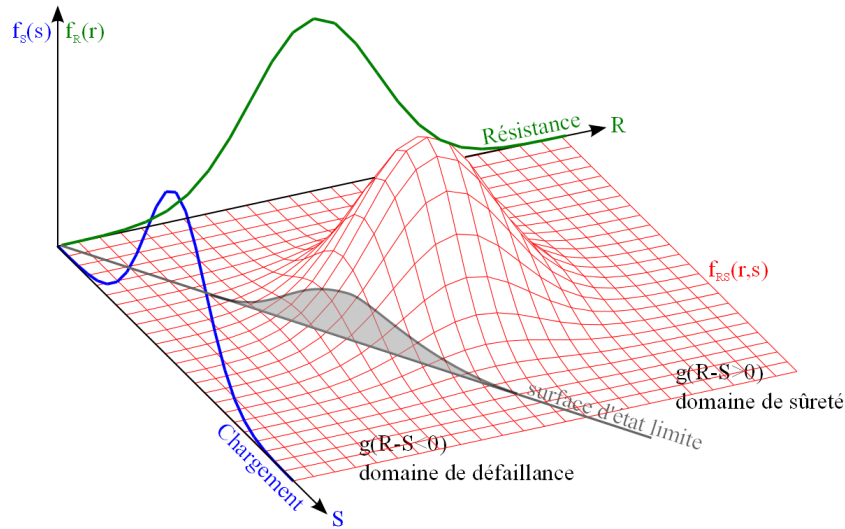


FIGURE 22.1: Illustration des concepts de fiabilité

En se plaçant dans l'espace de ces variables gaussiennes centrées réduites, dit espace réduit, la fonction d'état limite représente toujours la limite entre les domaines de sûreté et de défaillance, mais de sorte que le point de cette hypersurface le plus proche de l'origine, appelé point de conception, est le point pour lequel la probabilité de défaillance est maximale. Cette distance à l'origine est appelée indice de sûreté β , et on a donc :

$$P_f = \Phi(-\beta) \quad (22.57)$$

soit :

$$\beta = -\Phi^{-1}(P_f) \quad (22.58)$$

où Φ est la fonction de répartition gaussienne centrée réduite. Toutefois, déterminer ce point (i.e. le minimum de la fonction d'état limite) n'est pas si facile. Tous les algorithmes d'optimisation convergent vers un minimum, mais aucun ne peut affirmer que ce minimum est global.

Histoire

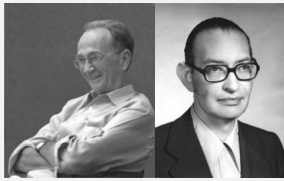
Le tout premier essai dans l'utilisation des concepts statistiques pour la fiabilité des structures date de 1926, par Max Mayer dans *Die Sicherheit der Bauwerke und ihre Berechnung nach Grenzkraften anstatt nach zulässigen Spannungen*. De nombreux progrès ont été faits par M. Prot (*Note sur la notion de coefficient de sécurité*, 1936), W. Weibull (*Investigations into strength of properties of brittle materials*, 1938, *A statistical theory of strength of materials*, 1939), W. Kjellman (*Säkerhetsproblemet ur principiell och teoretisk synpunkt*, 1940), et G. Wästlund (*Säkerhetsproblemet ur praktisk-konstruktiv synpunkt*, 1940), mais globalement, très peu d'articles avaient été publiés dans le domaine avant la seconde guerre mondiale.

Après 1945, le nombre d'article croît constamment. A l'université de Columbia, A.M. Freudenthal créa un institut pour l'étude de la fatigue et de la fiabilité, et produisit un grand nombre d'articles. L'évolution des codes modernes a été grandement influencée par les articles de 1949 à 1952 de E. Torroja et A. Pérez (par exemple *Calcul du coefficient de sécurité*, 1952). En 1953, dans sa thèse *Strength, safety and economical dimension of structures*, A.I. Johnson suggéra d'utiliser les distributions de valeurs extrêmes, celles-ci étant étudiées par E.J. Gumbel (*Statistics of Extremes*, 1958).



Freudenthal Gumbel

J. Ferry Borges en 1952 pointa l'importance de prendre en compte le caractère aléatoire des dimensions et des propriétés mécaniques dans celui du comportement structurel. Dans un rapport établi en 1962 par A.M. Freudenthal il est suggéré de représenter le chargement par une distribution en valeurs extrêmes et la résistance pour une distribution normale logarithmique.



Cornell Rosenblueth

En 1967, C.A. Cornell dans deux articles (*Bounds on the reliability of structural systems* et *A probability-based structural code*) propose une représentation du chargement, de la résistance et des dimensions par leurs moyenne et variance, ce qui devient la base des méthodes dites de niveau II^a. D'importantes contributions dans ce domaine seront dues à E. Rosenblueth et L. Esteva en 1972, O. Ditlevsen en 1973, A.M. Hasofer et N.C. Lind en 1974 et D. Veneziano en 1974. Nous allons en dire quelques mots un peu plus bas (indices de fiabilité).

En 1968, dans un article sur la conception sous chargement sismique (*Probabilistic models for seismic force design*) J.R. Benjamin défend l'utilisation des concepts probabilistes Bayésiens. Quant aux recherches des scientifiques russes sur le sujet, elles sont publiées par V.V. Bolotin en 1969 dans *Statistical methods in structural mechanics*.

Depuis 1970, de très nombreux articles ont été publiés sur le sujet. N'en mentionner que quelques uns est faire offense à de nombreux très bons autres articles. Nous citerons encore quelques grands noms du domaine tels que J. Armit (*wind structures*, 1976), F. Moses (*Reliability of structural systems*, 1976), R. Rackwitz (*First order reliability theories and stochastic models*, 1977), H. Karadeniz (logiciel SAPOS), A. Der Kiureghian, M. Shinozuka, M. Lemaire...



Rackwitz Karadeniz Der Kiureghian Shinozuka Lemaire

Revenons un peu sur L'indice de fiabilité utilisé en analyse de risque. C'est une mesure de la sûreté qui est élevée lorsque la probabilité de défaillance P_f est faible. Il s'agit d'un outil plus rudimentaire que la probabilité de défaillance qui est utilisé si celle-ci est trop grande ou lorsque son calcul est trop incertain à cause de l'approximation faite ou parce que l'on ne dispose pas d'informations suffisantes pour la calculer.

En 1969, Cornell définit l'indice de fiabilité β_c (dans le cas de variables gaussiennes) comme :

$$\beta_c = \frac{E[M]}{D[M]} \quad (22.59)$$

L'idée est que la distance du point de mesure $E[M]$ à la surface d'état limite est une bonne mesure de la fiabilité. Cette distance est mesurée par rapport à une unité d'incertitude $D[M]$ qui peut être la variance par exemple. On appelle M la marge de sécurité.

Si l'on note r la résistance et s le chargement correspondant, alors $g(r, s) = r - s$ et $M = R - S$. On peut alors écrire :

$$\beta_c = \frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}}$$

En 1972, Rosenblueth et Esteva proposent un nouvel indice de fiabilité. Dans le cas où $R > 0$

et $S > 0$, on peut écrire $g(r, s) = \log(r/s)$, et $M = \log(r/s)$, d'où :

$$\beta_{RE} = \frac{E[\log(r/s)]}{D[\log(r/s)]} \quad (22.60)$$

qui est calculé en linéarisant la marge de sécurité M . Un développement de Taylor au premier ordre donne :

$$M \approx \frac{\log \mu_R - \log \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}}$$

En 1974, Hasofer et Lind proposent de passer de l'espace des variables X à celui de variables normalisées et non corrélées. Ce sont eux qui donnent la définition de l'indice de fiabilité β_{HL} comme la distance minimale de l'origine à la surface d'état limite. En 1976, Ditlevsen propose un autre indice de fiabilité, plus sélectif, en introduisant une mesure de fiabilité obtenue en intégrant une fonction pondérée sur le domaine de sûreté. En pratique, on a souvent $\beta_G = \beta_{HL}$.

a. On distingue les méthodes de :

- niveau I : l'aspect probabiliste est introduit en donnant aux variables aléatoires une « valeur caractéristique » associée à un facteur de sécurité partiel ;
- niveau II : ce sont les méthodes fiabilistes utilisant deux paramètres pour décrire chaque variable aléatoire (moyenne et variance) ;
- niveau III : elles procèdent à l'analyse complète du problème et impliquent l'intégration de la fonction de densité de probabilité conjointe multidimensionnelle des variables aléatoires étendue sur le domaine de la sécurité. La fiabilité est exprimée en termes d'indices de sécurité adéquates, à savoir : indice de fiabilité et probabilité de défaillance ;
- niveau IV : elles concernent les structures qui sont d'une importance économique majeure, et impliquent l'utilisation des principes de l'analyse économique de l'ingénierie sous incertitude. Elles tiennent compte des coûts et des bénéfices de la construction, de la maintenance, de la réparation, des conséquences de la défaillance, des intérêts sur le capital... Elles doivent être utilisées pour les projets sensibles comme les projets nucléaires, les tours de transmission, les ponts routiers.

22.5.2 Méthodes FORM/SORM

La probabilité de défaillance P_f donnée à l'équation (22.57) est rarement utilisable directement, le domaine d'intégration étant défini implicitement. La méthode FORM (First Order Reliability Method) permet d'obtenir une approximation de cette intégrale.

On commence par réécrire l'équation (22.57) dans l'espace réduit. Pour cela on utilise une transformation isoprobabiliste $T : X \rightarrow U(X)$. Les transformations couramment utilisées sont celles de Rosenblatt ou Nataf¹. En utilisant ce changement de variable, on obtient :

$$P_f = \int_{T^{-1}(U), S(T^{-1}(U))} \varphi_n(U) du_1 \dots du_n \quad (22.61)$$

avec φ_n la densité de probabilité multinormale de dimension n :

$$\varphi_n(u) = \frac{1}{(\sqrt{2\pi})^n} \exp\left(-\frac{1}{2}(u_1^2 + \dots + u_n^2)\right) \quad (22.62)$$

Cette densité de probabilité est maximale à l'origine et décroît exponentiellement avec $\|u\|^2$. Les points contribuant le plus à l'intégrale sont donc ceux appartenant au domaine de défaillance les plus proches de l'origine.

1. La transformation de Rosenblatt transforme un vecteur X de loi quelconque en un vecteur U de même dimension mais à composantes indépendantes, gaussiennes, centrées et réduites. La transformation de Nataf transforme un vecteur ayant une loi à copule elliptique en un vecteur de loi sphérique associée au représentant elliptique de la copule. Une copule est une fonction de répartition définie sur $[0, 1]^N$ dont les lois marginales sont égales à la loi uniforme sur $[0, 1]$. Nous nous passerons volontiers de cela dans ce document de présentation. Pour la petite histoire, Murray Rosenblatt a passé sa thèse sous la direction de Mark Kac.

Puis on détermine le point de conception P, i.e. le point D_f le plus proche de l'origine :

$$P = \min_{\|u\|} \{g(T^{-1}(U), S(T^{-1}(U))) \leq 0\} \quad (22.63)$$

en utilisant un algorithme d'optimisation pour résoudre ce problème de minimisation.

Enfin on approche le domaine d'intégration D_f dans (22.61) par le demi-espace défini par l'hyperplan tangent à D_f en P. L'intégration est alors analytique, et on obtient l'équation (22.57).

Les méthodes SORM (Second Order Reliability Method) sont des extensions de la méthode FORM. L'une d'elle consiste à approcher la surface d'état limite par un hyperparaboloïde d'ordre 2 au voisinage du point de conception. Si $(\kappa_1, \dots, \kappa_{n-1})$ désignent les courbures principales de cet hyperparaboloïde, l'approximation SORM de la probabilité de défaillance conduit à la formule de Breitung :

$$P_f = \Phi(-\beta) \prod_{i=1}^{n-1} \frac{1}{\sqrt{1 - \beta \kappa_i}} \quad (22.64)$$

22.5.3 Tirages d'importance

Les simulations de Monte Carlo sont les méthodes les plus simples à mettre en œuvre pour calculer la probabilité de défaillance d'un système, mais ce sont aussi les plus coûteuses. Après avoir généré des réalisations pour les variables aléatoires d'entrée selon leur densité conjointe de probabilité, on évalue la fonction d'état limite pour ces réalisations. On compte le nombre total de cas défaillants parmi les calculs effectués, ce qui permet d'estimer la probabilité en fonction du nombre de cas défaillants par rapport au nombre total des réalisations effectuées. Pour s'assurer de la convergence des méthodes de simulation, il faut calculer le coefficient de variation de la simulation : la convergence est atteinte pour un coefficient de variation d'environ 5%.

Les simulations d'importance (ou Importance sampling) sont des méthodes efficaces pour estimer la probabilité de défaillance, car elles permettent de générer des tirages qui conduisent plus fréquemment à la défaillance et permettent de gagner plus d'information sur le domaine de défaillance.

Comme le poids de la probabilité de défaillance est généralement situé au voisinage du point de conception, il est plus efficace de concentrer les tirages autour de ce point. On commence donc par l'approcher, par exemple par la méthode FORM. On introduit la fonction $\psi(u)$, dite densité d'importance, que l'on peut prendre égale à la densité multinormale réduite centrée au point de conception de coordonnées u^* . Dans ce cas, pour simuler U suivant la densité $\psi(U)$, on génère d'abord une variable multinormale centrée réduite U_r et on effectue le changement de variables $U = U_r + U^*$. Pour la simulation de variables aléatoires, on pourra consulter [49].

Chapitre 23

Quelques méthodes dérivées

Résumé — Dans ce court chapitre, nous survolons quelques méthodes également utilisées en simulation numérique. Nous n’entrons pas dans le détail, mais si les notions d’éléments finis, de formulations mixtes et hybrides et les multiplicateurs de Lagrange ont été comprises, alors nos courtes explications doivent suffire.

23.1 Méthode des éléments frontières

La méthode des éléments finis et la méthode des éléments frontière (Boundary Element Method) peuvent être considérées comme issues des méthodes de Ritz et de Trefftz respectivement. Dans les deux cas, il s’agit de résoudre un problème décrit par des équations aux dérivées partielles dans un domaine Ω et sur sa frontière $\Gamma = \partial\Omega$ en remplaçant le problème continu par un nombre fini de paramètres inconnus dans la résolution numérique.

La méthode de Ritz, et donc la méthode des éléments finis, se base sur l’existence d’un principe variationnel pour lequel la fonction inconnue sera recherchée comme une combinaison de fonctions de base définies dans tout le domaine Ω .

La méthode de Trefftz a été proposée en 1926 dans un article intitulé « une alternative à la méthode de Ritz ». Le passage domaine/frontière se fait en appliquant le théorème de Green à la formulation variationnelle considérée et en proposant des fonctions d’interpolation linéairement indépendantes qui satisfont les équations aux dérivées partielles d’intérieur de domaine *a priori*. Trefftz montre que l’intégrale d’intérieur de domaine du principe variationnel disparaît et que seules subsistent des intégrations aux frontières. Cependant, cette méthode discrétisant la frontière seule peut être employée uniquement si le problème physique considéré est gouverné par des équations différentielles linéaires et homogènes. On obtient encore un système de type $\mathbf{Kq} = \mathbf{F}$, mais les intégrales se font uniquement sur Γ .

Cette méthode est préférée à la méthode des éléments finis lorsque le domaine de propagation devient infini puisque seule la surface de la frontière du domaine doit être discrétisée (i.e. on n’a pas besoin de mailler tout le domaine de propagation comme avec les éléments finis) et on gagne ainsi une dimension dans la discrétisation du problème (le champ considéré (acoustique par exemple) en tout point de l’espace étant dû au rayonnement de ses frontières). En outre la condition de Sommerfeld de rayonnement à l’infini est satisfaite automatiquement dans les méthodes d’éléments finis de frontière via le noyau des formulations intégrales qui répond exactement aux conditions aux limites pour des domaines infinis, contrairement aux méthodes d’éléments finis. Toutefois, on pourra appliquer la méthode des éléments finis en utilisant les éléments infinis ou des techniques de type « perfectly matched layer » et champs lointains, comme vu au chapitre 19.

La méthode des éléments finis et la méthode des éléments frontières sont toutes deux des méthodes très robustes pour de nombreuses applications industrielles, ce qui en fait des outils de choix pour les simulations numériques. Cependant, leur utilisation pose des problèmes lorsque

l'on a évolution des surfaces internes à cause du « maillage » de ces surfaces : description correcte, évolution de ces surfaces, fissures, front de sollicitation, interfaces... Une des motivations pour les méthodes sans maillage est de s'affranchir de ces difficultés.

23.2 Méthodes particulières ou sans maillage

En grandes déformations par exemple, le maillage va nécessairement être très déformé au cours du calcul, conduisant à une dégénérescence du jacobien ayant pour conséquence une perte de précision, des problèmes de convergence... et il est nécessaire alors de recourir à des méthodes de remaillage adaptatif qui sont coûteuses en temps de calcul et posent encore des problèmes dans le cas tridimensionnel. Sans raffinement de maillage, des phénomènes de localisation de la déformation, qui suit le bord des éléments se produit, et on constate, par exemple dans le cas de la simulation de fissures, que le chemin de fissuration varie fortement avec le maillage... il faut alors développer des éléments spéciaux.

Pour palier ces problèmes ont été développées, dès les années 70 puis améliorées et popularisées dans les années 90, de nombreuses méthodes où la discrétisation ne repose plus sur un maillage mais sur un ensemble de points¹. L'idée est de construire une fonction définie sur un espace continu à partir de l'ensemble des valeurs discrètes prises par cette fonction sur un nuage de points du domaine. On parle alors de méthode particulière (dont une autre application sera donnée au paragraphe 23.5). Les problèmes de déformation de maillage ne peuvent survenir puisqu'il n'y a plus de connectivité entre les nœuds ; le remaillage et la gestion des frontières internes (fissures...) sont résolus puisqu'il s'agit alors simplement d'ajouter des nœuds au grès des besoins. Chaque point possède un domaine d'influence de forme simple, comme un cercle, sur lequel les fonctions d'approximation sont construites. Les différentes approches se distinguent, entre autres, par les techniques utilisées pour la construction des fonctions d'approximation. Ces fonctions sont construites de manière à pouvoir représenter tous les modes rigides et de déformation sur le domaine d'influence (c'est une condition nécessaire à la convergence de ces méthodes), et sont nulles en dehors. De manière générale, on écrira alors :

$$u(x) = \sum_{i \in N_s(x)} \sum_{\alpha=1}^{N_f(i)} a_i^\alpha \phi_i^\alpha(x) \quad (23.1)$$

où $N_s(x)$ est l'ensemble des points i dont le support contient le point x et $N_f(i)$ est le nombre de fonctions d'interpolation définies sur le support associé au point i .

Une fois les fonctions d'interpolation construites, il est possible d'en ajouter d'autres par enrichissement. L'enrichissement de l'approximation permet ainsi de représenter un mode de déplacement donné $F(x)\phi_i^\alpha(x)$:

$$u(x) = \sum_{i \in N_s(x)} \sum_{\alpha=1}^{N_f(i)} a_i^\alpha \phi_i^\alpha(x) + \sum_{i \in N_s(x) \cap N_f} \sum_{\alpha=1}^{N_f(i)} b_i^\alpha \phi_i^\alpha(x) F(x) \quad (23.2)$$

L'enrichissement des fonctions d'interpolation a par exemple permis de résoudre des problèmes de propagation de fissure en deux et trois dimensions sans remaillage : la fissure se propage à travers un nuage de points et est modélisée par enrichissement, à savoir des fonctions $F(x)$ discontinues sur la fissure ou représentant la singularité en fond de fissure.

1. On peut par exemple nommer les méthode des noyaux régularisants, des éléments diffus ou discrets (DEM), de Galerkin sans éléments (Element Free Galerkin Method, EFGM), Smoothed Particule Hydrodynamics (SPH), Reproducing Kernel Particle Method (RKPM), $h - p$ Cloud Method, Moving Particle Simulation (MPS), Particle Finite Element Method (PFEM), Material Point Method (MPM), Molecular Dynamics (MD), Lattice-Boltzmann-Method (LBM)... Les méthodes que nous verrons dans les paragraphes suivants : partition de l'unité, éléments finis étendue, treillis de Boltzmann sont issues, dans une certaine mesure, des méthodes sans maillage.

En résumé, pour développer une méthode sans maillage, on a besoin : 1) d'un nuage de points, 2) d'une méthode d'interpolation, comme par exemple la formule (23.1), et 3) d'une formule de quadrature permettant d'exprimer la formulation faible que l'on souhaite traiter sous forme matricielle. Les méthodes sans maillage (Meshless Methods) sont flexibles dans le choix de l'approximation et de l'enrichissement. Elles comportent néanmoins quelques inconvénients. Tout d'abord, on remarquera que la méthode d'interpolation n'est pas, contrairement au cas de la méthode des éléments finis, interpolante. Il s'en suit que dans nombre de méthodes sans maillage, l'imposition des conditions aux limites de Dirichlet est plus compliquée à réaliser. On ne peut pas non plus utiliser les schémas d'intégration numérique de la méthode des éléments finis car les fonctions de forme ne sont plus polynomiales et que l'on ne dispose pas d'un maillage pour placer les points d'intégration. Le choix du nombre de points d'intégration et de la taille du domaine d'influence dans un nuage arbitraire de points d'approximation n'est pas trivial : alors que dans le cas de la méthode des éléments finis, la matrice de rigidité est obtenue par assemblage des contributions élémentaires, pour les méthodes sans maillage, l'assemblage se fait en couvrant le domaine de points d'intégration et en ajoutant leur contribution. De plus, chaque point ayant généralement plus de voisins que dans la méthode des éléments finis, la matrice de « rigidité » est moins creuse. Enfin, dans nombre de cas, il est nécessaire de faire varier la taille du domaine d'influence dans l'espace et le temps pour mieux suivre le problème : la raréfaction des particules en cours de calcul dans certaines zones conduit à des instabilités dues par exemple à l'augmentation des erreurs numériques dans le calcul de dérivées (mais augmenter la zone d'influence correspond à augmenter le nombre de voisins donc à alourdir le calcul...).

23.3 Partition de l'unité

La base des éléments finis classiques peut elle-aussi être enrichie (afin de ne pas recourir aux méthodes sans maillage) de manière à représenter une fonction donnée sur un domaine donné. Melenk et Babuška, en 1996, ont appelé cette technique : la Partition de l'Unité (PU). Ils ont remarqué en fait que si $N_s(x)$, autrement dit l'ensemble des points i dont le support contient le point x , était remplacé par $N_n(x)$, c'est-à-dire l'ensemble des nœuds contenant x , alors on retomberait sur la formulation éléments finis usuelle, qui, sous forme « assemblée », s'écrivait bien alors :

$$u(x) = \sum_{i \in N_n(x)} \sum_{\alpha} a_i^{\alpha} \phi_i^{\alpha}(x) \quad (23.3)$$

Cela veut simplement dire que, dans le cas des éléments finis, on considère les supports comme liés à des nœuds, alors que dans le cas sans maillage, ils sont liés à des points. En d'autres termes, si l'on considère un nœud à chaque point, la méthode des éléments finis est un cas particulier de la méthode sans maillage. On peut alors également écrire l'approximation de partition de l'unité par :

$$u(x) = \sum_{i \in N_n(x)} \sum_{\alpha} a_i^{\alpha} \phi_i^{\alpha}(x) + \sum_{i \in N_n(x) \cap N_f} \sum_{\alpha} b_i^{\alpha} \phi_i^{\alpha}(x) F(x) \quad (23.4)$$

23.4 Méthode des éléments finis étendue

Dans le cas particulier où la méthode de partition de l'unité est appliquée à la modélisation de discontinuités ou de vides au sein même des éléments, on obtient alors la méthode des éléments

finis étendue (X-FEM) :

$$u(x) = \sum_{i \in I} u_i \phi_i(x) + \sum_{i \in L} a_i \phi_i(x) H(x) + \sum_{i \in K_1} \phi_i(x) \left(\sum_{j=1}^4 b_{i,1}^j F_1^j(x) \right) + \sum_{i \in K_2} \phi_i(x) \left(\sum_{j=1}^4 b_{i,2}^j F_2^j(x) \right) \quad (23.5)$$

où I est l'ensemble des nœuds du maillage, L l'ensemble des nœuds enrichis pour modéliser la fissure coupant de part en part un élément, K_1 et K_2 les nœuds enrichis pour modéliser le fond de fissure.

L'avantage en X-FEM est qu'il n'est plus demandé au maillage de se conformer à des surfaces, qu'elles soient intérieures ou extérieures, et qu'il peut alors être conservé lors de leur évolution. Les surfaces ne sont plus maillées et sont localisées sur le maillage grâce à la notion de fonction de niveau. À chaque nœud au voisinage de cette surface, on associe la distance signée à cette surface. Cette fonction « distance » peut être interpolée sur chaque élément avec les fonctions classique de premier ordre. Les surfaces sont ainsi stockées par un champ élément fini défini au voisinage de la surface qui participe au calcul au même titre que les autres champs physiques. En particulier, la X-FEM permet la modélisation des trous, sans avoir à forcer le maillage à se conformer à ceux-ci. Un nœud dont le support est complètement à l'intérieur du trou ne donne pas lieu à la création de degrés de liberté. Pour un nœud dont le support coupe la frontière du trou, la fonction d'interpolation classique est multipliée par une fonction valant 1 dans la matière et 0 dans le trou.

23.5 Méthodes de treillis de Boltzmann

La dénomination « méthodes particulières » recouvre deux types différents de modèles pour la mécanique du solide et pour la mécanique des fluides. D'un côté, on trouve des concepts de discrétisation dans lequel la réponse d'un continuum est projetée sur les « particules » véhiculant l'information mécanique au cours de leurs déformations (SPH, MPS, PFEM, MPM, et Lattice-Boltzmann-Method). De l'autre côté, cette notion exprime la représentation de calcul de particules physiques existantes à différentes échelles (MD, DEM). On peut alors considérer les cas où les particules existent physiquement (comme pour les matières granulaires) ainsi que les cas où elles évoluent au cours du processus de chargement, ces deux cas pouvant même éventuellement être couplés.

Dans les méthodes de treillis de Boltzmann (Lattice Boltzmann Methods, LBM), on ne cherche plus à simuler un problème de dynamique des fluides par les équations de Navier-Stokes (nous avons déjà évoqué certains problèmes que l'on rencontre alors), mais en se servant de l'équation de Boltzmann discrète pour un fluide newtonien avec modèle de collision (par exemple Bhatnagar-Gross-Krook). L'idée est de simuler l'écoulement ainsi que les processus de collision entre un nombre limité de particules. Les interactions entre particules conduisent à un comportement de l'écoulement visqueux applicables à une échelle plus grande. On peut donc les voir également comme des méthodes de décomposition de domaine ou d'homogénéisation [paragraphe 13.4.2 et chapitre 15].

Les LBM sont particulièrement bien adaptées à la simulation de fluides autour de géométries complexes et ont l'avantage de pouvoir être implémentées sur des machines parallèles. Bien qu'historiquement développées pour les gaz en treillis, elles s'obtiennent également directement à partir des équations simplifiées de Boltzmann BGK (Bhatnagar-Gross-Krook).

On considère un treillis discret dont les nœuds portent des particules. Ces particules sautent d'un nœud au suivant selon leur vitesse : c'est la phase de propagation. Puis les particules se choquent et acquièrent une nouvelle vitesse : c'est la phase de collision. La simulation procède en alternant les phases de propagations et de collisions des particules. On montre que de tels gaz suivent les

équations des fluides de Navier-Stokes. L'inconvénient majeur de cette méthode appliquée à la dynamique des fluides est l'apparition de « bruit ». Si l'on ne cherche qu'un champ relativement calme (peu de variations, en tous cas, pas de variations brutales), il est nécessaire de pouvoir prendre la moyenne sur un treillis relativement grand et sur une grande période de temps. Dans ce cas, les LBM contournent le problème en pré-moyennant le gaz : on considère la distribution des particules sur le treillis plutôt que les particules elles-mêmes.

La forme générale de l'équation de treillis de Boltzmann est :

$$f_i(x + \Delta_t \mathbf{c}_i, t + \Delta_t) = f_i(x, t) + \Omega_i \quad (23.6)$$

où f_i est la concentration de particules se déplaçant avec la vitesse \mathbf{c}_i jusqu'au nœud suivant pendant un temps Δ_t . Ω_i représente l'opérateur de collision, et c'est lui qui change d'une méthode à l'autre. Dans le modèle BGK (Bhatnagar-Gross-Krook), la distribution des particules après propagation est relaxée vers la distribution à l'équilibre f_i^{eq} , et on a :

$$\Omega_i = \frac{1}{\tau} (f_i(x, t) - f_i^{eq}(x, t)) \quad (23.7)$$

avec τ le paramètre de relaxation, qui détermine la viscosité cinématique ν du fluide selon la relation $\nu = (2\tau - 1)/6$. La distribution à l'équilibre est une fonction de la densité locale ρ et de la vitesse locale \mathbf{u} , qui sont les moments d'ordre 1 et 2 de la distribution des particules :

$$\rho(x, t) = \sum_i f_i(x, t) \quad \text{et} \quad \mathbf{u}(x, t) = \frac{\sum_i f_i(x, t) \mathbf{c}_i}{\rho(x, t)} \quad (23.8)$$

Cette distribution à l'équilibre est calculée par la relation :

$$f_i^{eq}(\rho, \mathbf{u}) = t_p \rho \left(1 + \frac{\mathbf{c}_i \cdot \mathbf{u}}{c_s^2} + \frac{(\mathbf{c}_i \cdot \mathbf{u})^2}{2c_s^4} - \frac{\mathbf{u} \cdot \mathbf{u}}{2c_s^2} \right) \quad (23.9)$$

où c_s est la vitesse du son, $p = \mathbf{c}_i \cdot \mathbf{c}_i$ et t_p est la densité à l'équilibre pour $\mathbf{u} = 0$.

23.6 FEEC

Le Finite element exterior calculus, introduit par Douglas N. Arnold, Richard S. Falk et Ragnar Winther en 2006, est une approche visant à expliquer (et à développer) des solutions éléments finis pour une grande variété d'équations aux dérivées partielles. Il s'agit de mettre à profit les outils de la géométrie différentielle, de la topologie algébrique et de l'algèbre homologique afin de développer des discrétisations compatibles avec les structures géométriques, topologiques et algébriques nécessaires pour que le problème aux équations aux dérivées partielles considéré soit bien posé.

Dans cette version de ce document, nous n'entrons pas plus avant dans cette méthode, qui reste sans doute trop mathématique dans sa présentation pour le public visé. Néanmoins, les articles, notamment ceux de 2006 [7] et de 2010 [8], bien que longs, sont extrêmement pédagogiques et leur lecture ne peut qu'être un plus.

23.7 Systèmes multi-corps

Le concept de système multicorps est utilisé en mécanique du solide, plus particulièrement dans les domaines de la robotique, de l'automobile, de la biomécanique... pour modéliser le comportement dynamique de corps rigides et/ou flexibles connectés les uns aux autres par des liaisons mécaniques, chacun de ces corps décrivant de grands déplacements à la fois en translation et en rotation. Une analyse peut inclure plusieurs milliers de corps rigides. Dans une telle approche, ce n'est plus

le comportement local qui est visé mais plutôt le comportement de plusieurs corps formant un « mécanisme ».

Un corps représente donc une partie rigide ou flexible d'un système mécanique. Un lien désigne une connexion entre au moins deux corps ou entre un corps et le sol : on retrouve donc les liaisons mécaniques classiques comme l'appui ponctuel, la rotule, le glissière, le cardan...

Dans cette approche, le terme de degré de liberté désigne le nombre de mouvements cinématiques possibles, autrement dit le nombre de rotations ou déplacements qu'il reste à fixer pour définir complètement la position dans l'espace. De manière complémentaire, une condition de contrainte désigne une restriction des libertés de mouvement du corps. Ce terme désigne également les contraintes pouvant porter sur les vitesses ou les accélérations de ces mouvements. Enfin, pour parfaire l'analyse, des contraintes supplémentaires peuvent être introduites comme des contraintes de glissement et de contact, entre autres. La dynamique d'un système multicorps est décrite par les

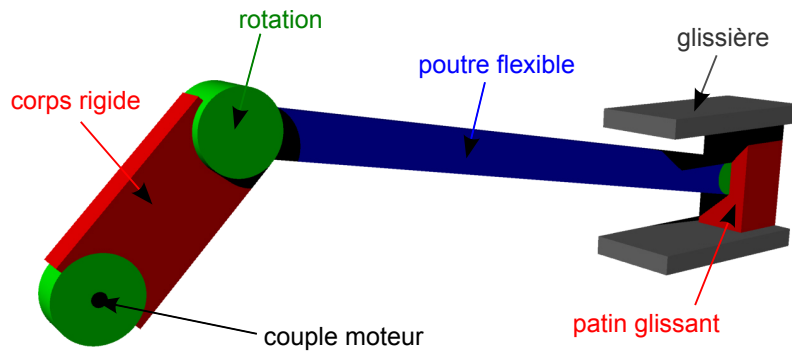


FIGURE 23.1: Exemple de système multi-corps

équations du mouvement. On obtient alors un système de la forme :

$$\begin{cases} M\ddot{u} - Q_{\dot{u}} + C_q\lambda = F \\ C(u, \dot{u}) = 0 \end{cases} \quad (23.10)$$

où M est la matrice de masse, C , la matrice des conditions de contraintes, C_u la matrice jacobienne (dérivée de C par rapport à u) permettant d'appliquer les forces λ correspondant à des multiplicateurs de Lagrange et $Q_{\dot{u}}$, le vecteur de vitesse quadratique utilisé pour introduire les termes de Coriolis et les termes centrifuges.

Nous n'entrons pas plus en avant, car le système à résoudre doit sembler suffisamment simple pour le lecteur à ce niveau du document et parce que cette méthode est d'application très particulière.

Chapitre 24

Quelques mots sur les singularités

Résumé — Les modélisations basées sur la mécanique des milieux continus conduisent, dans un certain nombre de cas particuliers, à des contraintes « infinies » en certains points : les singularités. Ces valeurs infinies sortent du domaine de validité de la plupart des modélisations et, dans le cadre des simulations par éléments finis, pourraient mener un concepteur peu averti à des erreurs d'analyse.

24.1 Qu'est-ce qu'une singularité ?

Comme nous l'avons vu au cours de ce document, la mise en œuvre de la mécanique des milieux continus conduit à la construction d'un problème mathématique dont la solution est constituée d'un champ des déplacements et d'un champ des contraintes. Concernant le problème continu (i.e. non discrétisé, i.e. tel que présenté à la partie II), ces champs sont généralement des fonctions spatiales relativement régulières.

Cependant, dans certains cas, il existe des points où la solution n'est pas entièrement définie : ces points sont nommés singularités. D'une manière très pragmatique, la contrainte et la déformation tendent vers l'infini lorsque l'on s'approche du point singulier. Le déplacement, quant à lui, garde généralement une valeur finie. Nous avons par exemple abordé cela au chapitre 20 : la contrainte tend vers l'infini au voisinage de la pointe d'une fissure.

Il est important de noter que les singularités ne proviennent ni d'erreurs de calcul, ni d'erreurs dans l'application de la théorie, ni de modèles physiques spécialisés : elles proviennent de la mécanique des milieux continus (ou, plus généralement, de toute autre théorie physique basée sur la notion de milieu continu), et leur existence est prédite par l'étude mathématique de ces théories.

24.2 Singularités et éléments finis

En pratique, la plupart des simulations de mécanique des milieux continus sont basées sur la méthode des éléments finis. Or, celle-ci, comme toute méthode numérique, a la fâcheuse et dangereuse habitude de toujours retourner des valeurs finies, ce qui masque par conséquent la présence éventuelle de singularités.

En effet, un solveur éléments finis ne calcule les contraintes et déformations qu'aux points d'intégration (ou points de Gauss) des éléments, qui sont situés à l'intérieur des éléments. Or, dans une simulation par éléments finis, les points singuliers sont toujours des nœuds du maillage, et sont donc situés au bord des éléments. Les contraintes ne sont donc jamais calculées aux points singuliers, et ne présentent pas de valeurs infinies qui permettraient de détecter la singularité. Ce que l'on observe ressemble plutôt à une simple concentration de contraintes et les valeurs obtenues n'ont souvent rien de choquant à première vue.

C'est pourquoi il est indispensable de savoir quand ces singularités doivent se produire afin d'être en mesure d'effectuer les corrections nécessaires et ainsi redonner du sens à l'analyse effectuée.

24.3 Quand les singularités se produisent-elles ?

Si les singularités avaient la gentillesse de ne se produire que dans certains cas particuliers bien spécifiques, elles ne constitueraient alors pas vraiment un problème. Malheureusement, c'est loin d'être le cas, et de nombreux modèles courants de lois de comportements ou de conditions aux limites conduisent à des singularités. Même en restant dans le cadre de l'élasticité linéaire, on trouve des cas fréquents conduisant à des singularités, parmi lesquels les plus fréquemment rencontrés sont :

- les modèles comportant un angle rentrant (i.e. inférieur à 180° entre deux faces extérieures). Une fissure peut d'ailleurs être considérée comme un angle rentrant d'angle nul ;
- les modèles de lois comportements discontinues, comme à l'interface entre deux matériaux (dont nous avons souvent parlé tout au long de ce document) ;
- les modèles de chargements contenant des efforts ponctuels, qui est de loin le cas le plus fréquemment rencontré.

En tout honnêteté, il est difficile de nier le fait que ces trois cas sont vraiment couramment employés, ce qui permet de prendre toute la mesure du problème : on observe régulièrement des dimensionnements réalisés à partir de contraintes calculées au fond d'un angle rentrant ou sous une force ponctuelle, alors que les valeurs de ces contraintes n'y sont absolument pas fiables...

Cette liste n'est naturellement pas exhaustive et il existe d'autres cas pouvant entraîner des singularités, comme la présence d'encastremements ou de déplacements imposés dans certaines configurations géométriques particulières. Inversement, et contrairement à ce que d'aucuns croient, les théories des poutres, plaques et coques présentent généralement moins de cas singuliers que la mécanique des milieux continus tridimensionnels (car certains de ces aspects peuvent être pris en compte dans la construction du modèle).

24.4 Comment éviter les singularités

Nous avons vu que les singularités proviennent de limitations intrinsèques de la mécanique des milieux continus : cette dernière donne des résultats non valides en présence d'un certain nombre de configurations. Cela signifie que ces configurations n'appartiennent pas au domaine de validité de la mécanique des milieux continus tridimensionnelle, et que leur emploi peut donc mener à des résultats non pertinents (en l'occurrence, singuliers).

Schématiquement, la singularité provient du fait que la mécanique des milieux continus postule l'existence d'une densité volumique d'énergie, et s'accommode donc mal du caractère « ponctuel » de ces modèles (angle ponctuel, force ponctuelle, interface d'épaisseur nulle) qui conduit à des densités d'énergie infinies. Pour éviter la singularité, il faut donc utiliser des modèles non ponctuels comme :

- remplacer un angle rentrant par un congé de raccordement possédant un rayon de courbure non nul ;
- remplacer une discontinuité entre lois de comportement par une zone de transition dans laquelle les paramètres varient de façon continue ;
- remplacer une force ponctuelle par une pression de contact appliquée sur une surface non nulle...

Ces configurations ne créent pas de singularités, mais de simples concentrations de contraintes : les contraintes et les déformations restent finies dans leur voisinage. Dans les faits, leur usage est

indispensable à chaque fois que l'objectif de la simulation est de calculer une contrainte ou une déformation localisée dans la zone incriminée.

24.5 Singularités et pertinence d'un résultat

Le problème des configurations suggérées pour éviter les singularités est qu'elles sont plus riches, qu'elles demandent plus d'informations que les modélisations ponctuelles. Or, le concepteur ne dispose pas toujours de ces informations, notamment aux premiers stades de la conception d'un produit, et peut donc être tenté d'utiliser un modèle plus simple, quitte à violer le domaine de validité de la mécanique des milieux continus.

En réalité, il peut même être tout à fait légitime d'utiliser un modèle « non valide » (entraînant des singularités), à condition d'avoir la certitude que ces singularités perturberont peu le résultat que l'on cherche à calculer.

C'est typiquement le cas lorsque le résultat est une quantité située suffisamment loin de la zone singulière : les singularités sont des anomalies très localisées, et leur effet direct décroît rapidement avec la distance. La singularité n'influe alors sur le résultat que par le biais des redistributions de contraintes, et cette influence est souvent (mais pas toujours !) négligeable.

En tout état de cause, il appartient au concepteur d'évaluer le caractère gênant ou non d'une singularité à l'aide de son expérience et de son esprit critique.

24.6 Conclusion

Nous espérons avoir pu mettre en évidence les points suivants :

- En mécanique des milieux continus, de nombreuses modélisations courantes mènent à des contraintes infinies en un ou plusieurs points : angles rentrants dans les modèles géométriques, discontinuités dans les modèles de comportements des matériaux, efforts ponctuels dans les modèles de chargement...
- Ces contraintes infinies sont prédites par les mathématiques, mais sortent du domaine de validité de la mécanique des milieux continus.
- Dans les simulations par éléments finis, les contraintes restent finies au voisinage des singularités, mais leur valeur n'est pas pertinente pour autant : elle dépend uniquement de la taille et de la forme des éléments et augmente indéfiniment lorsque l'on raffine le maillage.
- Un concepteur qui ignore l'existence de ces singularités risque donc de dimensionner une pièce par rapport à un résultat non fiable, sauf s'il prend la peine de raffiner successivement le maillage, ce qui permet de diagnostiquer le problème (vous savez, la fameuse étude de convergence que l'on fait toujours tant que l'on est élève ingénieur et que l'on a tendance à oublier, ou à négliger par la suite).
- Si l'on souhaite simuler l'état de contraintes au voisinage de la région singulière, il est nécessaire de modéliser celle-ci plus finement pour faire disparaître la singularité. Cela nécessite généralement des connaissances supplémentaires sur le produit, son environnement ou le comportement de ses matériaux dans la région concernée.
- Si l'état de contraintes ou de déformations au voisinage de la singularité ne fait pas partie des objectifs du calcul, alors celle-ci n'est pas gênante. Mieux vaut néanmoins être conscient du problème pour ne pas en tirer de fausses conclusions...

Nous réitérons ce que nous avons déjà dit : la plupart des ingénieurs pratiquant le calcul sont conscients que ces problèmes de singularités existent. Nous n'avons donc fait ici que rappeler des choses connues... et c'est très bien ainsi si c'est le cas.

Conclusion

En guise de conclusion à ce document, nous souhaitons synthétiser les problèmes qui peuvent survenir lors d'un calcul et ouvrir sur quelques perspectives.

Sur la fiabilité des résultats

La fiabilité d'un résultat dépend de celle de toute la chaîne d'approximation réalisées depuis la modélisation du phénomène physique jusqu'à l'obtention des résultats numériques fournis par le programme de calcul. Nombre de ces problèmes ont été mentionnés au cours de ce document. Rappelons les ici :

erreurs de modélisation : aussi bien au niveau du choix des équations mathématiques décrivant le phénomène, que de la représentativité des conditions aux limites choisies. Mais au-delà même, il est nécessaire de se rappeler que les modèles mathématiques ne fonctionnent bien qu'avec les plus simples des situations, ou avec des situations complexes ne dépendant que quelques facteurs dominants simples. La plupart des modèles sont des modèles linéaires, mais mêmes ceux-ci peuvent dissimuler de la complexité : le chaos linéaire existe et les simulations numériques d'un tel système peuvent conduire à des résultats qui peuvent surprendre si l'on n'en est pas averti (voir la linéarisation du pendule double, ou la suite logistique). En d'autres termes, et pour bien insister, la réalité physique que l'on se propose de modéliser, repose sur des propriétés qui ne sont pas connues exactement, dont les lois de probabilité ne sont pas forcément connues non plus, où tous les phénomènes sont non-linéaires et où ils sont tous couplés. Les mathématiques proposent des modèles simplifiés correspondant aux situations les plus simples (voire même n'ayant pas d'existence réelle), et il est du ressort de l'ingénieur, et c'est en cela que son travail est plus compliqué que celui du mathématicien, de savoir juger de la pertinence de tel ou tel modèle retenu pour modéliser la problématique considérée ;

erreurs de discrétisation : elles sont liées aux choix des méthodes numériques (méthode des éléments finis ou autre), aux problèmes d'intégration, de représentation du domaine...

erreurs de verrouillage numérique : elles concernent des problèmes survenant lors du traitement de paramètres introduits dans le calcul tels que les pas de temps, des modes parasites, des instabilités numériques...

incertitudes des données : connaissance approximative des lois de comportement, des efforts, des liaisons... Il faut alors procéder à une approche fiabiliste ;

erreurs d'arrondis : la manière dont un ordinateur traite un nombre est soumise à des contraintes (représentation en base 2 par exemple, approximation à n décimales...). Ces erreurs, mêmes infimes, peuvent dans certains cas se cumuler pour aboutir à un résultat faux.

Lorsque cela est possible, on n'hésitera donc pas à effectuer des comparaisons avec des essais, ce que l'on nomme recalage calcul-essais.

Quelques perspectives

Le calcul scientifique d'une manière générale, et la méthode des éléments finis en particulier, sont utilisés de manière intensive dans tous les secteurs. Toute amélioration de la performance des méthodes numériques est donc un enjeu majeur : rapidité, précision, fiabilité...

On pourra citer le développement de codes de calculs parallèles, ou l'amélioration des techniques d'optimisation par les algorithmes génétiques. Nous avons parlé des modèles multi-échelle essentiellement pour les matériaux, mais le développement de modèles pour les structures minces (plaques, coques) rentre dans cette catégorie... et si les théories de plaques sont aujourd'hui assez bien maîtrisées (d'un point de vue théorique et dans les codes), il reste des problèmes ouverts concernant les coques minces.

Les problèmes inverses constituent également un défi d'intérêt. Ils permettent par exemple de remonter aux caractéristiques d'un matériau (qui seront utilisées dans d'autres calculs), à partir d'essais sur un échantillon. C'est également souvent la seule voie d'identification des paramètres et comportement pour tout ce qui concerne la biomécanique et la modélisation du corps humain en général, car il n'est pas possible d'effectuer des mesures réelles.

Enfin, et peut-être surtout, les modèles développés aujourd'hui se veulent de plus en plus réalistes. Il devient alors indispensable de coupler des modèles numériques différents, ce qui n'est possible que si, au préalable, on a su établir des passerelles, des lieux de travail commun, entre des spécialistes de plusieurs disciplines.

Annexe A

Interpolation et approximation

L'interpolation est une opération consistant à approcher une courbe qui n'est connue que par la donnée d'un nombre fini de points (ou une fonction à partir de la donnée d'un nombre fini de valeurs).

Ainsi, l'interpolation numérique sert souvent à « faire émerger une courbe parmi des points ». Il s'agit de toutes les méthodes développées afin de mieux prendre en compte les erreurs de mesure, i.e. d'exploiter des données expérimentales pour la recherche de lois empiriques. Nous citerons par exemple la régression linéaire et la méthode des moindres carrés souvent bien maîtrisées. On demande à la solution du problème d'interpolation de passer par les points prescrits, voire, suivant le type d'interpolation, de vérifier des propriétés supplémentaires (de continuité, de dérivabilité, de tangence en certains points...). Toutefois, parfois on ne demande pas à ce que l'approximation passe exactement par les points prescrits. On parle alors plutôt d'approximation.

Histoire

Le jour du Nouvel An de 1801, l'astronome italien Giuseppe Piazzi a découvert l'astéroïde Cérès (il a suivi sa trajectoire jusqu'au 14 février 1801).

Durant cette année, plusieurs scientifiques ont tenté de prédire sa trajectoire sur la base des observations de Piazzi (à cette époque, la résolution des équations non linéaires de Kepler de la cinématique est un problème très difficile). La plupart des prédictions furent erronées ; et le seul calcul suffisamment précis pour permettre au baron Franz Xaver von Zach (astronome allemand) de localiser à nouveau Cérès à la fin de l'année, fut celui de Gauß, (alors âgé de 24 ans). Gauß avait déjà réalisé l'élaboration des concepts fondamentaux en 1795, lorsqu'il avait 18 ans. Mais sa méthode des moindres carrés ne fut publiée qu'en 1809, lorsqu'elle parut dans le tome 2 de ses travaux sur la Mécanique céleste *Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium*. Le mathématicien français Adrien-Marie Legendre a développé indépendamment la même méthode en 1805. Le mathématicien américain Robert Adrain a publié en 1808 une formulation de la méthode.



Piazzi



Gauß



Legendre



Adrain



Markov

En 1829, Gauß a pu donner les raisons de l'efficacité de cette méthode : celle-ci est optimale à l'égard de bien des critères. Cet argument est maintenant connu sous le nom de théorème de Gauß-Markov.

Ce théorème dit que dans un modèle linéaire dans lequel les erreurs ont une espérance nulle, sont non corrélées et dont les variances sont égales, le meilleur estimateur linéaire non biaisé des coefficients est l'estimateur des moindres carrés. Plus généralement, le meilleur estimateur linéaire non biaisé d'une combinaison linéaire des coefficients est son estimateur par les moindres carrés. *On ne suppose pas* que les erreurs possèdent une loi normale, ni qu'elles sont indépendantes (seulement non corrélées), ni qu'elles possèdent la même loi de probabilité.

L'approximation étant l'utilisation de méthodes permettant d'approcher une fonction mathématique par une suite de fonctions qui convergent dans un certain espace fonctionnel, on voit donc que

ce qui a été fait dans la deuxième partie ressort bien de cela : on cherche une fonction généralement notée u qui n'est pas connue explicitement mais solution d'une équation différentielle ou d'une équation aux dérivées partielles, et l'on cherche à construire une suite de problèmes plus simples, que l'on sait résoudre à chaque étape, et telle que la suite des solutions correspondantes converge vers la solution cherchée. L'approximation peut servir aussi dans le cas où la fonction considérée est connue : on cherche alors à la remplacer par une fonction plus simple, plus régulière ou ayant de meilleures propriétés. L'intégration numérique sera détaillée un peu plus au chapitre B. Ce sont les méthodes d'approximation numériques qui sont utilisées.

Pour en revenir à l'interpolation, la méthode des éléments finis est en elle-même une méthode d'interpolation (globale, basée sur des interpolations locales). On peut citer quelques méthodes d'interpolation telle que l'interpolation linéaire (dans laquelle deux points successifs sont reliés par un segment), l'interpolation cosinus (dans laquelle deux points successifs sont considérés comme les pics d'un cosinus. L'interpolation cubique ou spline (dans laquelle un polynôme de degré 3 passe par quatre points successifs : selon le type de continuité demandée plusieurs variantes existent) et de manière générale, l'interpolation polynomiale, est abordée ci-dessous.

Faisons d'emblée une mise en garde : la plus connue des interpolations polynomiale, l'interpolation lagrangienne (approximation par les polynômes de Lagrange, découverte initialement par Waring et redécouverte par Euler) peut fort bien diverger même pour des fonctions très régulières. C'est le phénomène de Runge : contrairement à l'intuition, l'augmentation du nombre de points d'interpolation ne constitue pas nécessairement une bonne stratégie d'approximation avec certaines fonctions (même infiniment dérivables).

Dans le cas où l'on travaille sur le corps des complexes, une méthode d'approximation d'une fonction analytique par une fonction rationnelle est l'approximant de Padé. Cela correspond à un développement limité qui approche la fonction par un polynôme. Tout comme les développements limités forment une suite appelée série entière, convergeant vers la fonction initiale, les approximants de Padé sont souvent vus comme une suite, s'exprimant sous la forme d'une fraction continue dont la limite est aussi la fonction initiale. En ce sens, ces approximants font partie de la vaste théorie des fractions continues. Les approximants offrent un développement dont le domaine de convergence est parfois plus large que celui d'une série entière. Ils permettent ainsi de prolonger des fonctions analytiques et d'étudier certains aspects de la question des séries divergentes. En théorie analytique des nombres, l'approximant permet de mettre en évidence la nature d'un nombre ou d'une fonction arithmétique comme celle de la fonction zêta de Riemann. Dans le domaine du calcul numérique, l'approximant joue un rôle, par exemple, pour évaluer le comportement d'une solution d'un système dynamique à l'aide de la théorie des perturbations. L'approximant de Padé a été utilisé pour la première fois par Euler pour démontrer l'irrationalité de e , la base du logarithme népérien. Une technique analogue a permis à Johann Heinrich Lambert de montrer celle de π .

A.1 Quelques bases polynomiales

A.1.1 Motivation

Histoire

Lorsque l'on souhaite approximer une courbe par une autre, recourir aux polynômes semble une voie naturelle.

Les séries de fonctions sont apparues à la fin du XVII^e siècle, lorsque Isaac Newton puis Brook Taylor décomposèrent des fonctions en séries (entières, donc sur \mathbb{C}) pour le calcul d'intégrales. Ce n'est qu'en 1821 que Cauchy fournira les critères précis de convergence dans ses notes de cours de Polytechnique *Analyse algébrique*, différents autres types de convergence étant introduits jusqu'en 1840.

La richesse de cette théorie vient de ce qu'elle permet d'étudier des fonctions qui ne s'expriment pas à l'aide de fonctions connues, comme le sont certaines solutions d'équations différentielles.

Le théorème de Taylor (1715) (théorème 66) montre qu'une fonction plusieurs fois dérivable au voisinage d'un point peut être approximée par une fonction polynôme dont les coefficients dépendent uniquement des dérivées de la fonction en ce point. En présentant cette formule, Taylor propose une méthode de développement en série, mais il se préoccupe peu de la nature du reste ; il faut attendre ses successeurs pour la caractériser rigoureusement. On désigne par théorèmes de Taylor ou formules de Taylor plusieurs résultats et expressions pour le reste $R_n(x)$, parfois renforcé par quelques hypothèses supplémentaires : Taylor-Young, Taylor-Lagrange, Taylor-Cauchy, Taylor avec reste intégral.



Taylor

Remarque. L'usage de chacune des formules de Taylor dépend du but recherché.

La formule de Taylor-Young précise seulement le comportement d'une expression lorsque h tend vers 0 ; elle est donc purement locale au voisinage du point a considéré. À l'ordre 1 (i.e. $k = 1$) elle se réduit à la définition de la différentielle. Pour $k = 2$ elle montre que le terme en $o(\|h\|)$ est en fait en $O(\|h\|^2)$ dès que la fonction est deux fois différentiable en a .

La formule avec reste intégral est la plus précise. On l'utilisera chaque fois que l'on a besoin d'informations détaillées sur le reste, notamment sur sa dépendance par rapport aux variables ou à d'éventuels paramètres.

La majoration de Taylor-Lagrange sera utilisée pour évaluer l'accroissement entre deux points, non nécessairement proches l'un de l'autre.

Théorème 66 — Développement en série de Taylor. Soit I un intervalle de \mathbb{R} , $a \in I$, E un espace vectoriel normé et $f : I \rightarrow E$ une fonction dérivable en a jusqu'à l'ordre n . Alors :

$$\forall x \in I, \quad f(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k + R_n(x) \quad (\text{A.1})$$

définit un reste $R_n(x)$ dont le comportement s'apparente au monôme $(x-a)^{n+1}$.

Le théorème d'approximation de Weierstrass en analyse réelle dit que toute fonction continue définie sur un segment peut être approchée uniformément par des fonctions polynômes. Le théorème de Stone-Weierstrass généralise ce résultat aux fonctions continues définies sur un espace compact et à valeurs réelles, en remplaçant l'algèbre des polynômes par une algèbre de fonctions qui sépare les points et contient au moins une fonction constante non nulle.

L'interpolation polynomiale consiste donc à trouver un polynôme passant par un ensemble de points donnés. Nous verrons qu'il est également possible de demander à ce que ce polynôme satisfasse à d'autres conditions.

A.1.2 Orthogonalité

Une suite de polynômes orthogonaux est une suite infinie de polynômes $P_0(x), P_1(x), \dots$ à coefficients réels, dans laquelle chaque $P_n(x)$ est de degré n , et telle que les polynômes de la suite

sont orthogonaux deux à deux pour un produit scalaire de fonctions donné. Le produit scalaire de fonctions le plus simple est l'intégrale du produit de ces fonctions sur un intervalle borné :

$$\langle f, g \rangle = \int_a^b f(x)g(x)dx \quad (\text{A.2})$$

Plus généralement, on peut ajouter une fonction de poids $\varpi(x)$ dans l'intégrale. On notera bien que sur l'intervalle d'intégration $]a, b[$, la fonction poids $\varpi(x)$ doit être à valeurs finies et strictement positives, et l'intégrale du produit de la fonction poids par un polynôme doit être finie (voir espaces L^p). Par contre, les bornes a et b peuvent être infinies. Il vient alors :

$$\langle f, g \rangle = \int_a^b f(x)g(x)\varpi(x)dx \quad (\text{A.3})$$

La norme associée est définie par $\|f\| = \sqrt{\langle f, f \rangle}$. Le produit scalaire fait de l'ensemble de toutes les fonctions de norme finie un espace de Hilbert. L'intervalle d'intégration est appelé intervalle d'orthogonalité.

A.1.3 Base naturelle

On rappelle que $1, X, \dots, X^n$ est une base de $\mathbb{K}_n[X]$ en tant que polynômes échelonnés.

A.1.4 Polynômes de Lagrange

Connaissant $n + 1$ points $(x_0, y_0), \dots, (x_n, y_n)$ d'abscisses distinctes, le polynôme de Lagrange est l'unique polynôme de degré n passant tous les points. Ce polynôme est trivialement défini par :

$$L(x) = \sum_{j=0}^n y_j \left(\prod_{i=0, i \neq j}^n \frac{x - x_i}{x_j - x_i} \right) \quad (\text{A.4})$$

Si l'on note :

$$L(x) = \sum_{j=0}^n y_j l_j(x) \quad (\text{A.5})$$

avec :

$$l_i(x) = \prod_{j=0, j \neq i}^n \frac{x - x_j}{x_i - x_j} = \frac{x - x_0}{x_i - x_0} \dots \frac{x - x_{i-1}}{x_i - x_{i-1}} \frac{x - x_{i+1}}{x_i - x_{i+1}} \dots \frac{x - x_n}{x_i - x_n} \quad (\text{A.6})$$

alors, on remarque que :

- l_i est de degré n pour tout i ;
- $l_i(x_j) = \delta_{ij}$, $0 \leq i, j \leq n$, i.e. $l_i(x_i) = 1$ et $l_i(x_j) = 0$ pour $j \neq i$

On en déduit immédiatement que $\forall i, L(x_i) = y_i$ qui est bien la propriété recherchée par construction.

A.1.5 Polynômes d'Hermite

Les polynômes d'Hermite sont définis sous la forme dite *probabiliste* :

$$H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2} \quad (\text{A.7})$$

ou sous la forme dite *physique* :

$$\hat{H}_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} \quad (\text{A.8})$$

Les deux définitions sont liées par la propriété d'échelle suivante :

$$\hat{H}_n(x) = 2^{n/2} H_n(x\sqrt{2}) \quad (\text{A.9})$$

Les premiers polynômes d'Hermite sont les suivants :

$H_0 = 1$	$\hat{H}_0 = 1$	(A.10)
$H_1 = x$	$\hat{H}_1 = 2x$	
$H_2 = x^2 - 1$	$\hat{H}_2 = 4x^2 - 2$	
$H_3 = x^3 - 3x$	$\hat{H}_3 = 8x^3 - 12x$	
$H_4 = x^4 - 6x^2 + 3$	$\hat{H}_4 = 16x^4 - 48x^2 + 12$	
$H_5 = x^5 - 10x^3 + 15x$	$\hat{H}_5 = 32x^5 - 160x^3 + 120x$	
$H_6 = x^6 - 15x^4 + 45x^2 - 15$	$\hat{H}_6 = 64x^6 - 480x^4 + 720x^2 - 120$	

H_n est un polynôme de degré n . Ces polynômes sont orthogonaux pour la mesure μ de densité :

$$\frac{d\mu(x)}{dx} = \frac{e^{-x^2/2}}{\sqrt{2\pi}} = \varphi(x) \quad (\text{A.11})$$

avec $\varphi(x)$ la densité de probabilité d'une variable aléatoire gaussienne centrée réduite (moyenne nulle, écart-type égal à 1). Ils vérifient :

$$\int_{-\infty}^{+\infty} H_n(x) H_m(x) \varphi(x) dx = n! \delta_{nm} \quad (\text{A.12})$$

où δ_{nm} est le symbole de Kronecker. C'est pour ces intéressantes propriétés que cette base a été choisie en mécanique stochastique... on utilise alors la jolie expression de base de chaos polynomial. Pour une application à la prise en compte de l'aléa dans la méthode des éléments finis, voir le chapitre 22. Ces polynômes forment une base orthogonale de l'espace de Hilbert $L^2(\mathbb{C}, \mu)$ des fonctions boréliennes telles que :

$$\int_{-\infty}^{+\infty} |f(x)|^2 \varphi(x) dx < +\infty \quad (\text{A.13})$$

dans lequel le produit scalaire est donné par l'intégrale :

$$\langle f, g \rangle = \int_{-\infty}^{+\infty} f(x) \overline{g(x)} \varphi(x) dx \quad (\text{A.14})$$

Des propriétés analogues sont vérifiées par les polynômes de Hermite sous leur forme physique.

A.1.6 Polynômes de Legendre

On appelle équation de Legendre l'équation :

$$\frac{d}{dx} \left((1-x^2) \frac{dy}{dx} \right) + n(n+1)y = 0 \quad (\text{A.15})$$

On définit le polynôme de Legendre P_n par :

$$\frac{d}{dx} \left((1-x^2) \frac{dP_n(x)}{dx} \right) + n(n+1)P_n(x) = 0, \quad P_n(1) = 1 \quad (\text{A.16})$$

La manière la plus simple de les définir est par la formule de récurrence de Bonnet : $P_0(x) = 1$, $P_1(x) = x$ et :

$$\forall n > 0, \quad (n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x) \quad (\text{A.17})$$

Les premiers polynômes de Legendre sont :

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ P_2(x) &= \frac{1}{2}(3x^2 - 1) \\ P_3(x) &= \frac{1}{2}(5x^3 - 3x) \\ P_4(x) &= \frac{1}{8}(35x^4 - 30x^2 + 3) \\ P_5(x) &= \frac{1}{8}(63x^5 - 70x^3 + 15x) \end{aligned} \quad (\text{A.18})$$

Le polynôme P_n est de degré n . La famille $(P_n)_{n \leq \mathbb{N}}$ est une famille de polynômes à degrés étagés, elle est donc une base de l'espace vectoriel $\mathbb{R}_n[X]$. On remarquera la propriété suivante :

$$P_n(-x) = (-1)^n P_n(x) \quad (\text{A.19})$$

qui donne en particulier $P_n(-1) = (-1)^n$ et $P_{2n+1}(0) = 0$. Les polynômes orthogonaux les plus simples sont les polynômes de Legendre pour lesquels l'intervalle d'orthogonalité est $[-1; 1]$ et la fonction poids est simplement la fonction constante de valeur 1 : ces polynômes sont orthogonaux par rapport au produit scalaire défini sur $\mathbb{R}[X]$ par :

$$\langle P, Q \rangle = \int_{-1}^{+1} P(x)Q(x)dx \quad \langle P_m, P_n \rangle = \int_{-1}^{+1} P_m(x)P_n(x)dx = 0 \quad \text{pour} \quad m \neq n \quad (\text{A.20})$$

De plus, comme $(P_n)_{n \leq \mathbb{N}}$ est une base de $\mathbb{R}_N[X]$, on a $P_{N+1} \in (\mathbb{R}_N[X])^\perp$:

$$\forall Q \in \mathbb{R}_N[X], \quad \int_{-1}^{+1} P_{N+1}(x)Q(x)dx = 0 \quad (\text{A.21})$$

Le carré de la norme, dans $L^2([-1; 1])$, est :

$$\|P_n\|^2 = \frac{2}{2n+1}. \quad (\text{A.22})$$

Ces polynômes peuvent servir à décomposer une fonction holomorphe, une fonction lipschitzienne ou à retrouver l'intégration numérique d'une fonction par la méthode de quadrature de Gauss-Legendre (voir chapitre B).

A.1.7 Polynômes de Tchebychev

Les polynômes de Tchebychev servent pour la convergence des interpolations de Lagrange. Ils sont également utilisés dans le calcul de filtres de Tchebychev en électronique analogique.

Les polynômes de Tchebychev constituent deux familles de polynômes (notés T_n pour la première espèce et U_n pour la seconde) définis sur l'intervalle $[-1 ; 1]$ par les relations trigonométriques :

$$T_n(\cos(\theta)) = \cos(n\theta) \quad (\text{A.23})$$

$$U_n(\cos(\theta)) = \frac{\sin((n+1)\theta)}{\sin \theta} \quad (\text{A.24})$$

Ces deux suites sont définies par la relation de récurrence :

$$\forall n \in \mathbb{N}, \quad P_{n+2}(X) = 2X P_{n+1}(X) - P_n(X) \quad (\text{A.25})$$

et les deux premiers termes :

$$T_0 = 1, \quad T_1 = X \quad \text{pour la suite } T \quad (\text{A.26})$$

$$U_0 = 1, \quad U_1 = 2X \quad \text{pour la suite } U \quad (\text{A.27})$$

Chacune de ces deux familles est une suite de polynômes orthogonaux par rapport à un produit scalaire de fonctions assorti d'une pondération spécifique.

Propriétés des polynômes de Tchebychev de première espèce

$$\forall n > 0, \quad T_n(x) = \frac{n}{2} \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^k \frac{(n-k-1)!}{k!(n-2k)!} (2x)^{n-2k} \quad (\text{A.28})$$

Les polynômes T_n sont orthogonaux pour le produit scalaire suivant :

$$\int_{-1}^1 \frac{T_n(x)T_m(x)}{\sqrt{1-x^2}} dx = \begin{cases} 0 & \text{si } n \neq m \\ \pi & \text{si } n = m = 0 \\ \pi/2 & \text{si } n = m \neq 0. \end{cases} \quad (\text{A.29})$$

$$\forall n, \quad T_n(1) = 1, \quad \forall n, m \in \mathbb{N}, \quad \forall x \in \mathbb{R}, \quad T_n(T_m(x)) = T_{mn}(x) \quad (\text{A.30})$$

Les premiers polynômes de Tchebychev de première espèce sont :

$$\begin{aligned} T_0 &= 1 \\ T_1 &= x \\ T_2 &= 2x^2 - 1 \\ T_3 &= 4x^3 - 3x \\ T_4 &= 8x^4 - 8x^2 + 1 \\ T_5 &= 16x^5 - 20x^3 + 5x \\ T_6 &= 32x^6 - 48x^4 + 18x^2 - 1 \\ T_7 &= 64x^7 - 112x^5 + 56x^3 - 7x \\ T_8 &= 128x^8 - 256x^6 + 160x^4 - 32x^2 + 1 \\ T_9 &= 256x^9 - 576x^7 + 432x^5 - 120x^3 + 9x \end{aligned} \quad (\text{A.31})$$

Propriétés des polynômes de Tchebychev de seconde espèce

$$\forall n \geq 0, \quad U_n(x) = \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^k \binom{n-k}{k} (2x)^{n-2k} \quad (\text{A.32})$$

Les polynômes U_n sont orthogonaux pour le produit scalaire suivant :

$$\int_{-1}^1 U_n(x) U_m(x) \sqrt{1-x^2} dx = \begin{cases} 0 & \text{si } n \neq m \\ \pi/2 & \text{si } n = m \end{cases} \quad \text{et} \quad \forall n, \quad U_n(1) = n+1 \quad (\text{A.33})$$

Les premiers polynômes de Tchebychev de deuxième espèce sont :

$$\begin{aligned} U_0 &= 1 \\ U_1 &= 2x \\ U_2 &= 4x^2 - 1 \\ U_3 &= 8x^3 - 4x \\ U_4 &= 16x^4 - 12x^2 + 1 \\ U_5 &= 32x^5 - 32x^3 + 6x \\ U_6 &= 64x^6 - 80x^4 + 24x^2 - 1 \\ U_7 &= 128x^7 - 192x^5 + 80x^3 - 8x \\ U_8 &= 256x^8 - 448x^6 + 240x^4 - 40x^2 + 1 \\ U_9 &= 512x^9 - 1024x^7 + 672x^5 - 160x^3 + 10x \end{aligned} \quad (\text{A.34})$$

A.1.8 Polynômes de Laguerre

Les polynômes de Laguerre apparaissent en mécanique quantique dans la partie radiale de la solution de l'équation de Schrödinger pour un atome à un électron. Ces polynômes sont les solutions de l'équation de Laguerre :

$$xy'' + (1-x)y' + ny = 0 \quad (\text{A.35})$$

qui est une équation différentielle linéaire du second ordre possédant des solutions non singulières si et seulement si n est un entier positif.

Traditionnellement notés L_0, L_1, \dots ces polynômes forment une suite de polynômes qui peut être définie par la formule de Rodrigues :

$$L_n(x) = \frac{e^x}{n!} \frac{d^n}{dx^n} (e^{-x} x^n). \quad (\text{A.36})$$

Ils sont orthogonaux les uns par rapport aux autres pour le produit scalaire :

$$\langle f, g \rangle = \int_0^\infty f(x) g(x) e^{-x} dx \quad (\text{A.37})$$

Cette propriété d'orthogonalité revient à dire que si X est une variable aléatoire distribuée exponentiellement avec la fonction densité de probabilité suivante :

$$f(x) = \begin{cases} e^{-x} & \text{si } x > 0 \\ 0 & \text{si } x < 0 \end{cases} \quad (\text{A.38})$$

alors $E(L_n(X), L_m(X)) = 0$ si $n \neq m$.

Les premiers polynômes de Laguerre sont :

$$\begin{aligned}
L_0 &= 1 \\
L_1 &= -x + 1 \\
L_2 &= \frac{1}{2}(x^2 - 4x + 2) \\
L_3 &= \frac{1}{6}(-x^3 + 9x^2 - 18x + 6) \\
L_4 &= \frac{1}{24}(x^4 - 16x^3 + 72x^2 - 96x + 24) \\
L_5 &= \frac{1}{120}(-x^5 + 25x^4 - 200x^3 + 600x^2 - 600x + 120) \\
L_6 &= \frac{1}{720}(x^6 - 36x^5 + 450x^4 - 2400x^3 + 5400x^2 - 4320x + 720)
\end{aligned} \tag{A.39}$$

Il existe des polynômes de Laguerre généralisés dont l'orthogonalité peut être liée à une densité de probabilité faisant intervenir la fonction Gamma. Ils apparaissent dans le traitement de l'oscillateur harmonique quantique. Ils peuvent être exprimés en fonction des polynômes d'Hermite.

A.1.9 Polynômes de Bernstein

Les polynômes de Bernstein permettent de donner une démonstration constructive du théorème de Stone-Weierstrass. Dans le cadre de ce cours, nous les présentons surtout car ils sont utilisés dans la formulation générale des courbes de Bézier. Pour un degré n , il y a $n + 1$ polynômes de Bernstein B_0^n, \dots, B_n^n définis, sur l'intervalle $[0; 1]$ par :

$$B_i^n(u) = \binom{n}{i} u^i (1-u)^{n-i} \quad \text{où les } \binom{n}{i} \text{ sont les coefficients binomiaux.} \tag{A.40}$$

Ces polynômes présentent quatre propriétés importantes :

— Partition de l'unité :

$$\sum_{i=0}^n B_i^n(u) = 1, \quad \forall u \in [0; 1] \tag{A.41}$$

— Positivité :

$$B_i^n(u) \geq 0, \quad \forall u \in [0; 1], \quad \forall i \in 0, \dots, n \tag{A.42}$$

— Symétrie :

$$B_i^n(u) = B_{n-i}^n(1-u), \quad \forall u \in [0; 1], \quad \forall i \in 0, \dots, n \tag{A.43}$$

— Formule de récurrence :

$$B_i^n(u) = \begin{cases} (1-u)B_i^{n-1}(u), & i = 0 \\ (1-u)B_i^{n-1}(u) + uB_{i-1}^{n-1}(u), & \forall i \in 1, \dots, n-1, \\ uB_{i-1}^{n-1}(u), & i = n \end{cases} \quad \forall u \in [0; 1] \tag{A.44}$$

On notera la grande ressemblance de ces polynômes avec la loi binomiale.

A.2 Interpolation polynomiale

A.2.1 Interpolation de Lagrange

Dans la version la plus simple (interpolation lagrangienne), on impose simplement que le polynôme passe par tous les points donnés. On obtient les polynômes de Lagrange tels que présentés juste avant. Le théorème de l'unisolvance (voir paragraphe 12.1.1) précise qu'il n'existe qu'un seul polynôme de degré n au plus défini par un ensemble de $n + 1$ points.

L'erreur d'interpolation lors de l'approximation d'une fonction f (donnée par les points $(x_i, y_i = f(x_i))$) par un polynôme de Lagrange p_n est donnée par une formule de type Taylor-Young : Si f est $n + 1$ fois continûment différentiable sur $I = [\min(x_0, \dots, x_n, x), \max(x_0, \dots, x_n, x)]$, alors :

$$f(x) - p_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{i=0}^n (x - x_i) \quad \text{avec} \quad \xi \in I \quad (\text{A.45})$$

Dans le cas particulier où $x_i = x_0 + ih$ (points uniformément répartis), il se produit en général une aggravation catastrophique de l'erreur d'interpolation, connue sous le nom de phénomène de Runge lorsque l'on augmente le nombre de points pour un intervalle $[x_0, x_n]$, donné (on a alors $\xi \in]-1, 1[$).

Pour limiter le phénomène de Runge, i.e. pour minimiser l'oscillation des polynômes interpolateurs, on peut utiliser les abscisses de Tchebychev au lieu de points équirépartis pour interpoler. Dans ce cas, on peut montrer que l'erreur d'interpolation décroît lorsque n augmente. On peut aussi préférer utiliser des splines pour approximer la fonction f (ce sont des polynômes par morceaux définis plus bas). Dans ce cas, pour améliorer l'approximation, on augmente le nombre de morceaux et non le degré des polynômes.

A.2.2 Interpolation par Spline

Une spline est une fonction définie par morceaux par des polynômes. Comme mentionné au dessus, la méthode des splines est souvent préférée à l'interpolation polynomiale, car on obtient des résultats similaires en se servant de polynômes ayant des degrés inférieurs, tout en évitant le phénomène de Runge. De plus, leur simplicité d'implémentation les rend très populaires et elles sont fréquemment utilisées dans les logiciels de dessin.

Une courbe spline est une fonction polynomiale par morceaux définie sur un intervalle $[a; b]$ divisé en sous intervalles $[t_{i-1}; t_i]$ tels que $a = t_0 < t_1 < \dots < t_{k-1} < t_k = b$. On la note $S : [a; b] \rightarrow \mathbb{R}$. Sur chaque intervalle $[t_{i-1}; t_i]$ on définit un polynôme $P_i : [t_{i-1}; t_i] \rightarrow \mathbb{R}$. Cela entraîne pour une spline à k intervalles : $S(t) = P_1(t), t_0 \leq t < t_1, S(t) = P_2(t), t_1 \leq t < t_2, \dots, S(t) = P_k(t), t_{k-1} \leq t \leq t_k$.

Le degré de la spline est défini comme étant celui du polynôme $P_i(t)$ de plus haut degré. Si tous les polynômes ont le même degré, on dit que la spline est uniforme. Dans le cas contraire, elle est non uniforme.

Tout polynôme étant C^∞ , la continuité d'une spline dépend de la continuité au niveau de la jointure des courbes polynômes. Si $\forall i$ tel que $1 \leq i \leq k$ et $\forall j$ tel que $0 \leq j \leq n$ l'égalité suivante est vérifiée :

$$P_i^{(j)}(t_i) = P_{i+1}^{(j)}(t_i) \quad (\text{A.46})$$

alors la spline est C^n .

A.2.3 Interpolation d'Hermite

L'interpolation d'Hermite consiste à chercher un polynôme qui non seulement prend les valeurs fixées en les abscisses données, mais dont également la dérivée, donc la pente de la courbe, prend

une valeur imposée en chacun de ces points. Naturellement, il faut pour cela un polynôme de degré supérieur au polynôme de Lagrange. On peut aussi imposer encore la valeur des dérivées secondes, troisièmes, etc. en chaque point. La démarche de l'interpolation newtonienne utilisant les différences divisées est particulièrement adaptée pour construire ces polynômes.

A.3 Méthodes d'approximation

A.3.1 Courbe de Bézier

Histoire

Pierre Bézier (ingénieur de l'École Nationale Supérieure d'Arts et Métiers en 1930 et de l'École Supérieure d'Électricité en 1931, docteur en mathématiques de l'université de Paris en 1977) est connu pour son invention des courbes et surfaces de Bézier, couramment utilisées en informatique.

Entré chez Renault en 1933, il y fera toute sa carrière jusqu'en 1975 au poste de directeur des méthodes mécaniques. Il y conçoit, en 1945, des machines transferts pour la ligne de fabrication des Renault 4CV, et, en 1958, l'une des premières machines à commande numérique d'Europe, une fraiseuse servant aux maquettes. Sa préoccupation était de créer un moyen simple et puissant pour modéliser des formes et faciliter la programmation des machines à commande numérique. Le problème auquel il s'attaque est celui de la modélisation des surfaces en trois dimensions, les commandes numériques se contentant jusqu'alors de courbes en deux dimensions. La solution qu'il cherche est celle d'une interface intuitive accessible à tout utilisateur. Il décide de considérer classiquement les surfaces comme une transformation de courbes. Son exigence de s'adapter au dessinateur et non de contraindre le dessinateur à devenir calculateur, l'amène à une inversion géniale, déduire le calcul à partir du dessin et non le dessin à partir du calcul. Il invente alors la poignée de contrôle, curseur de déplacement des courbes d'un dessin informatisé transmettant automatiquement les variations de coordonnées au processeur. Ces poignées de contrôle sont toujours utilisées aujourd'hui.



Bézier

Les courbes de Bézier sont des courbes polynomiales paramétriques. Elles ont de nombreuses applications dans la synthèse d'images et le rendu de polices de caractères (Pierre Bézier a travaillé sur les deux sujets). Ses recherches aboutirent à un logiciel, Unisurf, breveté en 1966. Il est à la base de tous les logiciels créés par la suite. Les concepts de CAO et de CFAO venaient de prendre forme. Ultérieurement, l'un des développeurs d'Apple, John Warnock, réutilise les travaux de Pierre Bézier pour élaborer un nouveau langage de dessin de polices : Postscript. Il crée ensuite en 1982, avec Charles M. Geschke, la société Adobe pour lancer un logiciel de dessin dérivé de ces résultats : Illustrator.

Notons que les splines existaient avant Bézier, mais leur défaut était de changer d'aspect lors d'une rotation de repère, ce qui les rendait inutilisables en CAO. Bézier partit d'une approche géométrique fondée sur la linéarité de l'espace euclidien et la théorie, déjà existante, du barycentre : si la définition est purement géométrique, aucun repère n'intervient puisque la construction en est indépendante. Les splines conformes aux principes de Bézier seront par la suite nommées B-splines.

Pour $n + 1$ points de contrôle (P_0, \dots, P_n) on définit une courbe de Bézier par l'ensemble des points :

$$b(t) = \sum_{i=0}^n B_i^n(t) \cdot P_i \quad \text{avec} \quad t \in [0; 1] \quad (\text{A.47})$$

où les B_i^n sont les polynômes de Bernstein. La suite des points P_0, \dots, P_n forme le polygone de contrôle de Bézier.

Chaque point de la courbe peut être vu alors comme un barycentre des $n + 1$ points de contrôle pondérés d'un poids égal au polynôme de Bernstein. Les principales propriétés des courbes de Bézier sont les suivantes :

- la courbe est à l'intérieur de l'enveloppe convexe des points de contrôle ;
- la courbe commence par le point P_0 et se termine par le point P_n , mais ne passe pas *a priori* par les autres points de contrôle ;

- $\mathbf{P}_0\mathbf{P}_1$ est le vecteur tangent à la courbe en \mathbf{P}_0 et $\mathbf{P}_{n-1}\mathbf{P}_n$ au point \mathbf{P}_n ;
- une courbe de Bézier est C^∞ ;
- la courbe de Bézier est un segment si et seulement si les points de contrôle sont alignés ;
- chaque restriction d'une courbe de Bézier est aussi une courbe de Bézier ;
- un arc de cercle, ni même aucun arc de courbe conique (en dehors du segment de droite) ne peut être décrit par une courbe de Bézier, quel que soit son degré ;
- le contrôle de la courbe est global : modifier un point de contrôle modifie toute la courbe, et non pas un voisinage du point de contrôle ;
- pour effectuer une transformation affine de la courbe, il suffit d'effectuer la transformation sur tous les points de contrôle.

A.3.2 B-Spline

Une B-spline est une combinaison linéaire de splines positives à support compact minimal. Les B-splines sont la généralisation des courbes de Bézier, elles peuvent être à leur tour généralisées par les NURBS. Étant donné $m + 1$ points t_i dans $[0; 1]$ tels que $0 \leq t_0 \leq t_1 \leq \dots \leq t_m \leq 1$, une courbe spline de degré n est une courbe paramétrique $S : [0; 1] \rightarrow \mathbb{R}^d$, composée de fonctions B-splines de degré n :

$$S(t) = \sum_{i=0}^{m-n-1} b_{i,n}(t) \cdot \mathbf{P}_i, \quad t \in [0; 1] \quad (\text{A.48})$$

où les \mathbf{P}_i forment un polygone appelé polygone de contrôle. Le nombre de points composant ce polygone est égal à $m - n$. Les $m - n$ fonctions B-splines de degré n sont définies par récurrence sur le degré inférieur :

$$b_{j,0}(t) = \begin{cases} 1 & \text{si } t_j \leq t < t_{j+1} \\ 0 & \text{sinon} \end{cases} \quad (\text{A.49})$$

$$b_{j,n}(t) = \frac{t - t_j}{t_{j+n} - t_j} b_{j,n-1}(t) + \frac{t_{j+n+1} - t}{t_{j+n+1} - t_{j+1}} b_{j+1,n-1}(t) \quad (\text{A.50})$$

Quand les points sont équidistants, les B-splines sont dites uniformes. C'est le cas des courbes de Bézier qui sont des B-splines uniformes, dont les points t_i (pour i entre 0 et m) forment une suite arithmétique de 0 à 1 avec un pas constant $h = 1/m$, et où le degré n de la courbe de Bézier ne peut être supérieur à m).

Par extension, lorsque deux points successifs t_j et t_{j+1} sont confondus, on pose $0/0 = 0$: cela a pour effet de définir une discontinuité de la tangente, pour le point de la courbe paramétré par une valeur de t , donc d'y créer un sommet d'angle non plat. Toutefois il est souvent plus simple de définir ce B-spline étendu comme l'union de deux B-splines définis avec des points distincts, ces splines étant simplement joints par ce sommet commun, sans introduire de difficulté dans l'évaluation paramétrique ci-dessus des B-splines pour certaines valeurs du paramètre t . Mais cela permet de considérer alors tout polygone simple comme une B-spline étendue.

La forme des fonctions de base est déterminée par la position des points. La courbe est à l'intérieur de l'enveloppe convexe des points de contrôle. Une B-spline de degré n , $b_{i,n}(t)$ est non nulle dans l'intervalle $[t_i; t_{i+n+1}]$:

$$b_{i,n}(t) = \begin{cases} > 0 & \text{si } t_i \leq t < t_{i+n+1} \\ 0 & \text{sinon} \end{cases} \quad (\text{A.51})$$

En d'autres termes, déplacer un point de contrôle ne modifie que localement l'allure de la courbe. Par contre, les B-splines ne permettent pas de décrire un arc de courbe conique.

A.3.3 B-splines rationnelles non uniformes

Ces objets couramment nommés NURBS, pour Non-Uniform Rational Basis Splines, correspondent à une généralisation des B-splines car ces fonctions sont définies avec des points en coordonnées homogènes. Les coordonnées homogènes, introduites par Möbius, rendent les calculs possibles dans l'espace projectif comme les coordonnées cartésiennes le font dans l'espace euclidien. Ces coordonnées homogènes sont largement utilisées en infographie ou en CAO car elles permettent la représentation de scènes en trois dimensions. Les NURBS parviennent à ajuster des courbes qui ne peuvent pas être représentées par des B-splines uniformes. Ils permettent même une représentation exacte de la totalité des arcs coniques ainsi que la totalité des courbes et surfaces polynomiales, avec uniquement des paramètres entiers ou rationnels si les NURBS passent par un nombre limité mais suffisant de points définis dans un maillage discret de l'espace.

Les fonctions NURBS de degré d sont définies par la formule doublement récursive de Cox-de Boor (formulation trouvée de manière indépendante par M.G. Cox en 1971 et C. de Boor en 1972) :

$$\begin{cases} b_{j,0}(t) = \begin{cases} 1 & \text{si } t_j \leq t < t_{j+1} \\ 0 & \text{sinon} \end{cases} \\ b_{j,d}(t) = \frac{t - t_j}{t_{j+d} - t_j} b_{j,d-1}(t) + \frac{t_{j+d+1} - t}{t_{j+d+1} - t_{j+1}} b_{j+1,d-1}(t) \end{cases} \quad (\text{A.52})$$

où les t_j sont des points. Lorsque plusieurs points t_j sont confondus, on peut encore poser $0/0 = 0$ comme pour les B-splines.

Annexe B

Intégration numérique

La méthode des éléments finis conduit à la discrétisation d'une formulation faible où la construction des matrices constitutives du système à résoudre nécessite le calcul d'intégrales. Dans certains cas particuliers, ou en utilisant des codes de calcul formel, ces intégrations peuvent être réalisées de manière exacte. Cependant, dans la plupart des cas et dans la plupart des codes de calcul, ces intégrations sont calculées numériquement. On parle alors de méthodes d'intégration numérique et de formules de quadrature.

B.1 Méthodes de Newton-Cotes

Soit à calculer l'intégrale suivante :

$$I = \int_a^b f(x) dx \quad (\text{B.1})$$

L'idée consiste à construire un polynôme pour interpoler $f(x)$ et à intégrer ce polynôme. Plusieurs types de polynômes peuvent être utilisés pour cette interpolation. Les principales méthodes d'interpolations ont été détaillées au chapitre A.

Méthode des rectangles

La méthode des rectangles consiste à interpoler $f(x)$ par un polynôme de degré 0, i.e. par la constante valant, selon les variantes de la méthode, soit $f(a)$, soit $f((a+b)/2)$. Comme cette approximation est très brutale, il est possible de subdiviser l'intervalle $[a; b]$ en plusieurs intervalles et d'appliquer la méthode sur chacun des intervalles, i.e. d'approcher f par une fonction en escalier. Si l'on subdivise l'intervalle $[a; b]$ en n intervalles égaux, il vient alors :

$$I \approx h \sum_{i=0}^{n-1} f(x_i) \quad (\text{B.2})$$

où $h = (b-a)/n$ est la longueur de chaque sous intervalle et $x_i = a + ih$ le point courant.

Méthode des trapèzes

La méthode des trapèzes consiste à interpoler $f(x)$ par un polynôme de degré 1, i.e. par la droite passant par les points $(a, f(a))$ et $(b, f(b))$. On obtient alors :

$$I \approx h \frac{f(a) + f(b)}{2} \quad (\text{B.3})$$

où $h = b - a$ est la longueur de l'intervalle, et l'erreur commise vaut $-\frac{h^3}{12}f''(w)$ pour un certain $w \in [a; b]$ (sous réserve que f soit 2 fois dérivable). L'erreur étant proportionnelle à f'' , la méthode est dite d'ordre 2, ce qui signifie qu'elle est exacte (erreur nulle) pour tout polynôme de degré inférieur ou égale à 1.

Comme cette approximation peut sembler un peu brutale, il est possible de subdiviser l'intervalle $[a; b]$ en plusieurs intervalles et d'appliquer cette formule sur chacun des intervalles, i.e. d'approcher f par une fonction affine continue par morceaux. Si l'on subdivise l'intervalle $[a; b]$ en n intervalles égaux, il vient alors :

$$I \approx \frac{(b-a)}{n} \sum_{i=0}^n f(x_i) \quad (\text{B.4})$$

où $h = (b-a)/n$ est la longueur de chaque sous intervalle, $x_i = a + ih$ le point courant et l'erreur commise vaut $-\frac{h^3}{12n^2}f''(w)$ pour un certain $w \in [a; b]$.

Remarques.

- la méthode de Romberg permet d'accélérer la convergence de la méthode des trapèzes ;
- la méthode des trapèzes est une méthode de Newton-Cotes pour $n = 1$.

Méthode de Simpson

La méthode de Simpson consiste à interpoler $f(x)$ par un polynôme de degré 2, i.e. par la parabole passant par les points extrêmes $(a, f(a))$ et $(b, f(b))$ et le point milieu $(c, f(c))$ avec $c = (a+b)/2$. On obtient alors :

$$I \approx \frac{h}{6} (f(a) + f(b) + f(c)) \quad (\text{B.5})$$

où $h = b - a$ est la longueur de l'intervalle, et l'erreur commise vaut $-\frac{h^5}{2880}f^{(4)}(w)$ pour un certain $w \in [a; b]$ (sous réserve que f soit quatre fois dérivable). L'erreur étant proportionnelle à $f^{(4)}$, la méthode est dite d'ordre 4, ce qui signifie qu'elle est exacte (erreur nulle) pour tout polynôme de degré inférieur ou égale à 3.

Comme dans le cas précédent, il est possible de subdiviser l'intervalle $[a; b]$ en plusieurs intervalles et d'appliquer cette formule sur chacun des intervalles. Si l'on subdivise l'intervalle $[a; b]$ en n intervalles égaux, avec n pair, il vient alors :

$$I \approx \frac{h}{3} \left(f(a) + f(b) + 2 \sum_{i=1}^{n/2-1} f(x_{2i}) + 4 \sum_{i=1}^{n/2} f(x_{2i-1}) \right) \quad (\text{B.6})$$

où $h = (b-a)/n$ est la longueur de chaque sous intervalle, $x_i = a + ih$ le point courant et l'erreur commise vaut $-\frac{nh^5}{180}f^{(4)}(w)$ pour un certain $w \in [a; b]$.

Remarques.

- la parabole interpolant f est trouvée en utilisant l'interpolation de Lagrange ;
- la méthode de Simpson est un cas particulier de celle de Newton-Cotes pour $n = 2$.

Méthode de Newton-Cotes

Les formules de Newton-Cotes se proposent également d'approximer l'intégrale I et découpant l'intervalle $[a; b]$ en n intervalles identiques. On posera donc encore une fois $h = (b-a)/n$ la longueur de chaque sous intervalle, et $x_i = a + ih$ le point courant. La formule est :

$$I \approx \sum_{i=0}^n w_i f(x_i) \quad (\text{B.7})$$

où les ϖ_i sont appelés poids ou coefficients de la quadrature et sont construits à partir des polynômes de Lagrange. La méthode de Newton-Cotes intègre exactement un polynôme de degré $n - 1$ avec n points.

Remarque. Il est possible de construire une formule de Newton-Cotes de degré quelconque. Toutefois, une telle formule n'est pas inconditionnellement stable. C'est pourquoi, on se cantonnera aux plus bas degrés : $n = 0$ méthode du point médian (i.e. méthode des rectangle où la valeur est évaluée en milieu d'intervalle) ; $n = 1$ méthode des trapèzes ; $n = 2$ méthode de Simpson dite 1/3, i.e. celle présentée avant ; $n = 3$ méthode de Simpson 3/8 (il suffit de faire le calcul) ; $n = 4$ méthode de Boole. Lorsque le degré augmente, des instabilités apparaissent, dues au phénomène de Runge. En effet, avec certaines fonctions (même infiniment dérivables), l'augmentation du nombre n de points d'interpolation ne constitue pas nécessairement une bonne stratégie d'approximation. Carle Runge a montré qu'il existe des configurations où l'écart maximal entre la fonction et son interpolation augmente indéfiniment avec n . Pour remédier à cela on peut utiliser les abscisses de Tchebychev au lieu de points équirépartis pour interpoler, ou plus simplement utiliser des splines (i.e. des polynômes par morceaux), et donc augmenter le nombre de morceaux et non le degré des polynômes.

B.2 Méthodes de quadrature de Gauß

Le principe de la méthode reste le même que pour la méthode de Newton-Cotes, mais on va essayer d'améliorer un peu encore la qualité du résultat. Pour cela, on souhaite que :

$$I = \int_a^b \varpi(x) f(x) dx \approx \sum_{i=1}^n \varpi_i f(x_i) \quad (\text{B.8})$$

où $\varpi(x) : (a ; b) \rightarrow \mathbb{R}$ est une fonction de pondération, qui peut assurer l'intégrabilité de f . Les ϖ_i sont appelés les poids ou coefficients de quadrature. Les x_i sont réels, distincts, uniques et sont les racines de polynômes orthogonaux (et non plus uniquement de Lagrange) pour le produit scalaire :

$$\langle f, g \rangle = \int_a^b f(x)g(x)\varpi(x)dx \quad (\text{B.9})$$

Ils sont appelés points ou nœuds de Gauß. Les poids et les nœuds sont choisis de façon à obtenir des degrés d'exactitude les plus grands possibles. Cette fois-ci $(a ; b)$ peut être n'importe quel type d'intervalle (fermé, ouvert, fini ou non).

Intégration sur un intervalle type

Intervalle $(a ; b)$	Fonction de pondération $\varpi(x)$	Famille de polynômes orthogonaux
$[-1 ; 1]$	1	Legendre
$] -1 ; 1[$	$(1 - x)^\alpha (1 + x)^\beta$, $\alpha, \beta > -1$	Jacobi
$] -1 ; 1[$	$\frac{1}{\sqrt{1-x^2}}$	Tchebychev (premier type)
$] -1 ; 1[$	$\sqrt{1-x^2}$	Tchebychev (second type)
\mathbb{R}^+	e^{-x}	Laguerre
\mathbb{R}	e^{-x^2}	Hermite

Tableau B.1: Polynômes et intégration

On rappelle que les nœuds sont déterminés comme les n racines du n ème polynôme orthogonal associé à la formule de quadrature. Les méthodes de quadrature de Gauß intègrent exactement un polynôme de degré $2n - 1$ avec n points.

Changement d'intervalle d'intégration

Si on intègre sur $(a ; b)$ au lieu de $(-1 ; 1)$, alors on fait un changement de variable. Finalement, on obtient l'approximation :

$$\frac{b-a}{2} \sum_{i=1}^n \varpi_i f\left(\frac{b-a}{2}x_i + \frac{a+b}{2}\right) \quad (\text{B.10})$$

Remarques.

- pour la méthode des éléments finis, l'intégration se déroule sur l'élément de référence, donc on n'a pas besoin de faire ce changement. Il est fait par la transformation affine entre l'élément considéré et l'élément de référence ;
- le nombre de points de Gauß ainsi que leurs positions sur l'élément sont donnés dans les documentations des logiciels (bien que vous sachiez désormais les trouver).

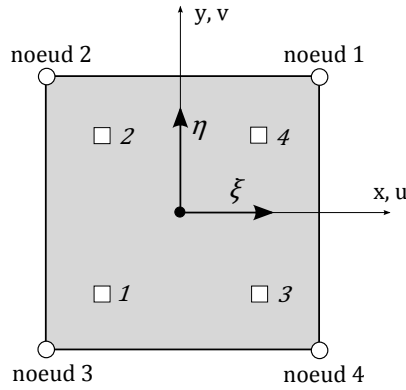


FIGURE B.1: Élément rectangulaire de référence Q_1 avec quatre points de Gauß

Intégration sur des carrés ou des cubes

Sur les carrés et les cubes, qui correspondent à ce qui nous intéresse en terme d'éléments de référence, on obtient les formules suivantes :

$$\int_{-1}^{-1} \int_{-1}^{-1} f(x, y) dx dy \approx \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \varpi_i \varpi_j f(x_i, x_j) \quad (\text{B.11})$$

$$\int_{-1}^{-1} \int_{-1}^{-1} \int_{-1}^{-1} f(x, y, z) dx dy dz \approx \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \sum_{k=1}^{n_z} \varpi_i \varpi_j \varpi_k f(x_i, x_j, x_k) \quad (\text{B.12})$$

où n_x , n_y et n_z sont les nombres de points de Gauß utilisés dans les directions x , y et z . Dans la pratique, on a souvent $n_x = n_y = n_z$.

Intégration sur un triangle ou un tétraèdre

Malheureusement, tous les éléments ne sont pas des segments, carrés ou cubes... on a également souvent à faire à des triangles et à des tétraèdres. Dans ce cas, on construit des formules spécifiques qui ne sont pas issues du cas unidimensionnel. L'élément triangulaire de référence est un triangle isocèle à côtés égaux de longueur 1. L'angle droit est à l'origine du repère. La forme générale d'intégration est :

$$I = \int_{\hat{K}} f(x, y) dx dy \approx \sum_{i=1}^n \varpi_i f(x_i, y_i) \quad (\text{B.13})$$

Les positions et poids des n points de Gauß sont choisis afin d'intégrer exactement un polynôme de degré N . Le tout est listé dans le tableau B.2. Le même travail peut être fait sur un tétraèdre.

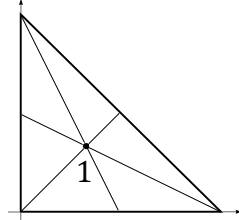
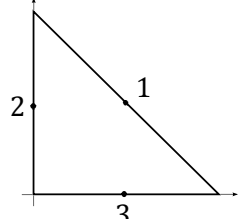
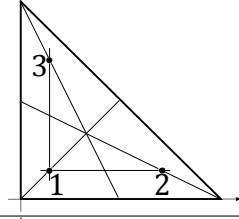
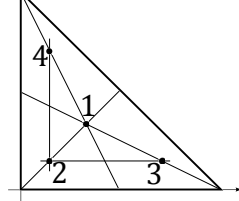
	n	x_i	y_i	ϖ_i	N
	1	1/3	1/3	1/2	1
	3	1/2	1/2	1/6	2
		0	1/2		
		1/2	0		
	3	1/6	1/6	1/6	2
		2/3	1/6		
		1/6	2/3		
	4	1/3	1/3	-27/96	4
		1/5	1/5	25/96	
		3/5	1/5	25/96	
		1/5	3/5	25/96	

Tableau B.2: Triangles et points de Gauß

Annexe C

Résolution des équations différentielles ordinaires

Résumé — La résolution exacte des équations différentielles fait partie des choses qui ont été demandées comme complément, même si elles ne correspondent pas vraiment au but de ce document. Toutefois, le paragraphe sur la résolution numérique des équations différentielles nous a permis d'introduire des méthodes qui sont employées également dans la méthode des éléments finis (notamment la méthode de Newmark).

C.1 Résolution exacte des équations différentielles linéaires

Une équation différentielle linéaire est de la forme suivante :

$$a_0(x)y + a_1(x)y' + a_2(x)y'' + \dots + a_n(x)y^{(n)} = f(x) \quad (\text{C.1})$$

où les coefficients $a_i(x)$ sont des fonctions numériques continues. Si les fonctions y dépendent d'une seule variable x , alors ces équations différentielles sont dites équations différentielles linéaires scalaires, si c'est un vecteur elles sont dites équations différentielles linéaires vectorielles ou système différentiel linéaire. Dans ce dernier cas, les a_i sont des applications linéaires. L'ordre d'une équation différentielle est le degré non nul le plus élevé des a_i , ici n . La méthode générale consiste à résoudre d'abord l'équation homogène, i.e. sans second membre, i.e. $f(x) = 0$, qui donne une solution contenant une ou des « constantes d'intégration » que l'on identifie ensuite en appliquant la forme générale trouvée à l'équation avec second membre.

C.1.1 Équation différentielle linéaire scalaire d'ordre 1

Il s'agit du cas particulier :

$$a(x)y' + b(x)y = c(x) \quad (\text{C.2})$$

où a , b et c sont des fonctions (connues).

Coefficients constants

Si a et b sont des constantes, alors l'équation homogène précédente s'écrit sous la forme :

$$y' = ky \quad (\text{C.3})$$

avec $k \in \mathbb{R}$, et la solution est :

$$f(x) = Ce^{kx} \quad (\text{C.4})$$

où la constante $C \in \mathbb{R}$ est déterminée à l'aide des conditions initiales :

- si pour x_0 on a $f(x_0) = y_0$ alors $C = y_0 e^{-k x_0}$.
- si $c = 0$, alors la solution du problème est celle de l'équation homogène et on a fini le travail.

Ce type d'équation se retrouve :

- $k < 0$: modélisation de la décroissance radioactive dans un milieu homogène et fermé ;
- $k > 0$: modélisation de la croissance d'une population (modèle simplifié, car en milieu fermé cette croissance ne peut durer indéfiniment).

Si $c \neq 0$, il faut déterminer une solution particulière de l'équation avec second membre. On appliquera les mêmes techniques que dans le cas des coefficients non constants (voir ci-après).

Coefficients non constants

1. Traitement de l'équation homogène

On réécrit l'équation homogène :

$$y' + \frac{b(x)}{a(x)}y = 0 \quad (\text{C.5})$$

sur un intervalle I où $a(x)$ ne s'annule pas. Ensuite, en notant $A(x)$, une primitive de la fonction $b(x)/a(x)$, il vient :

$$y' + A'(x)y = 0 \quad (\text{C.6})$$

puis :

$$y' e^{A(x)} + y \frac{b(x)}{a(x)} e^{A(x)} = 0 \quad (\text{C.7})$$

et :

$$y' e^{A(x)} + y A'(x) e^{A(x)} = 0 \quad (\text{C.8})$$

qui est de la forme $u'v + uv'$ et vaut $(uv)'$, d'où :

$$\frac{d}{dx}(y e^{A(x)}) = 0 \quad (\text{C.9})$$

soit :

$$y e^{A(x)} = C \quad (\text{C.10})$$

Les solutions sont alors les fonctions, définies sur I , de la forme :

$$y(x) = C e^{-A(x)} \quad (\text{C.11})$$

où encore une fois la constante $C \in \mathbb{R}$ est déterminée par les conditions initiales.

2. Solution particulière

Plusieurs cas se présentent :

- $c(x) = 0$: la solution du problème est celle de l'équation homogène et on a fini ;
- $c(x) \neq 0$: il faut déterminer une solution particulière de l'équation avec second membre. Cela n'est pas toujours simple car la forme de cette solution particulière varie en fonction de $c(x)$;
- $c(x) = C^e$: la quantité $f_0 = c/b$ est aussi une constante et l'ensemble des solutions est donc $f = y(x) + f_0$;
- $c(x)$ est une somme de fonctions : f_0 est alors la somme des solutions particulières obtenues pour chacun des termes de la somme constituant $c(x)$;
- $c(x)$ est un polynôme de degré n : f_0 est également un polynôme de degré n ;
- $c(x) = A \cos(\omega x + \varphi) + B \sin(\omega x + \varphi)$: on cherche f_0 comme combinaison linéaire de $\cos(\omega x + \varphi)$ et $\sin(\omega x + \varphi)$, i.e. sous la même forme que $c(x)$.

Dans le cas général, on utilise la méthode dite de la variation des constantes qui consiste à se ramener, par un changement de fonction variable, à un problème de calcul de primitive. On suppose que, sur l'intervalle d'étude, la fonction $a(x)$ ne s'annule pas¹. On connaît déjà la solution générale (C.11)

1. Il est possible de résoudre sur plusieurs intervalles de type I et essayer de « recoller » les solutions.

de l'équation homogène. On décide d'étudier le cas de la fonction $z(x)$ définie par :

$$z(x) = k(x)e^{-A(x)} \quad (\text{C.12})$$

en substituant dans $y(x)$ la fonction $k(x)$ à la constante C , d'où le nom de la méthode. En reportant dans l'équation différentielle initiale, il vient :

$$a(x)k'(x) = c(x)e^{A(x)} \quad (\text{C.13})$$

qui est une équation différentielle dépendant cette fois de la fonction $k(x)$. Si on note $B(x)$ une primitive de la fonction $c(x)e^{A(x)}/a(x)$, l'ensemble des solutions est alors :

$$k(x) = B(x) + C \quad (\text{C.14})$$

et la solution générale s'écrit alors sous la forme :

$$f(x) = (B(x) + C)e^{-A(x)} \quad (\text{C.15})$$

soit, finalement :

$$f = \exp\left(-\int \frac{b(x)}{a(x)} dx\right) \left\{ C + \int \frac{c(x)}{a(x)} \exp\left(\int \frac{b(x)}{a(x)} dx\right) dx \right\} \quad (\text{C.16})$$

On peut évidemment être confronté à un calcul d'intégral qui n'est pas simple (ou même pas possible à l'aide des fonctions usuelles).

C.1.2 Équation différentielle du premier ordre à variables séparées

Une équation différentielle d'ordre un à variables séparées est une équation différentielle qui peut se mettre sous la forme :

$$y' = f(x)g(y) \quad (\text{C.17})$$

Dans un tel problème, on commence par chercher les solutions régulières qui sont les solutions telles que $g(y)$ n'est jamais nul. Comme $g(y) \neq 0$, on peut écrire l'équation sous la forme :

$$\frac{1}{g(y(x))} y'(x) = f(x) \quad (\text{C.18})$$

par rapport à la variable x , ce qui conduit à :

$$\int_{x_0}^x \frac{1}{g(y(u))} y'(u) du = \int_{x_0}^x f(u) du \quad (\text{C.19})$$

et qui après changement de variable, est de la forme :

$$\int_{y_0}^y \frac{1}{g(v)} dv = \int_{x_0}^x f(u) du \quad (\text{C.20})$$

L'hypothèse $g(y) \neq 0$ écarte certaines solutions particulières. Par exemple, si y_0 est un point d'annulation de g , alors la fonction constante égale à y_0 est une solution maximale de l'équation. Une telle solution, dite solution singulière, est donc telle que $g(y)$ est toujours nul. Si la seule hypothèse faite sur g est la continuité, il peut exister des solutions « hybrides » constituées du raccordement de solutions régulières et singulières. D'une manière générale, pour une solution donnée, la quantité $g(y)$ sera soit toujours nulle, soit jamais nulle.

Il existe un cas particulier, qui est celui de l'équation différentielle d'ordre un à variables séparées autonome qui s'écrit :

$$y' = g(y) \quad (\text{C.21})$$

c'est-à-dire que la relation formelle ne dépend pas de x . Dans ce cas, si $x \mapsto y_0(x)$ est une solution, les fonctions obtenues par translation de la variable, de la forme $x \mapsto y_0(x + A)$, sont également solutions. Il y a en outre une propriété de monotonie, au moins pour les solutions régulières : puisque g ne s'annule pas, il garde alors un signe constant.

C.1.3 Équation différentielle linéaire d'ordre deux

Les équations différentielles linéaires d'ordre deux sont de la forme :

$$a(x)y'' + b(x)y' + c(x)y = d(x) \quad (\text{C.22})$$

où $a(x)$, $b(x)$, $c(x)$ et $d(x)$ sont des fonctions. Si elles ne peuvent pas toutes être résolues explicitement, beaucoup de méthodes existent.

Équation différentielle homogène

Pour l'équation différentielle homogène ($d(x) = 0$), une somme de deux solutions est encore solution, ainsi que le produit d'une solution par une constante. L'ensemble des solutions est donc un espace vectoriel et contient notamment une solution évidente, la fonction nulle.

Équation différentielle homogène à coefficients constants

On cherche des solutions sous forme exponentielle $f(x) = e^{\lambda x}$. Une telle fonction sera solution de l'équation différentielle si et seulement si λ est solution de l'équation caractéristique de l'équation différentielle :

$$a\lambda^2 + b\lambda + c = 0 \quad (\text{C.23})$$

Comme pour toute équation du second degré, il y a trois cas correspondant au signe du discriminant Δ :

1. $\Delta > 0$ — L'équation caractéristique possède deux solutions λ_1 et λ_2 , et les solutions de l'équation différentielle sont engendrées par $f_1(x) = e^{\lambda_1 x}$ et $f_2(x) = e^{\lambda_2 x}$, i.e. de la forme :

$$f(x) = C_1 f_1(x) + C_2 f_2(x) \quad (\text{C.24})$$

les constantes réelles C_1 et C_2 étant définies par :

- les conditions initiales : en un point (instant) donné x_0 , on spécifie les valeurs de $y_0 = y(x_0)$ et $y'_0 = y'(x_0)$. Dans ce cas l'existence et l'unicité de la solution vérifiant ces conditions initiales sont garanties.
- les conditions aux limites : pour de nombreux problèmes physiques, il est fréquent de donner des conditions aux limites en précisant les valeurs y_1 et y_2 aux instants x_1 et x_2 . Il y a alors fréquemment existence et unicité des solutions, mais ce n'est pas toujours vrai.

2. $\Delta = 0$ — L'équation caractéristique possède une solution double λ , et les solutions de l'équation différentielle sont de la forme :

$$f(x) = (C_1 x + C_2) e^{\lambda x} \quad (\text{C.25})$$

les constantes réelles C_1 et C_2 étant définies comme précédemment.

3. $\Delta < 0$ — L'équation caractéristique possède deux solutions λ_1 et λ_2 complexes conjuguées, et les solutions de \mathbb{R} dans \mathbb{C} de l'équation différentielle sont engendrées par $f_1(x) = e^{\lambda_1 x}$ et $f_2(x) = e^{\lambda_2 x}$, i.e. de la forme :

$$f(x) = C_1 f_1(x) + C_2 f_2(x) \quad (\text{C.26})$$

où cette fois C_1 et C_2 sont des complexes. Comme on cherche des solutions de \mathbb{R} dans \mathbb{R} , on note $\lambda_1 = u + iv$ (et donc $\lambda_2 = u - iv$), on exprime f_1 et f_2 , et on déduit que les

fonctions g_1 et g_2 , à valeurs dans \mathbb{R} cette fois, sont encore solutions :

$$g_1(x) = \frac{1}{2}(f_1(x) + f_2(x)) = e^{ux} \cos(vx) \quad (\text{C.27})$$

$$g_2(x) = \frac{1}{2i}(f_1(x) - f_2(x)) = e^{ux} \sin(vx) \quad (\text{C.28})$$

et engendrent encore l'ensemble des solutions. On a donc les solutions sous la forme :

$$f(x) = e^{ux}(C_1 \cos(vx) + C_2 \sin(vx)) \quad (\text{C.29})$$

les constantes réelles C_1 et C_2 étant définies comme précédemment. Notons que $f(x)$ s'écrit également :

$$f(x) = qe^{ux} \cos(vx + r) \quad (\text{C.30})$$

avec q et r deux réels à déterminer comme précédemment. Cette forme est parfois plus pratique selon les problèmes.

Équation différentielle homogène à coefficients non constants

Si les fonctions $a(x)$, $b(x)$ et $c(x)$ ne sont pas constantes, alors il n'existe pas d'expression générale des solutions. C'est pour cette raison qu'au XIX^e siècle furent introduites de nombreuses fonctions spéciales, comme les fonctions de Bessel ou la fonction d'Airy, définies comme solutions d'équations qu'il est impossible de résoudre explicitement.

Toutefois, dès lors qu'une solution particulière non nulle de l'équation est connue, il est possible de la résoudre complètement. En effet, le théorème de Cauchy-Lipschitz affirme que l'ensemble des solutions de l'équation constitue un espace vectoriel de dimension deux. Résoudre l'équation différentielle revient donc à exhiber deux fonctions solutions non proportionnelles : elles formeront une base de l'espace des solutions. Une telle base est appelée système fondamental de solutions.

Solution particulière et traitement du second membre

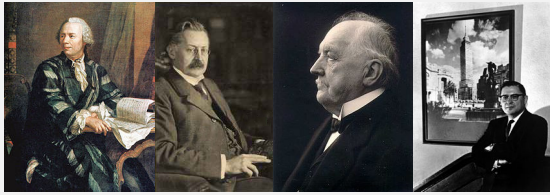
On peut agir de la même manière que pour les équations différentielles d'ordre un, et les mêmes remarques s'appliquent. On résout l'équation homogène puis on cherche une solution de l'équation avec second membre pour les connaître toutes.

Les équations différentielles d'ordre deux correspondent typiquement en physique aux problèmes dynamiques. Même si dans le cas réel on n'a rarement des phénomènes linéaires, des hypothèses de petits mouvements permettent de s'y ramener. Si cette hypothèse de petits déplacements ne peut être vérifiée, on aura alors recourt à des techniques dites de linéarisation, ou à des méthodes numériques comme la méthode de Newmark (qui sera présentée un petit peu plus loin).

C.2 Résolution numérique

Histoire

La première méthode numérique fut introduite en 1768 par Leonhard Euler.



Euler

Runge

Kutta

Newmark

Depuis, un grand nombre de techniques ont été développées : elles se basent sur la discrétisation de l'intervalle d'étude en un certain nombre de pas. Suivant le type de formule utilisé pour approcher les solutions, on distingue les méthodes numériques à un pas ou à pas multiples, explicites ou implicites.

Il existe plusieurs critères pour mesurer la performance des méthodes numériques : la consistance d'une méthode indique que l'erreur théorique effectuée en approchant la solution tend vers zéro avec les pas. La stabilité indique la capacité à contrôler l'accumulation des erreurs d'arrondi. Ensemble elles assurent la convergence, i.e. la possibilité de faire tendre l'erreur globale vers zéro avec le pas. Ces notions seront posées brièvement au chapitre E.

Dans le cas d'équations différentielles non linéaires, on passera forcément à une résolution numérique. Mais les méthodes numériques permettent évidemment aussi de résoudre numériquement les équations différentielles et les équations aux dérivées partielles.

L'idée générale est toujours la même : on approche la dérivée d'une fonction en un point par sa tangente (ce qui revient finalement à la définition de la dérivée). Pour une fonction $f(x)$, on écrit donc au point $x = a$ une relation de la forme :

$$f'(a) \approx \frac{f(b) - f(c)}{b - c} \quad (\text{C.31})$$

où b et c sont d'autres points. Par exemple, pour $c = a$ et $b = a + \varepsilon$ on obtient un schéma décentré à droite ; pour $c = a - \varepsilon$ et $b = a + \varepsilon$, on obtient un schéma centré.

C.2.1 Méthode d'Euler, Runge-Kutta d'ordre 1

Soit à résoudre l'équation différentielle suivante :

$$y' = f(t, y), \quad y(t_0) = y_0 \quad (\text{C.32})$$

D'après ce qui précède, on utilise une discrétisation de pas h , ce qui donne comme point courant $y_i = y_0 + ih$, et on fait l'approximation :

$$y' = \frac{y_{i+1} - y_i}{h} \quad (\text{C.33})$$

On obtient alors le schéma numérique :

$$y_{i+1} = y_i + hf(t, y_i) \quad (\text{C.34})$$

qui permet d'obtenir y_{i+1} uniquement en fonction de données au pas i . Cette méthode due à Euler, correspond également à la méthode de Runge-Kutta à l'ordre 1.

C.2.2 Méthode de Runge-Kutta d'ordre 2

La méthode de Runge-Kutta à l'ordre 2 est obtenue par amélioration de la méthode d'Euler en considérant le point milieu du pas h . Ainsi, on écrit cette fois :

$$y_{i+1} = y_i + h \cdot f\left(t + \frac{h}{2}, y_i + \frac{h}{2} f(t, y_i)\right) \quad (\text{C.35})$$

Mézalor me direz-vous, il manque des bouts... Les dérivées au milieu du pas d'intégration sont obtenues par :

$$y_{i+\frac{1}{2}} = y_i + \frac{h}{2} f(t, y_i) \quad \text{et} \quad y'_{i+\frac{1}{2}} = f\left(t + \frac{h}{2}, y_{i+\frac{1}{2}}\right) \quad (\text{C.36})$$

En réinjectant cela, on obtient sur le pas h complet :

$$y_{i+1} = y_i + h y'_{i+\frac{1}{2}} \quad (\text{C.37})$$

Notons qu'il s'agit du cas centré ($\alpha = 1/2$) de la formule plus générale :

$$y_{i+1} = y_i + h \left[\left(1 - \frac{1}{2\alpha}\right) f(t, y_i) + \frac{1}{2\alpha} f\left(t + \alpha h, y_i + \alpha h f(t, y_i)\right) \right] \quad (\text{C.38})$$

C'est une méthode d'ordre 2 car l'erreur est de l'ordre de h^3 .

C.2.3 Méthode de Runge-Kutta d'ordre 4

Aujourd'hui, le cas le plus fréquent est celui de l'ordre 4. L'idée est toujours d'estimer la pente de y , mais de façon plus précise. Pour cela, on ne prend plus la pente en un point (début ou milieu), mais on utilise la moyenne pondérée des pentes obtenues en 4 points du pas.

- $k_1 = f(t_i, y_i)$ est la pente au début de l'intervalle ;
- $k_2 = f\left(t_i + \frac{h}{2}, y_i + \frac{h}{2}k_1\right)$ est la pente au milieu de l'intervalle, en utilisant la pente k_1 pour calculer la valeur de y au point $t_i + h/2$ par la méthode d'Euler ;
- $k_3 = f\left(t_i + \frac{h}{2}, y_i + \frac{h}{2}k_2\right)$ est de nouveau la pente au milieu de l'intervalle, mais obtenue en utilisant la pente k_2 pour calculer y ;
- $k_4 = f(t_i + h, y_i + h k_3)$ est la pente en fin d'intervalle, avec la valeur de y calculée en utilisant k_3 .

On obtient finalement la discrétisation de Runge-Kutta à l'ordre 4 :

$$y_{i+1} = y_i + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4) \quad (\text{C.39})$$

La méthode est d'ordre 4, ce qui signifie que l'erreur commise à chaque étape est de l'ordre de h^5 , alors que l'erreur totale accumulée est de l'ordre de h^4 . Notons enfin que toutes ces formulations sont encore valables pour des fonctions à valeurs vectorielles.

C.2.4 Méthode de Crank-Nicolson

La méthode de Crank-Nicolson est un algorithme simple permettant de résoudre des systèmes d'équations différentielles partielles. Cette méthode utilise les différences finies en espace et la règle des trapèzes en temps pour approcher une solution du problème : elle est numériquement stable et quadratique pour le temps. On peut facilement la généraliser à des problèmes à deux ou trois dimensions.

Si l'on considère l'équation différentielle suivante :

$$\frac{\partial u}{\partial t} = F\left(u, x, t, \frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial x^2}\right) \quad (\text{C.40})$$

alors, en notant u_i^n le terme $u(i \Delta x, n \Delta t)$, le schéma est :

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} = \frac{1}{2} \left[F_i^{n+1}\left(u, x, t, \frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial x^2}\right) + F_i^n\left(u, x, t, \frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial x^2}\right) \right] \quad (\text{C.41})$$

où la fonction F est discrétisée à l'aide des différences finies centrées en espace.

Ce schéma est celui utilisé pour les équations paraboliques, comme celles de mécanique des fluides. On l'utilise également pour des problèmes de mécanique quantique, de thermodynamique hors-équilibre et d'électromagnétisme... et pour tout phénomène pouvant être ramenés à l'étude de l'équation de la chaleur (qui a été le cadre de son développement).

La méthode converge en $O(\Delta_t^2)$. L'étude de convergence est d'ailleurs plus aisée que dans le cas des équations hyperboliques pour lesquelles la méthode de Newmark est préférée.

C.2.5 Méthode de Newmark

La méthode de Newmark (1959) permet de résoudre numériquement des équation différentielle du second ordre. Elle convient, non seulement pour des systèmes différentiels linéaires, mais aussi pour des systèmes fortement non-linéaires avec une matrice de masse et une force appliquée qui peuvent dépendre à la fois de la position et du temps. Dans ce second cas, le calcul nécessite à chaque pas une boucle d'itération.

L'idée générale reste la même : on cherche à estimer les valeurs des dérivées (premières, secondes...) à l'instant t à partir des informations disponible à l'instant précédent (au pas de temps précédent). Pour cela, on va recourir à un développement limité. Considérons l'équation de la dynamique :

$$M \ddot{x}(t) + C \dot{x}(t) + K x(t) = f(t) \quad (C.42)$$

On fait un développement en série de Taylor :

$$u_{t+\Delta_t} = u_t + \Delta_t \dot{u}_t + \frac{\Delta_t^2}{2} \ddot{u}_t + \beta \Delta_t^3 \dddot{u} \quad \text{et} \quad \dot{u}_{t+\Delta_t} = \dot{u}_t + \Delta_t \ddot{u}_t + \gamma \Delta_t^2 \dddot{u} \quad (C.43)$$

et on fait l'hypothèse de linéarité de l'accélération à l'intérieur d'un pas de temps Δ_t :

$$\ddot{u} = \frac{\ddot{u}_{t+\Delta_t} - \ddot{u}_t}{\Delta_t} \quad (C.44)$$

Les différents schémas de Newmark correspondent à des valeurs particulières de β et γ . Dans le cas $\beta = 0$ et $\gamma = 1/2$, on retombe sur le schéma des différences finies centrées.

La méthode de Newmark fonctionne également pour les problèmes non linéaires, mais dans ce cas la matrice de rigidité devra être réévaluée à chaque pas de temps (ainsi que celle d'amortissement dans les cas les plus compliqués).

Annexe D

Méthode de Newton Raphson

Résumé — La méthode de Newton-Raphson est, dans son application la plus simple, un algorithme efficace pour trouver numériquement une approximation précise d'un zéro (ou racine) d'une fonction réelle d'une variable réelle.

D.1 Présentation

Histoire

La méthode de Newton fut décrite par Newton dans *De analysi per aequationes numero terminorum infinitas*, écrit en 1669 et publié en 1711 par William Jones. Elle fut à nouveau décrite dans *De metodis fluxionum et serierum infinitarum* (De la méthode des fluxions et des suites infinies), écrit en 1671, traduit et publié sous le titre *Methods of Fluxions* en 1736 par John Colson. Toutefois, Newton n'appliqua la méthode qu'aux seuls polynômes. Comme la notion de dérivée et donc de linéarisation n'était pas définie à cette époque, son approche diffère de l'actuelle méthode : Newton cherchait à affiner une approximation grossière d'un zéro d'un polynôme par un calcul polynomial.



Newton

Jones

Colson



Wallis



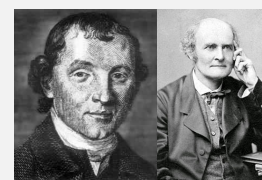
Raphson

Cette méthode fut l'objet de publications antérieures. En 1685, John Wallis en publia une première description dans *A Treatise of Algebra both Historical and Practical*. En 1690, Joseph Raphson en publia une description simplifiée dans *Analysis aequationum universalis*. Raphson considérait la méthode de Newton toujours comme une méthode purement algébrique et restreignait aussi son usage aux seuls polynômes. Toutefois, il mit en évidence le calcul récursif des approximations successives d'un zéro d'un polynôme au lieu de considérer

comme Newton une suite de polynômes.

C'est Thomas Simpson qui généralisa cette méthode au calcul itératif des solutions d'une équation non linéaire, en utilisant les dérivées (qu'il appelait fluxions, comme Newton). Simpson appliqua la méthode de Newton à des systèmes de deux équations non linéaires à deux inconnues, en suivant l'approche utilisée aujourd'hui pour des systèmes ayant plus de 2 équations, et à des problèmes d'optimisation sans contrainte en cherchant un zéro du gradient.

Arthur Cayley fut le premier à noter la difficulté de généraliser la méthode de Newton aux variables complexes en 1879, par exemple aux polynômes de degré supérieur à 3.



Simpson

Cayley

Sous sa forme moderne, l'algorithme se déroule comme suit : à chaque itération, la fonction dont on cherche un zéro est linéarisée en l'itéré (ou point) courant ; et l'itéré suivant est pris égal au zéro de la fonction linéarisée.

Cette description sommaire indique qu'au moins deux conditions sont requises pour la bonne marche de l'algorithme : la fonction doit être différentiable aux points visités (pour pouvoir y linéariser la fonction) et les dérivées ne doivent pas s'y annuler (pour que la fonction linéarisée ait un zéro) ; s'ajoute à ces conditions la contrainte forte de devoir prendre le premier itéré assez proche d'un zéro régulier de la fonction (i.e. en lequel la dérivée de la fonction ne s'annule pas), pour que la convergence du processus soit assurée.

L'intérêt principal de l'algorithme de Newton-Raphson est sa convergence quadratique locale. En termes imagés mais peu précis, cela signifie que le nombre de chiffres significatifs corrects des itérés double à chaque itération, asymptotiquement. Comme le nombre de chiffres significatifs représentables par un ordinateur est limité (environ 15 chiffres décimaux sur un ordinateur avec un processeur 32-bits), on peut simplifier grossièrement en disant que, soit il converge en moins de 10 itérations, soit il diverge. En effet, si l'itéré initial n'est pas pris suffisamment proche d'un zéro, la suite des itérés générée par l'algorithme a un comportement erratique, dont la convergence éventuelle ne peut être que le fruit du hasard (i.e. si l'un des itérés est par chance proche d'un zéro).

L'importance de l'algorithme de Newton-Raphson a incité les numériciens à étendre son application et à proposer des remèdes à ses défauts.

Par exemple, l'algorithme permet également de trouver un zéro d'une fonction de plusieurs variables à valeurs vectorielles, voire définie entre espaces vectoriels de dimension infinie ; la méthode conduit d'ailleurs à des résultats d'existence de zéro.

On peut aussi l'utiliser lorsque la fonction est différentiable dans un sens plus faible, ainsi que pour résoudre des systèmes d'inégalités non linéaires, des problèmes d'inclusion, d'équations différentielles ou d'équations aux dérivées partielles, d'inéquations variationnelles...

On a également mis au point des techniques de globalisation de l'algorithme, lesquelles ont pour but de forcer la convergence des suites générées à partir d'un itéré initial arbitraire (non nécessairement proche d'un zéro)

Dans les versions dites inexactes ou tronquées, on ne résout le système linéaire à chaque itération que de manière approchée.

Enfin, la famille des algorithmes de quasi-Newton (par exemple si l'on ne connaît pas l'expression analytique de la fonction dont on cherche une racine) propose des techniques permettant de se passer du calcul de la dérivée de la fonction.

Toutes ces améliorations ne permettent toutefois pas d'assurer que l'algorithme trouvera un zéro existant, quel que soit l'itéré initial.

Appliqué à la dérivée d'une fonction réelle, cet algorithme permet d'obtenir des points critiques (i.e. des zéros de la dérivée). Cette observation est à l'origine de son utilisation en optimisation sans ou avec contraintes.

D.2 Algorithme

Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ la fonction dont on cherche à construire une bonne approximation d'un zéro. pour cela, on se base sur son développement de Taylor au premier ordre.

Partant d'un point x_0 que l'on choisit de préférence proche du zéro à trouver (en faisant des estimations grossières par exemple), on approche la fonction au premier ordre, autrement dit, on la considère à peu près égale à sa tangente en ce point :

$$f(x) \simeq f(x_0) + f'(x_0)(x - x_0) \quad (\text{D.1})$$

Pour trouver un zéro de cette fonction d'approximation, il suffit de calculer l'intersection de la droite tangente avec l'axe des abscisses, i.e. de résoudre :

$$0 = f(x_0) + f'(x_0)(x - x_0) \quad (\text{D.2})$$

On obtient alors un point x_1 qui en général a de bonnes chances d'être plus proche du vrai zéro de f que le point x_0 précédent. Par cette opération, on peut donc espérer améliorer l'approximation par itérations successives.

Cette méthode requiert que la fonction possède une tangente en chacun des points de la suite que l'on construit par itération. Cela est évidemment vrai si f est dérivable.

Formellement, on construit donc la suite :

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \quad (\text{D.3})$$

à partir d'un point x_0 .

Bien que la méthode soit très efficace, certains aspects pratiques doivent être pris en compte :

- Avant tout, la méthode de Newton-Raphson nécessite que la dérivée soit effectivement calculée. Dans les cas où la dérivée est seulement estimée en prenant la pente entre deux points de la fonction, la méthode prend le nom de méthode de la sécante, moins efficace.
- Par ailleurs, si la valeur de départ est trop éloignée du vrai zéro, la méthode de Newton-Raphson peut entrer en boucle infinie sans produire d'approximation améliorée. À cause de cela, toute mise en œuvre de la méthode de Newton-Raphson doit inclure un code de contrôle du nombre d'itérations.

Annexe E

Schémas numériques

Résumé — Il nous est apparu qu'il était nécessaire d'ajouter quelques mots encore sur les schémas numériques, notamment après la présentation des chapitres C et D. Nous resterons évidemment brefs, tant le sujet est conséquent.

Nous nous sommes efforcés de présenter la méthode des éléments finis ainsi que la théorie mathématique la sous-tendant.

La partie II nous a permis de voir que l'existence de solutions, dans le cas continu, est assurée notamment par les théorèmes de Lax-Milgram, (paragraphe 7.3) Babuška (paragraphe 7.4) et Brezzi (paragraphe 7.5) selon le type de formulation. La partie III, nous a permis de s'assurer de la convergence de la solution approximée vers la solution continue, notamment par les lemmes de Céa (paragraphe 10.4) et de Strang (paragraphe 10.5).

Toutefois, nous avons vu, lorsque nous prenons en compte certains phénomènes un peu plus complexes (non stationnarité, non linéarité), que nous devons avoir recours à d'autres algorithmes tels que les différences finies, et les algorithmes de Newmark et Newton-Raphson par exemple. Or nous n'avons rien dit de très général sur la converges des schémas numériques, c'est le pourquoi de ce petit chapitre.

E.1 Les propriétés d'un schéma numérique

E.1.1 Problème bien posé

Selon une définition due à Hadamard à propos des modèles mathématiques de phénomènes physiques, un problème est bien posé s'il possède les propriétés suivantes :

1. ce problème possède une solution (existence) ;
2. cette solution est unique (unicité) ;
3. et elle dépend de façon continue des données dans le cadre d'une topologie raisonnable.

Typiquement, des problèmes issus de la physique et pour lesquels on peut observer (mesurer) les grandeurs physiques correspondantes sont bien posés. On pensera notamment au problème de Dirichlet pour l'équation de Laplace et à l'équation de la chaleur (avec conditions initiales connues).

Encore une fois, la mécanique nous montre que tout n'est pas si simple. Un problème de mécanique des milieux continus est bien posé dans un domaine Ω si la frontière $\Gamma = \partial\Omega$ de ce domaine admet une partition en deux sous-ensembles Γ_1 et Γ_2 sur lesquels portent des conditions aux limites sur les déplacements et sur les efforts respectivement (un théorème de Douglas N. Arnold existe dans le cas où les deux conditions existent simultanément, mais nous n'en parlerons pas ici).

Si un problème de mécanique est bien posé, la solution en contrainte existe et est unique, alors que la solution en déplacement ne l'est pas nécessairement lorsque des mouvements de corps rigides sont possibles.

Les problèmes inverses fournissent souvent des problèmes mal posés. Par exemple, le problème dit d'inversion du temps dans l'équation de la chaleur consistant à déduire une distribution passée de la température à partir d'un état final n'est pas bien posé. Sa solution est en effet très sensible à des perturbations de l'état final.

Comme nous l'avons déjà mentionné, la recherche de la solution d'un problème continu en utilisant des méthodes numériques sur la discrétisation de ce problème comporte de nombreux types d'erreurs, de la simple erreur d'arrondi dans une donnée au comportement d'un algorithme. C'est ce dont nous allons parler tout de suite.

E.1.2 Conditionnement

Le conditionnement mesure la dépendance de la solution d'un problème numérique par rapport aux données du problème, ceci afin de contrôler la validité d'une solution calculée par rapport à ces données. En effet, les données d'un problème numérique dépendent en général de mesures expérimentales et sont donc entachées d'erreurs.

De façon plus générale, on peut dire que le conditionnement associé à un problème est une mesure de la difficulté de calcul numérique du problème. Un problème possédant un conditionnement bas est dit bien conditionné et un problème possédant un conditionnement élevé est dit mal conditionné.

E.1.3 Stabilité ou robustesse

La stabilité est une propriété de la solution obtenue. Elle se réfère à la propagation des erreurs au cours des étapes du calcul, à la capacité de l'algorithme de ne pas trop amplifier d'éventuels écarts, à la précision des résultats obtenus, mais elle ne se limite pas aux erreurs d'arrondis et à leurs conséquences. Une solution est dite stable si elle est bornée dans l'espace et/ou le temps. La valeur de la stabilité peut parfois (souvent) être exprimée en fonction du pas de discrétisation.

Les algorithmes dédiés à la résolution d'équations différentielles ou d'équations aux dérivées partielles (en particulier la méthode des différences finies et la méthode des éléments finis) se basent sur une discrétisation ou un maillage de l'espace (et du temps) : dans ce cas, la stabilité se réfère à un comportement numérique robuste lorsque le pas de discrétisation ou la taille des mailles tend vers 0.

La stabilité d'un schéma n'a aucun lien avec la solution exacte du problème traité (convergence).

Une solution U_n est stable si l'on peut trouver une constante C (souvent nulle) telle que :

$$-1 \leq \frac{U_{n+1} - C}{U_n - C} \leq 1, \quad \forall n \quad (\text{E.1})$$

Ce critère simple permet donc de s'assurer que la solution ne s'éloigne pas d'une valeur de référence.

E.1.4 Consistance

La consistance est une propriété de la discrétisation. Un schéma numérique (d'une équation aux dérivées partielles par exemple) sera dit consistant par rapport au problème qu'il discrétise si celui-ci tend vers le problème considéré lorsque la discrétisation tend vers 0. La consistance concerne essentiellement la capacité du schéma à représenter une solution régulière satisfaisant localement les équations aux dérivées partielles, ceci lorsque les pas de discrétisation (Δt , Δx , etc.) tendent tous vers 0. Plus précisément, si les données d'une étape du traitement algorithmique sont issues d'une solution exacte, les résultats de ce traitement tendent vers cette solution.

La différence entre l'équation discrétisée et l'équation réelle est appelée l'erreur de troncature

La consistance d'une discrétisation s'analyse en effectuant un développement en série de Taylor de l'équation discrétisée et en vérifiant que celle-ci tend vers l'équation originale lorsque le pas de discrétisation tend vers 0.

Sur le calcul de l'erreur de troncature. Le calcul de l'erreur de troncature d'un schéma est souvent délicat à mener. Si on ne procède pas de manière soignée et méthodique, on peut aisément se retrouver englué dans un calcul inextricable, dont le coût croît exponentiellement en fonction de l'ordre à déterminer. Quelques règles simples permettent en général d'éviter ce travers.

L'erreur de troncature se calcule en développant tous les termes du schéma au même point à l'aide des formules de Taylor. Le point choisi n'a évidemment aucune influence sur le résultat obtenu (l'ordre du schéma ne dépend pas du point considéré). Par contre, ce choix influe sur la taille du calcul qui en résulte.

Il est recommandé de diviser le calcul en plusieurs étapes. Les développements calculés lors d'une étape pouvant être réutilisés à une autre. Il faut absolument utiliser l'équation vérifiée par la solution (par exemple remplacer les dérivées en temps par des dérivées en espace). Cela simplifie considérablement les calculs, et nous permet de déterminer l'ordre optimal du schéma. Il faut éviter à tout prix d'effectuer des calculs inutiles et ne pas manipuler des termes d'ordre non significatifs.

Enfin, un petit truc classique consiste à utiliser les symétries du schéma, qui peuvent impliquer que les termes non nuls du développement sont nécessairement soit pairs, soit impairs.

E.1.5 Convergence

Contrairement à la consistance, qui est une propriété locale, la convergence est de portée globale.

On dit qu'une solution numérique converge vers la solution analytique si elle tend vers elle en tout point du temps et/ou de l'espace lorsque les paramètres de discrétisation (Δt , Δx ...) tendent vers 0.

C'est évidemment bien la convergence (souvent difficile à prouver) d'un schéma numérique que nous visons, mais stabilité et consistance (plus faciles à prouver) sont les outils nous permettant de nous en assurer, et ceci via les très beaux théorèmes de Lax (celui du théorème de Lax-Milgram) et de Lax-Wendroff, que nous présentons maintenant.

E.2 Théorèmes de convergence

Théorème 67 — Théorème de Lax. Pour résoudre un problème évolutif avec condition initiale qui est supposé être bien posé, ceci à l'aide d'un schéma numérique consistant, la stabilité du schéma est une condition nécessaire et suffisante pour assurer sa convergence. Les notions de consistance, de stabilité et de convergence se réfèrent ici à une même norme.

En d'autres termes : consistance + stabilité \implies convergence

Ce théorème est parfois également appelé théorème de Lax-Richtmyer

Dans les applications physiques que nous considérons, le théorème suivant permet d'assurer la convergence du schéma numérique utilisé.

Théorème 68 — Théorème de Lax-Wendroff. Pour résoudre un problème aux dérivées partielles basé sur une loi de conservation, un schéma numérique qui est à la fois conservatif, consistant et convergent (lorsque l'on raffine les pas de temps et d'espace, i.e. lorsque $\Delta t \rightarrow 0$ et $\Delta x \rightarrow 0$), alors la solution numérique converge vers une solution faible des équations.

Notons que Burton Wendroff a passé sa thèse sous la direction de Peter Lax.

Annexe F

Index des noms propres

A

Abel (Niels Henrik), 1802-1829, Norvégien, 27, 29
Adrain (Robert), 1775-1843, Américain, 371
Airy (George Biddell, Sir -), 1801-1892, Anglais, 395
Alart (Pierre), ?-, Français, 314
Allaire (Grégoire), ?-, Français, 210, 228, 229
Allard (Jean-François), ?-, Français, 283
Ambrose (Warren Arthur), 1914-1995, Américain, 82
Archimède de Syracuse, -287– -212, Grec, 36, 141
Argyris (John Hadji), 1913-2004, Grec, 130
Aristote (dit le Stagirite), -384– -322, Grec, 81, 141
Arnold (Douglas Norman), ?-, Américain, 361, 403
Arnold (Vladimir Igorevitch), 1937-2010, Russe, 340
Ascoli (Giulio), 1843-1896, Italien, 17

B

Babuška (Ivo Milan), 1926-, Tchèque, 100, 136, 359, 403
Bachet (Claude-Gaspard, dit - de Mézirac), 1581-1638, Français, 103
Bakhvalov (Nikolai Sergeevich), 1934-2005, Russe, 210
Bampton (Mervyn Cyril Charles), ?, Américain, 256, 257
Banach (Stephan), 1892-1945, Polonais, 20, 55, 69, 70
Beck (Henry Charles, dit Harry), 1902-1974, Anglais, 26
Bell (Alexander Graham), 1847-1922, Britannique, 273
Bellman (Richard Ernest), 1920-1984, Américain, 218
Bérenger (Jean-Pierre), ?-, Français, 285
Bernoulli (Daniel), 1700-1782, Suisse, 146
Bernoulli (Jacques), 1654-1705, Suisse, 146, 218
Bernoulli (Jean), 1667-1748, Suisse, 33, 96, 146, 218
Bernstein (Sergeï Natanovich), 1880-1968, Russe, 379, 381
Bessel (Friedrich Wilhelm), 1784-1846, Allemand, 395
Beyle (Henri), dit Stendhal, 1783-1842, Français, 80
Bézier (Pierre), 1910-1999, Français, 379, 381, 382
Bhatnagar (Prabhu Lal), 1912-1976, Indien, 360, 361
Bingham (Eugene Cook), 1878-1945, Américain, 304
Biot (Maurice Anthony), 1905-1985, Américain, 283

Bochner (Salomon), 1899-1982, Américain, 57
Bolotin (Vladimir V.), 1926-2008, Russe, 353
Boltzmann (Ludwig), 1944-1906, Autrichien, 82, 85, 305, 340, 360
Bolzano (Bernard Placidus Johann Nepomuk), 1781-1848, Allemand, 33
Bonaparte (Napoléon), 1769-1821, Français, 80
Bonnet (Pierre-Ossian), 1819-1892, Français, 376
Boole (George), 1815-1864, Anglais, 22
Borel (Félix Édouard Justin Émile), 1871-1956, Français, 23, 25
Boussinesq (Joseph Valentin), 1842-1929, Français, 313
Brezzi (Franco), 1945-, Italien, 100, 101, 112, 121–123, 137, 144, 163, 403
Brown (Robert), 1773-1859, Écossais, 79

C

Cameron (Robert Horton), 1908-1989, Américain, 345
Cantor (Georg Ferdinand Ludwig Philip), 1845-1918, Allemand, 17, 24, 249
Carathéodory (Constantin), 1873-1950, Grec, 25
Castigliano (Carlo Alberto), 1847-1884, Italien, 130
Cauchy (Augustin Louis, baron -), 1789-1857, Français, 19, 27, 55, 69, 76, 249, 298, 373, 395
Cayley (Arthur), 1821-1895, Anglais, 249, 250, 399
Céa (Jean), ?-, Français, 135–137, 403
Champollion (Jean-François), 1790-1832, Français, 80
Charpit de Villecourt (Paul), ?-1784, Français, 77
Chasles (Michel), 1793-1880, Français, 307
Cherepanov (Genady P.), 1937-, Américain, 329
Chevreuil (Mathilde), ?-, Française, 259
Chladni (Ernst Florens Friedrich), 1756-1827, Allemand, 258
Cholesky (André-Louis), 1875-1918, Français, 135, 251, 349
Clairaut (Alexis Claude), 1713-1765, Français, 75
Clough (Ray William), 1920-?, Américain, 130
Colson (John), 1680-1760, Anglais, 399
Cornell (Carl Allin), 1938-2007, Américain, 353
Cotes (Roger), 1682-1716, Anglais, 385, 386
Cottrell (Alan Howard, Sir -), 1919-2012, Anglais, 332

- Coulomb (Charles-Augustin), 1736-1806, Français, 303, 308, 310, 313, 315, 316
- Courant (Richard), 1888-1972, Américain, 57, 130, 261, 284
- Cousin (Pierre), 1867-1933, Français, 48
- Cox (Gertrude Mary), 1900-1978, Américaine, 383
- Craig (Roy R. Jr), ?, Américain, 256, 257
- Crank (John), 1916-2006, Américain, 397
- Curnier (Alain), 1948-, Suisse, 314
- D**
- d'Alembert (Jean le Rond), 1717-1783, Français, 29, 75, 76, 97
- Dantzig (George Bernard), 1914-2005, Américain, 218
- de Boor (Carl-Wilhelm Reinhold), 1937-, Américain, 383
- De Giorgi (Ennio), 1928-1996, Italien, 82
- De Saxé (Géry), 1955-, Français, 314
- Delaunay (Boris Nikolaïevitch), 1890-1980, Russe, 201, 204, 207
- Denjoy (Arnaud), 1884-1974, Français, 48
- Der Kiureghian (Armen), 1947-, Arménien, 353
- Descartes (René), 1596-1650, Français, 36, 204, 218, 264
- Desvillettes (Laurent), 1966-, Français, 85
- Didon, ?, Carthaginoise, 218
- Dirac (Paul Adrien Maurice), 1902-1984, Anglais, 24, 56, 58
- Dirichlet (Johann Peter Gustav Lejeune), 1805-1859, Allemand, 45, 46, 65, 77, 81, 105, 106, 109-111, 113, 117, 204, 210, 215
- Ditlevsen (Ove Dalager), 1935-, Danois, 353, 354
- Drucker (Daniel Charles), 1918-2001, Américain, 308
- du Bois-Reymond (Paul David Gustave), 1831-1889, Allemand, 81
- Dugdale (Donald Stephen), 332
- Dynkin (Eugène Borisovitch), 1924-, Russe, 25
- E**
- Eiffel (Gustave), 1832-1923, Français, 146, 307
- Einstein (Albert), 1879-1955, Suisse-américain, 79
- Esteva (Luis), 1935-2012, Mexicain, 353
- Euclide, -325- -265, Grec, 19-21, 25, 40, 67, 69, 133
- Euler (Leonhard Paul), 1707-1783, Suisse, 17, 27, 33, 36, 76, 77, 81, 86, 103, 111, 146, 208, 218, 221, 372, 396
- Eyring (Carl Ferdinand), 1889-1951, Américain, 270, 276
- F**
- Falk (Richard S.), ?-, Américain, 361
- Fechner (Gustav Theodor), 1801-1887, Allemand, 273
- Feng (Zhi-Qiang), 1963-, Français, 314
- Fermat (Pierre, de -), ?-1665, Français, 36, 218
- Fick (Adolf Eugene), 1829-1901, Allemand, 79, 82
- Fix (George J. III), 1939-2002, Américain, 130
- Fletcher (Harvey C.), 1884-1981, Américain, 274
- Fourier (Jean Baptiste Joseph), 1768-1830, Français, 45, 79-81, 249, 259, 261, 345
- Fréchet (Maurice René), 1878-1973, Français, 17, 19, 22, 98, 211, 220
- Fredholm (Ivar), 1866-1927, Suédois, 344
- Freudenthal (Alfred Martin), 1906-1977, Américain, 353
- Friedrichs (Kurt Otto), 1901-1982, Allemand, 57, 67, 284
- Frobenius (Ferdinand Georg), 1849-1917, Allemand, 250
- G**
- Galerkine (Boris), 1871-1945, Russe, 97, 130, 136, 261
- Galilée (Galileo Galilei), 1564-1642, Italien, 141, 146, 218
- Galois (Évariste), 1811-1832, Français, 29
- Gauß (Johann Carl Friedrich), 1777-1855, Allemand, 135, 174, 249, 334, 371, 376, 387-389
- Geschke (Charles M.), 1939-, Américain, 381
- Ghanem (Roger G.), ?-, Libanais, 350
- Gibbs (Josiah Willard), 1839-1903, Américain, 81
- Goodyear (Charles), 1800-1860, Américain, 309
- Grassmann (Hermann Günther), 1809-1877, Allemand, 249
- Green (George), 1793-1841, Anglais, 44, 264, 281, 301, 302, 357
- Griffith (Alan Arnold), 1893-1963, Anglais, 325-327, 329, 334
- Gross (Eugène P.), 1926-1991, Américain, 360, 361
- Guest, 308
- Gumbel (Emil Julius), 1891-1966, Allemand, 353
- H**
- Haar (Alfréd), 1885-1933, Hongrois, 25
- Hadamard (Jacques Salomon), 1865-1963, Français, 17, 55, 403
- Halley (Edmond), 1656-1742, Anglais, 142
- Hamilton (Richard Streit), 1943-, Américain, 84
- Hamilton (William Rowan, Sir -), 1805-1865, Irlandais, 142, 218, 249, 250
- Hasofer (A. M.), 353, 354
- Hausdorff (Felix), 1868-1942, Allemand, 17, 18
- Heine (Eduard), 1821-1881, Allemand, 33
- Hellinger (Ernst David), 1883-1950, Allemand, 114, 144, 164, 172, 178, 181
- Helmholtz (Hermann Ludwig Ferdinand von -), 1821-1894, Allemand, 117, 265, 268, 269, 280, 294
- Hencky (Heinrich), 1885-1951, Allemand, 307
- Henstock (Ralph), 1923-2007, Anglais, 48
- Hermite (Charles), 1822-1901, Français, 20, 21, 60, 61, 97, 153, 158, 159, 171, 345, 374, 375, 379, 380, 387
- Héron d'Alexandrie, 1er siècle après JC, Grec, 218

- Hertz (Heinrich Rudolf), 1857-1894, Allemand, 313, 314
- Hilbert (David), 1862-1943, Allemand, 20, 21, 55, 57, 60, 61, 64, 66, 69, 84, 120, 249, 374, 375
- Hill (Rodney), 1921-2011, Anglais, 166, 309
- Hölder (Otto Ludwig), 1859-1937, Allemand, 35, 51
- Hoff (Nicholas J.), 1906-1997, Hongrois, 304
- Hofmann (Friedrich Carl Albert dit Fritz), 1866-1956, Allemand, 309
- Hooke (Robert), 1635-1703, Anglais, 89, 113, 143, 146, 164, 298, 299, 303-305
- L'Hôpital (Guillaume François Antoine de -), 1661-1704, Français, 218
- Howard (Ronald William), 1954- , Américain, 83
- Hrennikoff (Alexander), 1896-1984, Russe, 130
- Hu (Haichang), 1928-2011, Chinois, 113
- Huber (Tytus Maksymilian), 1872-1950, Polonais, 307
- Hugo (Victor), 1802-1885, Français, 80
- Huygens [ou Huyghens] (Christian), 1629-1695, Néerlandais, 218, 248, 264
- I**
- Inglis (Charles), 1875-1952, Anglais, 325
- Irwin (George Rankine), 1907-1998, Américain, 326-328, 331, 332
- Itō (Kiyoshi), 1915-2008, Japonais, 345
- J**
- Jacobi (Carl Gustav Jakob), 1804-1851, Allemand, 167, 218
- Johnson (A.I.), ?, Suédois, 353
- Jones (William, Sir -), 1675-1749, Gallois, 399
- Jordan (Marie Ennemond Camille), 1838-1922, Français, 249
- K**
- Kac (Mark ou Marek), 1914-1984, Américain, 344
- Kantorovitch (Leonid Vitalievitch), 1912-1986, Russe, 218
- Karadeniz (Halil), 1946-, Turquie, 353
- Karhunen (Kari), 1915-1992, Finlandais, 344, 345, 347-349
- Kelvin (William Thomson, connu sous le nom de Lord -), 1924-1907, Anglais, 306
- Kepler (Johannes), 1571-1630, Allemand, 371
- Kirchhoff (Gustav Robert), 1824-1887, Allemand, 146, 281, 300, 302
- Kirsch, 325
- Kjellman (Walter), 1905-1955, Suédois, 352
- Kolmogorov (Andreï Nikolaïevitch), 1903-1987, Russe, 23, 340
- Korn (Arthur), 1870-1945, Allemand, 67, 68
- Krook (Max), 1913-1985, Américain, 360, 361
- Krylov (Aleksey Nikolaevich), 1863-1945, Russe, 57
- Kuratowski (Kazimierz), 1896-1980, Polonais, 17
- Kurzweil (Jaroslav), 1926-, Tchèque, 48
- Kutta (Martin Wilhelm), 1867-1944, Allemand, 396, 397
- L**
- Ladevèze (Pierre), ?- , Français, 259
- Ladyjenskaïa (Olga Aleksandrovna), 1922-2007, Russe, 100
- Lagrange (Joseph Louis, comte de -), 1736-1813, Italien, 27, 29, 76, 77, 92, 97, 102, 103, 114, 142, 144, 145, 153, 159, 163, 164, 174, 176, 197, 218, 222, 260, 301, 302, 307, 314, 340, 372-374, 376, 380, 381, 386
- Laguerre (Edmond Nicolas), 1834-1886, Français, 378, 379, 387
- Lambert (Johann Heinrich), 1728-1777, Suisse, 372
- Lamé (Gabriel), 1795-1870, Français, 90, 303
- Landau (Lev Davidovitch), 1908-1968, Russe, 85
- Laplace (Pierre Simon de -), 1749-1827, Français, 79, 105-108, 307
- Laursen (Tod A), ?- , Américain, 314
- Lax (Peter), 1926-, Américain, 98, 120, 135, 211, 403, 405
- Lebesgue (Henri-Léon), 1875-1941, Français, 23-25, 46-48, 55, 70, 249
- Lee (Shaw M.), ?- , Américain, 338
- Legendre (Adrien-Marie), 1752-1833, Français, 218, 307, 371, 375, 376, 387
- Légrand (Mathias), ?- , Français, 4
- Leibniz (Gottfried Wilhelm), 1646-1716, Allemand, 27, 33, 36, 75, 218
- Leissa (Arthur W.), 1932-, Américain, 258
- Lemaire (Christophe), ?- , Français, 208
- Lemaire (Maurice), ?- , Français, 353
- Leray (Jean), 1906-1998, Français, 57
- Lévy (Paul Pierre), 1886-1971, Français, 345
- Lewy (Hans), 1904-1988, Américain, 284
- Lind (Niels Christian), 354
- Lind (Niels Christian), 1930-, Canadien, 353
- Liouville (Joseph), 1809-1882, Français, 76
- Lipschitz (Rudolph Otto Sigismund), 1832-1903, Allemand, 35, 76, 395
- Listing (Johann Benedict), 1808-1882, Allemand, 17
- Loève (Michel), 1907-1979, Français, 344, 345, 347-349
- Lorentz (Hendrik Antoon), 1853-1928, Néerlandais, 209
- Lotka (Alfred James), 1880-1949, Américain, 76
- Love (Augustus Edward Hough), 1863-1940, Anglais, 146
- M**
- Markov (Andreï Andreïevitch), 1856-1922, Russe, 371
- Martin (William T.), 1911-2004, Américain, 345
- Maxwell (James Clerk), 1831-1879, Écossais, 130, 209, 304-307
- Mayer (Max), ?, Allemand, 352
- Melenk (Jens Markus), ?- , Autrichien, 359

- Metropolis (Nicholas Constantine), 1915-1999, Américain, 350
- Meyers (Norman George), 1930-, Américain, 60
- Milgram (Arthur Norton), 1912-1961, Américain, 98, 120, 135, 211, 403
- Mindlin (Raymond David), 1906-1987, Américain, 146
- Miner (M.A.), 335, 336
- Minkowski (Hermann), 1864-1909, Allemand, 51
- Mises (Richard Edler, von -), 1883-1953, Autrichien, 166, 299, 307, 308
- Möbius (August (us) Ferdinand), 1790-1868, Allemand, 383
- Mohr (Christian Otto), 1835-1918, Allemand, 130, 308
- Montel (Paul Antoine Aristide Montel), 1876-1975, Français, 58
- Mooney (Melvin), 1893-1968, Américain, 310
- Moser (Jürgen Kurt), 1928-1999, Allemand, 340
- Moses (Fred), ?-, Américain, 353
- Mossotti (Ottaviano-Fabrizio), 1791-1863, Italien, 209
- Mullins (Leonard), 1918-1997, Anglais, 309
- Munson (Wilden A.), 1902-1982, Américain, 274
- N**
- Nash (John Forbes), 1928-, Américain, 82, 83
- Nataf (André), ?, Français, 354
- Navier (Claude Louis Marie Henri), 1785-1836, Français, 57, 83, 84, 110, 307, 360, 361
- Neper [ou Napier] (John, Baron de Merchiston), 1550-1617, Écossais, 264
- Neumann (Carl Gottfried), 1832-1925, Allemand, 65, 77, 107, 108, 111, 113, 117, 119, 142, 215
- Neumann (John, von -), 1903-1957, Hongrois, 20, 256, 347, 348
- Newmark (Nathan Mortimore), 1910-1981, Américain, 130, 239-241, 260, 391, 395, 398
- Newton (Isaac, Sir -), 1643-1727, Anglais, 36, 77, 142, 218, 303, 305, 316, 373, 381, 385, 386, 399-401
- Nguetseng (Gabriel), ?-, Camerounais, 210
- Nicolson (Phyllis), 1917-1968, Anglaise, 397
- Nirenberg (Louis), 1925-, Américain, 82
- Norton (Frederick H.), 304
- O**
- Ogden (Ray W.), ?-, Écossais, 310
- Oresme (Nicole), 1325-1382, Allemand, 36
- Ostrogradsky (Mikhaïl Vassilievitch), 1801-1862, Ukai-Sien, 44
- P**
- Padé (Henri Eugène), 1863-1953, Français, 372
- Páez (A.), ?, Espagnol, 353
- Palmgren (A), 335
- Panasenko (Grigory), ?-, Russe, 210
- Parent (Antoine), 1660-1726, Français, 146
- Paris (Paul C.), ?-, Américain, 336
- Perelman (Grigori Iakovlevitch), 1966-, Russe, 84
- Perron (Oskar), 1880-1975, Allemand, 48
- Pian (Theodore H.H.), 1919-2009, Américain, 114, 144, 164
- Piazzi (Giuseppe), 1746-1826, Italien, 371
- Piola (Gabrio), 1794-1850, Italien, 300, 302
- Poincaré (Henri), 1854-1912, Français, 17, 20, 48, 67, 71, 84, 208, 340
- Poisson (Siméon Denis), 1781-1840, Français, 79, 80, 90, 105-108, 209, 210, 214, 215, 303
- Pontriaguine (Lev Semionovitch), 1908-1988, Russe, 218
- Poritsky (H.), 314
- Prager (William), 1903-1980, Américain, 308
- Prandtl (Ludwig), 1875-1953, Allemand, 307
- R**
- Rackwitz (Rüdiger), 1941-2012, Allemand, 353
- Raphson (Joseph), 1648-1715, Anglais, 303, 316, 399-401
- Rayleigh (John William Strutt, troisième baron -), 1842-1919, Anglais, 130, 146, 256, 258, 280
- Reissner (Max Erich, dit Eric), 1913-1996, Américain, 114, 144, 146, 163, 164, 172, 178, 181
- Reuss, 307
- Reynolds (Osborne), 1842-1912, Irlandais, 85, 86, 111
- Riccati (Jacopo Francesco), 1676-1754, Italien, 75
- Ricci-Curbastro (Gregorio), 1853-1925, Italien, 84
- Rice (James R.), 1940-, Américain, 329, 331
- Richtmyer (Robert Davis), 1910-2003, Américain, 405
- Riemann (Georg Friedrich Bernhard), 1826-1866, Allemand, 44-47
- Riesz (Frigyes), 1880-1956, Hongrois, 55, 98, 211, 220
- Ritz (Walther), 1878-1909, Suisse, 130, 258, 357
- Rivlin (Ronald Samuel), 1915-2005, Américain, 310
- Robin (Victor Gustave), 1855-1897, Français, 78, 117
- Rodrigues (Benjamin-Olinde), 1795-1851, Français, 378
- Rosenblatt (Murray), 1926-, Américain, 354
- Rosenblueth (Emilio), 1926-1994, Mexicain, 353
- Runge (Carl David Tolmé), 1856-1927, Allemand, 372, 380, 387, 396, 397
- Sabine (Wallace Clement), 1868-1919, Américain, 270, 276
- Saint-Venant (Adhémar Jean Claude Barré de -), 1797-1886, Français, 147, 193, 304
- San Martín (Jorge), ?-, Chilien, 132
- Schrödinger (Erwin Rudolf Josef Alexander), 1887-1961, Autrichien, 378
- Schur (Issai), 1875-1941, Russe, 173, 174, 260

- Schwartz (Laurent), 1915-2002, Français, 55, 56
 Schwarz (Hermann Amandus), 1843-1921, Allemand, 39, 69
 Serrin (James Burton), 1926-2012, Américain, 60
 Sezawa (Katsutada), 1895-1944, Japonais, 258
 Shinozuka (Masanobu), 1930-, Japonais, 353
 Siegert (Arnold John Frederick), 1911-1995, Américain, 344
 Sierpiński (Wacław Franciszek), 1882-1969, Polonais, 25
 Signorini (Antonio), 1888-1963, Italien, 315
 Simo (Juan-Carlos), 1952-1994, Espagnol, 314
 Simpson (Thomas), 1710-1761, Anglais, 386, 399
 Snell (Snell van Royen (Willebrord), dit Willebrordus Snellius), 1581-1626, Néerlandais, 264
 Snow (John), 1813-1858, Anglais, 204
 Sobolev (Sergueï Lvovitch), 1908-1989, Russe, 55, 57, 60, 63, 64, 66
 Soize (Christian), 1948-, Français, 262
 Sommerfeld (Arnold Johannes Wilhelm), 1868-1951, Allemand, 357
 Spanos (Pol D.), ?-, Grec, 350
 Stevens (Stanley Smith), 1906-1973, Américain, 273
 Stokes (George Gabriel), 1819-1903, Anglais, 57, 83-85, 109, 110, 307, 360, 361
 Stone (Marshall Harvey), 1903-1989, Américain, 373
 Strang (William Gilbert), 1934-, Américain, 130, 137, 138, 403
 Sturm (Jacques Charles François), 1803-1855, Français, 76
 Sylvester (James Joseph), 1814-1897, Anglais, 249
- T**
 Taylor (Brook), 1685-1731, Anglais, 340, 344, 373, 380, 398, 400
 Tchebychev (Pafnouti Lvovitch), 1821-1894, Russe, 376-378, 380, 387
 Thiessen (Alfred H.), 1872-1956, Américain, 204
 Timoshenko (Stephen), 1878-1972, Russe, 146
 Tisserand (François Félix), 1845-1896, Français, 340
 Tong (Pin), ? , Chinois, 114, 144, 164
 Torroja (Eduardo), ? , Espagnol, 353
 Toscani (Giuseppe), ?- , Italien, 85
 Trefftz (Erich Immanuel), 1888-1937, Allemand, 263, 357
 Tresca (Henri Édouard), 1814-1885, Français, 307, 308, 332
 Tsai (Stephen W), ?-, Américain, 166, 309
 Tschirnhaus (Ehrenfried Walther von -), 1651-1708, Allemand, 218
 Tsotsis (Thomas K.), ?- , Américain, 338
- U**
 Uflyand (Yakov Solomonovic), Russe, 146
 Ulam (Stanisław Marcin), 1909-1984, Américain, 351
 Uzawa (Hirofumi), 1928-, Japonais, 316
- V**
 Villani (Cédric), 1973- , Français, 85
 Vinci (Léonard, de-), 1452-1519, Italien, 146
 Voigt (Woldemar), 1850-1919, Allemand, 214, 215, 298, 304-306
 Volterra (Vito), 1860-1940, Italien, 76, 345
 Voronoï (Gueorgui Feodossievitch), 1868-1908, Russe, 201, 203, 204
- W**
 Wallis (John), 1616-1703, Anglais, 399
 Warburton (Geoffrey Barratt), 1924-2009, Anglais, 258
 Waring (Edward), 1736-1798, Anglais, 372
 Warnock (John Edward), 1940- , Américain, 381
 Washizu (Kyuichiro), 1921-1981, Japonais, 113
 Wästlund (Karl Georg), 1905-1980, Suédois, 352
 Weber (Ernst Heinrich), 1795-1878, Allemand, 273
 Weibull (Ernst Hjalmar Waloddi), 1887-1979, Suédois, 337, 352
 Weierstrass (Karl Theodor Wilhelm), 1815-1897, Allemand, 218, 373
 Weill (Jean-Christophe), ?-, Français, 208
 Wells (Alan Arthur), 1924-2005, Anglais, 332
 Wendroff (Burton), 1930-, Américain, 405
 Wheatstone (Charles), 1802-1875, Anglais, 258
 Wiener (Norbert), 1894-1964, Américain, 345
 Wilson (John), 1741-1793, Anglais, 103
 Winther (Ragnar), 1949-, Norvégien, 361
 Wirtinger (Wilhelm), 1865-1945, Autrichien, 67
 Wisner (Alain), 1923-2004, Français, 277
 Wöhler (August), 1819-1914, Allemand, 335
 Wu (Edward Ming-Chi), 1938-2009, Américain, 166
- Y**
 Young (Thomas), 1773-1829, Anglais, 90, 169, 303, 310
 Young (William Henry), 1863-1942, Anglais, 373, 380
- Z**
 Zach (Franz Xaver, Baron von -), 1754-1832, Allemand, 371
 Zienkiewicz (Olgierd Cecil), 1921-2009, Anglais, 130

Annexe G

Index des concepts

A

analyse ondulatoire de l'énergie 263
 ansatz 212, 213, 215, 216
 application 21, 28
 approche ascendante 191
 approche descendante 191
 approximation
 conforme 134, 135, 138
 non conforme 134, 137, 138

B

bijection 28
 borne
 de Reuss 214, 215
 de Voigt 214, 215

C

chaos polynomial 345
 coefficients de Lamé 90, 303
 coefficients de Poisson 90, 214, 303
 compacité 18
 complément de Schur 173, 174, 260
 Component Modal Synthesis 262
 condensation statique 195
 condition aux limites 77
 de Dirichlet ... 65, 77, 105, 106, 108–111, 113, 117, 210, 215
 de Neumann .. 65, 77, 107, 108, 111, 113, 117, 215
 de Robin 78, 117
 dynamique 78
 essentielle 107
 mêlée 78, 107, 108
 naturelle 107
 condition BBL 100
 condition de périodicité 192
 condition inf-sup 100, 123
 continuité
 C^0 33
 C^k 37
 Hölder 35
 Lipschitz 35
 contrainte de von Mises 299
 contrainte effective 299
 couplage des cellules microscopiques 192, 193

courbe de Wöhler (ou S-N) 335, 337
 critère de Cauchy 19
 critère de plasticité
 Drucker-Prager 308
 Hill 309
 Mohr-Coulomb 308
 Tresca 308
 Tsai 309
 von Mises 307, 308
 crochet de dualité 31

D

d'Alembertien 40
 décomposition de domaine 193, 194
 décomposition modale 244, 254, 255
 densité spectrale 248
 dérivée
 au sens de Fréchet 220
 au sens des distributions 58
 classique 35
 directionnelle 220
 d'ordre supérieur 198, 199
 généralisée 70
 normale 41, 61
 par rapport à la géométrie 199
 partielle 37
 diagramme de Voronoï 203
 différentielle 37, 220
 distance 19
 issue d'une norme 20, 69
 distribution 56
 de Dirac 58
 divergence 40, 44
 domaine
 fréquentiel 259, 261
 temporel 260
 dual d'un espace vectoriel 30
 dual topologique d'un espace vectoriel ... 31, 58, 62

E

ED-EDP
 de Faraday 89
 de Helmholtz 117
 de la chaleur 80, 89, 108, 249, 252
 de la déformation d'une membrane 80

de la diffusion moléculaire	89	propre	249, 261
de l'acoustique	80, 92, 116	régulier	18
de Laplace	79, 105–108	séparé	18
de l'élasticité plane	89	topologique	18
de Navier-Stokes	83, 110	trace	61–63, 71
de Poisson	79, 105–108, 210, 215	vectorel	
de Stokes	85, 109	normé	20, 69
des ondes	80	topologique	20, 61
d'Euler	86, 111	exposant conjugué	
d'Ohm	89	de Lebesgue	49
dynamique de population	76	de Sobolev	66
Lotka-Volterra	76	F	
relation fondamentale de la dynamique . 77, 89,		facteur de forme du maillage	133, 202
111, 113, 239, 250, 252–254		facteur d'intensité de contrainte K	328
système proie-prédateur	76	fonction test	70
effectif	209, 213	fonctionnelle	
élément fini		de Hu-Washizu	113
de Lagrange	153, 154, 159	de l'énergie complémentaire	113
de référence	156	de l'énergie potentielle totale	112
d'Hermite	158, 159, 171	de Pian et Tong	114
diamètre d'un	133, 202	d'Hellinger-Reissner . 114, 144, 164, 172, 178,	
enrichi	262	181	
famille affine d'	156	duale	113
frontière	262, 357	hybride	114, 144
hiérarchique	262	mixte	114, 144, 172, 178, 181
rondeur d'un	133, 202	primale	112
singulier	333	forme	
stochastique	346, 350	bilinéaire	21, 69
énergie potentielle	97	linéaire	30, 58
ensemble mesurable	23	formulation faible	95
équation		fréquence propre	247–249, 251, 252, 255, 261
d'Euler	221	FRF	261
équations		G	
de Laguerre	378	G-FEM	262
de Legendre	375	gradient	40, 43
de Navier-Stokes	360, 361	H	
espace		homogénéisation	192, 209, 210, 213
compact	18	I	
de Banach	20, 60, 69, 70	image	31
de Hausdorff	18	inégalité	
de Hilbert	20, 60, 61, 64, 66, 69, 374, 375	de Cauchy-Schwarz	69
de Lebesgue	48, 70, 375	de Korn	67
de Sobolev		de Poincaré	67, 71
$H(\text{div})$	66	de Poincaré-Friedrichs	67
$H^1(\Omega)$	64	de Poincaré-Wirtinger	67
$H^{-1}(\Omega)$	64	injection	28
$H^m(\Omega)$	60, 71	intégrale J	329, 331
$H^{-m}(\Omega)$	60, 61	intégration	
$H_0^1(\Omega)$	64, 71	Newton-Cotes	385, 386
$H_0^m(\Omega)$	60, 61	quadrature de Gauß	387
$W^{m,p}(\Omega)$	60	rectangles	385
d'énergie	64	Simpson	386
euclidien	69	trapèzes	385
mesurable (probabilisable)	22		
mesuré	23, 24		
métrique	19		
complet	19, 69		
préhilbertien	69		

J	
jacobien	167, 168
jacobienne	167, 168
L	
lagrangien	221
laplacien	40, 79, 105–108
lemme de	
Céa	135–137
Strang	137, 138
loi	
de Paris	336
de Weibull	337
des mélanges	214, 215
loi de comportement	89, 111, 113
anisotrope	312
auxétique	91
Bingham-Norton	304
élastomères	309
Hooke	89, 113, 164, 303–305
hyperélasticité	310
incompressible	91, 164
Kelvin-Voigt	306
Maxwell	304–306
Newton	305
Norton-Hoff	304
plasticité	306
Rivlin généralisé	310
Saint-Venant	304
viscoélasticité	305
viscoplasticité	305
Voigt	304, 305
M	
macro-élément	195
maillage de Delaunay-Voronoi	203, 204
mesure	23
complète	24
de Dirac	24
de Lebesgue	24, 25
régulière	25
méthode de développement asymptotique infini	213
méthode de développement régulier	212
méthode de diffusion d'énergie	263
méthode de Galerkin	
discontinue	261
avec multiplicateurs de Lagrange	262
Gradient Least Squares	262
Least Squares	262
sans maillage	262
méthode de la couche limite	212
méthode de Lattice-Boltzmann	360
méthode de Monte-Carlo	350, 355
méthode de Newton-Raphson	303, 316, 399–401
méthode de Runge-Kutta	
ordre 1	396
ordre 2	396
ordre 4	397
méthode de Trefftz	263
méthode d'enrichissement discontinu	262
méthode des éléments discontinus	262
méthode des éléments finis étendue	262, 334, 359, 360
méthode des rayons	264
méthode d'Euler	396
méthode h	177, 262
méthode hp	177
méthode p	177, 262
méthode particulière	358
méthode r	177
méthode sans maillage	262, 334, 358, 359
méthode spectrale	250
méthodes Énergétiques Simplifiées	263
métrique	19
mode propre	247–249, 251, 255, 261–263
modèle macroscopique	191–193, 209
modèle microscopique	191–194, 209
module de Coulomb	303, 310
module d'Young	90, 169, 303, 310
morphisme	29
automorphisme	30
endomorphisme	30
isomorphisme	30
multi-échelles	190, 191, 194
multi-niveaux	192, 193
multiplicateurs de Lagrange	92, 102, 103, 114, 163, 164, 174, 176, 197, 222, 260, 314
N	
normale	41
norme	20, 69, 70
induite par un produit scalaire	69, 71, 376
sur $H(\text{rot})$	66
sur $H^m(\Omega)$	71
sur $H^{-m}(\Omega)$	61
sur $W^{m,p}(\Omega)$	60
notation de Voigt	298
noyau	31
O	
opérateur	32
optimisation	217
oscillateur	247
P	
partition de l'unité	262, 359
patch-test	164, 165
perturbation	196, 199
polynôme	
caractéristique	249, 250
de Bernstein	379, 381
de Lagrange	374, 380, 381
de Laguerre	378, 379, 387
de Legendre	375, 376, 387
de Tchebychev	376–378, 387
d'Hermite	342, 345, 349, 350, 374, 375, 379, 380, 387

- orthogonalité 342, 373, 375–378, 387
- post-traitement 163, 164, 177
- principe
 - de d'Alembert 97
 - de Saint-Venant 147, 193
 - des puissances virtuelles 96
 - des travaux virtuels 96
- produit scalaire 20, 21, 69, 374–378
 - de $H(\text{div})$ 66
 - de $H^1(\Omega)$ 64
 - de $H^m(\Omega)$ 60, 71
 - de L^2 51, 70, 375, 376
- propriété de la boule ouverte 204
- propriété d'orthogonalité 255
- pseudo-inversion 195, 196, 198
- Q**
- quotient de Rayleigh 256
- R**
- représentation modale 248
- résidus pondérés 97
- résonance 247, 248
- rotationnel 40, 43
- S**
- schéma
 - de Newmark 240, 241, 260, 398
 - des différences finies centrées .. 240, 241, 261, 398
- SEA 259, 263
- série de Karhunen-Loève 344, 345, 347–350
- série de von Neumann 347, 348
- sous-structuration 190, 194, 195
- sous-structure 194, 195
- super-élément 194, 195
- support 24, 29
- surjection 28
- synchronisation 248, 249
- T**
- taux de restitution d'énergie G 327, 329, 331
- tenseur
 - des contraintes de Cauchy 298
 - des contraintes de Kirchhoff 300
 - des contraintes de Piola-Kirchhoff ... 300, 302
 - des déformations de Green-Lagrange . 301, 302
 - gradient de déformation 300
 - linéarisé des déformations 301
- théorème
 - de Babuška 100, 136
 - de Borel-Lebesgue 25
 - de Brezzi 100, 121–123, 137
 - généralisé 101
 - de Cauchy 298
 - de Cayley-Hamilton 250
 - de changement de variables dans les intégrales multiples 168
 - de complétion (sur un Banach) 21
 - de densité 64
 - de dérivation des fonctions composées dans les espaces de Sobolev 65
 - de Dirichlet 81
 - de Green-Ostrogradsky 44
 - de Green-Riemann 44
 - de Kirchhoff 281
 - de Lax 405
 - de Lax-Milgram 98, 120, 135, 211
 - de Lax-Wendroff 405
 - de Meyers-Serrin 60
 - de projection (dans un Hilbert) 21, 51
 - de représentation de Riesz-Fréchet 98, 211, 220
 - de Stone-Weierstrass 373
 - de Taylor 373
 - de Weierstrass 373
 - d'injections continues de Sobolev 66
 - d'inversion locale 168
 - du flux-divergence 44
 - du gradient 43
 - du rang 32
 - du rotationnel 43
- théorie des structures floues 262
- Théorie Variationnelle des Rayons Complexes . 259
- topologie 18, 31
- trace 178, 193
- transformée de Fourier 259, 261, 264
- tribu
 - borélienne 23
- TXFEM 261
- U**
- unisolvance 153
- V**
- valeur propre voir fréquence propre
- vecteur propre voir mode propre
- vibration
 - libre 250, 252
 - périodique 250, 253
 - transitoire 250, 254
- volume élémentaire représentatif 209, 213
- W**
- WBT 263
- X**
- X-FEM 262, 334, 359, 360

Annexe H

Références bibliographiques

- [1] Grégoire ALLAIRE. « Homogeneization and two-scale convergence ». In : *SIAM J. Math. Anal.* 23 (1992), pages 1482–1518. URL : <http://www.cmap.polytechnique.fr/~allaire/two-scale.pdf> (cf. page 213).
- [2] Grégoire ALLAIRE. *Analyse numérique et optimisation*. Eds de l'École Polytechnique, 2007.
- [3] Grégoire ALLAIRE. *Optimal design of structures*. École Polytechnique, Cours MAP 562. 2014. URL : http://www.cmap.polytechnique.fr/~allaire/course_map562.html.
- [4] Grégoire ALLAIRE. *Boîte à outils FreeFem++ pour l'optimisation de formes (géométrie et topologie)*. École Polytechnique, CMAP. Janvier 2008. URL : <http://www.cmap.polytechnique.fr/~allaire/freefem.html> (cf. pages 228, 229).
- [5] Grégoire ALLAIRE, Stéphane GAUBERT et Olivier PANTZ. *Exercices corrigés - Analyse numérique et optimisation - Une introduction à la modélisation mathématique et à la simulation numérique 2011*. École Polytechnique, CMAP. 2011. URL : <http://www.cmap.polytechnique.fr/~allaire/map431/corrections.pdf>.
- [6] Pierre ALLIEZ et al. *Anisotropic Polygonal Remeshing*. INRIA. URL : <http://hal.inria.fr/docs/00/07/17/78/PDF/RR-4808.pdf> (cf. page 201).
- [7] Douglas N. ARNOLD, Richard S. FALK et Ragnar WINTHER. « Finite element exterior calculus, homological techniques, and applications ». In : *Acta Numerica* (2006), pages 1–155. URL : <http://www.ima.umn.edu/~arnold/papers/acta.pdf> (cf. page 361).
- [8] Douglas N. ARNOLD, Richard S. FALK et Ragnar WINTHER. « Finite element exterior calculus : from Hodge theory to numerical stability ». In : *Bulletin of AMS* 47.2 (2010), pages 281–354. URL : <http://www.ima.umn.edu/~arnold/papers/bulletin.pdf> (cf. page 361).
- [9] Nadia BAHLOULI. *Cours Matériaux Composites*. Univ. Strasbourg. URL : <http://www-ipst.u-strasbg.fr/cours/materiaux-composites/tm.htm>.
- [10] Stéphanie BASSEVILLE. *Méthodes de résolutions en éléments finis*. École Nationale Supérieure des Mines de Paris, Enseignement spécialisé « Éléments finis » S3733-S3735. 2005-2006. URL : http://mms2.ensmp.fr/ef_paris/technologie/polycop/f_coursMR.pdf.
- [11] Jean-Louis BATOZ et Gouri DHATT. *Modélisation des structures par éléments finis. Vol. 1 – solides élastiques*. Hermes, 1995.
- [12] Marc BERVEILLER. « Éléments finis stochastiques : approches intrusive et non intrusive pour des analyses de fiabilité ». Thèse de doctorat. Univ. Blaise Pascal - Clermont II, 2005. URL : <http://tel.archives-ouvertes.fr/docs/00/36/62/25/PDF/Berveiller.pdf>.

- [13] Éric BLAYO. *Notes de cours sur la méthode des éléments finis*. M1 Mathématiques Appliquées et Industrielles, Laboratoire Jean Kuntzmann, INRIA. 2010. URL : <http://www-ljk.imag.fr/membres/Eric.Blayo/Documents/Poly-EF.pdf>.
- [14] Éric BLAYO. *Notes de cours sur la méthode des éléments finis*. M1 Mathématiques Appliquées et Industrielles, Laboratoire Jean Kuntzmann, INRIA. 2010. URL : <http://www-ljk.imag.fr/membres/Eric.Blayo/Documents/Poly-EF.pdf>.
- [15] Éric BLAYO. *Notes de cours sur la méthode des éléments finis*. M1 Mathématiques Appliquées et Industrielles, Laboratoire Jean Kuntzmann, INRIA. 2010. URL : <http://www-ljk.imag.fr/membres/Eric.Blayo/Documents/Poly-EF.pdf>.
- [16] Franck BOYER. *Analyse numérique des EDP elliptiques*. Aix-Marseille universités, M2 Mathématiques et Applications. 2011. URL : http://www.cmi.univ-mrs.fr/~fboyer/_media/enseignements/cours_m2_anedp_fboyer.pdf.
- [17] Franco BREZZI. « On the existence, uniqueness and approximation of saddle-point problems arising from lagrangian multipliers ». In : *RAIRO* 2 (1974), pages 129–151. URL : http://archive.numdam.org/ARCHIVE/M2AN/M2AN_1974__8_2/M2AN_1974__8_2_129_0/M2AN_1974__8_2_129_0.pdf.
- [18] Stéphane CANU. *Introduction à l'optimisation*. INSA de Rouen, Cours ASI 4. 5 novembre 2013. URL : https://moodle.insa-rouen.fr/pluginfile.php/18601/mod_resource/content/3/Canu/Slides_01_intro.pdf.
- [19] Laurent CHAMPANEY. *Les éléments finis par l'exemple : modèles 1D, 2D ou 3D ?* ENS Cachan ; Département de génie mécanique, Laboratoire de mécanique et technologie. URL : http://www.si.ens-cachan.fr/accueil_V2.php?page=affiche_ressource&id=173.
- [20] Laurent CHAMPANEY. *Problèmes spécifiques à la méthode des éléments finis*. ENS Cachan ; Département génie mécanique, Laboratoire de mécanique et technologie. URL : <http://laurent.champaney.free.fr/perso/cours/EF-PbSpecifiques-Champaney.pdf>.
- [21] Laurent CHAMPANEY. *Simulation par éléments finis - attention aux singularités*. ENS Cachan. URL : http://www.si.ens-cachan.fr/accueil_V2.php?page=affiche_ressource&id=175.
- [22] Stéphane CHAUSSEDENT. *Mécanique des milieux continus*. NUMEPHY, Univ. du Maine, Le Mans. URL : http://res-nlp.univ-lemans.fr/NLP_C_M03_G01/co/NLP_C_M03_G01_web.html.
- [23] Shiyi CHEN, Gary D. DOOLEN et Kenneth G. EGGERT. « Lattice-Boltzmann Fluid Dynamics : a versatile tool for multiphase and other complicated flows ». In : *Los Alamos Science Number* 22 (1994). URL : <http://www.fas.org/sgp/othergov/doe/lanl/pubs/00285550.pdf>.
- [24] Mathilde CHEVREUIL. « Sur une nouvelle approche en calcul dynamique transitoire, incluant les basses et les moyennes fréquences ». Thèse de doctorat. ENS Cachan, 2005. URL : <http://hal.archives-ouvertes.fr/docs/00/05/24/41/PDF/Chevreuil.pdf> (cf. page 259).
- [25] ? « Chapter 9 : Fundamentals of Structural Reliability ». In : *Lectures on Structural Reliability 1980. One-day Conference on Structural Reliability, Aalborg University Centre 1980. Proceedings*. Sous la direction de P.T. CHRISTENSEN, Ålborg Universitetscenter. Instituttet for BYGNINGSTEKNIK et AUC. Institute of Building Technology et Structural Engineering, Aalborg University Centre, AUC : Conference on Structural Reliability 1980, 1980. URL : http://www.google.fr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&ved=0CDEQFjAA&url=http://vbn.aau.dk/files/17204972/Chapter_9_

Fundamentals_of_Structural_Reliability&ei=_40EUvbOKI6T0QWLt4HICA&usg=AFQjCNF_pNM7iDYWfoi8jrD0SNQEEZ0mJg&sig2=4vorKeIJfiErvHef5QhJgw&bvm=bv.50500085,d.d2k.

- [26] Ray W. CLOUGH et Edward L. WILSON. *Early Finite Element Research at Berkeley*. URL : http://www.ce.memphis.edu/7111/notes/class_notes/papers/fe-history.pdf (cf. page 130).
- [27] Centre de Mathématiques Appliquées (CMAP). *Exemples avec Freefem++*. École Polytechnique. URL : <http://www.cmap.polytechnique.fr/>.
- [28] Didier COMBESURE. *Éléments de dynamique des structures. Illustrations à l'aide de CAST3M*. Site de Cast3M. septembre 2006. URL : <http://www-cast3m.cea.fr/html/Combescure/DynamiqueCAST3MSeptembre2006.pdf>.
- [29] Richard COURANT, KURT et Hans LEWY. « Über die partiellen Differenzengleichungen der mathematischen Physik ». In : *Mathematische Annalen* 100.1 (1928), pages 32–74. URL : http://www.digizeitschriften.de/dms/img/?PPN=PPN235181684_0100&DMDID=dmdlog5&LOGID=log5&PHYSID=phys38#navi (cf. page 284).
- [30] Amy DAHAN-DALMEDICO et Jeanne PEIFFER. *Une histoire des mathématiques. Routes et dédales*. Eds du Seuil, 1986.
- [31] David DUREISSEIX. *Méthodes numériques appliquées à la conception par éléments finis*. 10 janvier 2009. URL : <http://hal.archives-ouvertes.fr/docs/00/35/17/13/PDF/polyMNb.pdf>.
- [32] Michel DYSLI. *B3-5 Introduction aux éléments finis*. École Polytechnique Fédérale de Lausanne ; cycle postgrade : géologie appliquée à l'ingénierie et à l'environnement. URL : http://lms.epfl.ch/files/content/sites/lms/files/shared/Publications/Soil%20and%20Rock/Intro_aux_elements_finis_GAIE.pdf.
- [33] Bernard FAVRE. « Méthode pour établir la relation entre le bruit émis par les véhicules routiers et l'impact de ce bruit sur les populations exposées ». Thèse de doctorat. Université Claude Bernard Lyon I, 1984 (cf. page 266).
- [34] Zhi-Qiang FENG. *Mécanique non linéaire M2-MN91*. UFR Science et Technologies, Univ. d'Evry Val d'Essonne. URL : <http://lmee.univ-evry.fr/~feng/MN91.pdf>.
- [35] Pascal FREY et Frédéric HECHT. *Cours Maillage 2D, surfacique et 3D*. Université Pierre et Marie Curie (Paris 6). URL : <https://www.ljll.math.upmc.fr/~hecht/ftp/old/IST/NoteCours-v1.pdf>.
- [36] Roger R. GHANEM et Pol D. SPANOS. *Stochastic Finite Elements. A spectral approach*. Dover, 2005.
- [37] Gene H. GOLUB et Charles F. van LOAN. *Matrix Computations, 3e Ed*. Johns Hopkins University Press, 1996.
- [38] K.K. GUPTA et J.L. MEEK. « A bried history of the beginning of the Finite Element Method ». In : *Int. J. Num Meth in Eng.* 39 (1996), pages 3761–3774. URL : http://people.sc.fsu.edu/~jpeterson/history_fem.pdf (cf. page 130).
- [39] Jean-François HAMET. *Les mécanismes de génération du bruit de roulement et l'influence des caractéristiques de chaussée*. URL : http://www.infobruit.com/revues/78_09956.PDF (cf. page 266).
- [40] Jean-François HAMET. *Modélisation du bruit de contact pneu / chaussée. Quels sont les facteurs clés d'une réduction des nuisances sonores ?* URL : <http://www.inrets.fr/ur/lte/publi-autresactions/lesposters/contactpneuchauss.pdf> (cf. page 266).

- [41] Bertrand HAUCHECORNE. *Les contre-exemples en mathématiques : 522 Contre-exemples (2ème Ed)*. Ellipses, 2007.
- [42] Frédéric HECHT. « New development in freefem++ ». In : *J. Numer. Math.* 20.3-4 (2012), pages 251–265. ISSN : 1570-2820.
- [43] Frédéric HESCHUNG. « Vers une construction et une adaptation automatiques des maillages ». Thèse de doctorat. Université de Reims Champagne Ardenne, 26 mai 2005. URL : <http://ebureau.univ-reims.fr/slide/files/quotas/SCD/theses/exl-doc/GED00000155.pdf>.
- [44] Franck HETROY. *Paramétrisation de maillages*. Université de Grenoble. URL : <https://www.ljll.math.upmc.fr/~hecht/ftp/old/IST/NoteCours-v1.pdf>.
- [45] Jean JACOD. *Théorie de la mesure et intégration (LM365)*. Université Pierre et Marie Curie (Paris 6), L3 Maths. 2010-11. URL : <http://www.proba.jussieu.fr/pageperso/amaury/LM364/jacod.pdf>.
- [46] Steven G. JOHNSON. *Notes on Perfectly Matched Layers (PMLs)*. URL : <http://math.mit.edu/~stevenj/18.369/pml.pdf>.
- [47] Patrick JOLY. *Introduction au calcul scientifique pour les EDP de la physique*. INRIA-Rocquencourt. URL : <http://pages.saclay.inria.fr/houssem.haddar/Cours/ENSTA/ma201-cours-01.pdf>.
- [48] Tosio KATO. *Perturbation Theory for Linear Operators*. Classics in Mathematics, Springer, 1995 (reprint de 1980).
- [49] Raphaël KRIKORIAN. *Simulations des phénomènes aléatoires (LM346)*. Université Pierre et Marie Curie (Paris 6), L3 Maths. URL : <http://www.proba.jussieu.fr/dw/lib/exe/fetch.php?media=users:laruelle:lm346poly.pdf> (cf. page 355).
- [50] *La simulation multi-échelles*. ENS Cachan. URL : http://www.si.ens-cachan.fr/accueil_V2.php?page=affiche_ressource&id=164.
- [51] Pierre LADEVÈZE. « A new computational approach for structure vibrations in the medium frequency range ». In : *C. R. Acad. Sci. Paris, Série II* 322, no 12 (1996), pages 849–856 (cf. page 259).
- [52] *Lagrange (documentaire)*. Institut Henri Poincaré. décembre 2013. URL : http://www.youtube.com/watch?v=K_tAygfZnAE.
- [53] M. LEBRUN. *Encapsulage acoustique du compartiment GMP*. Rapport technique. Rapport Insavalor, 1998 (cf. page 266).
- [54] Christophe LEMAIRE et Jean-Christophe WEILL. « Partitioning the sphere with constant area quadrangles ». In : *12th Canadian Conference on Computational Geometry*. 2000, pages 227–232 (cf. page 208).
- [55] Daniel LI. *Mesure et Intégration : Ch6 Espaces L^p* . Université d'Artois ; Faculté des Sciences Jean Perrin ; L3 Mathématiques-Informatique. 2011. URL : <http://li.perso.math.cnrs.fr/textes/Integration/espaceLp.pdf>.
- [56] R. LUCQUIAUD. *Évaluation de niveaux de bruit de poids lourds, Rapport no 90.15.73.801/1967*. Rapport technique. Union Technique de l'Automobile du Motocycle et du Cycle (UTAC), 1967 (cf. page 266).
- [57] Jean-François MAITRE. « Cours d'éléments finis ». In : *DEA Analyse numérique, EDP et calcul scientifique ; Univ Lyon 1 et St-Étienne, ECL, ENS Lyon* (1996).
- [58] Vincent MANET. *Mesure de la beauté des bruits périodiques*. marmonier. URL : http://hal.archives-ouvertes.fr/docs/00/72/26/81/PDF/VM_suavity-fr.pdf (cf. page 278).

- [59] Vincent MANET. « Méthodes d'analyse par éléments finis des contraintes aux interfaces dans les structures sandwich ». Thèse de doctorat. Univ. Blaise Pascal - Clermont II, 17 juillet 1998. URL : <http://tel.archives-ouvertes.fr/docs/00/66/30/78/PDF/THESE.pdf>.
- [60] Vincent MANET. « Cours d'option Choix des matériaux - Application à la vibro-acoustique ». In : *École Nationale Supérieure des Mines de St-Étienne* (2009).
- [61] Régis MARCHIANO. *Rayonnement Acoustique des Structures*. URL : <http://www.dalembert.upmc.fr/home/marchiano/images/stories/rasnb.pdf>.
- [62] J.S. San MARTIN et al. « An initial and boundary value problem modeling fish-like swimming ». In : *Arch. Ration. Mech. Anal.* 188.3 (2008), pages 429–455. URL : <http://www.iecn.u-nancy.fr/~scheid/Publication/fish.pdf> (cf. page 132).
- [63] Norman G. MEYERS et James SERRIN. « $H = W$ ». In : *Proc. Nat. Acad. Sci USA* 51 (1964), pages 1055–1056. URL : <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC300210/pdf/pnas00180-0073.pdf> (cf. page 60).
- [64] Hakim NACEUR. *Mécanique de la rupture*. ENSIAME - Univ. de Valenciennes et du Hainaut Cambrésis. 2011.
- [65] Hakim NACEUR. *Modélisation de la rupture en dynamique rapide*. Univ. de Valenciennes et du Hainaut Cambrésis. 2012.
- [66] J. NAUMANN. *Remarks on the Prehistory of Sobolev Spaces*. Humboldt Universität. 25 mars 2002. URL : <http://edoc.hu-berlin.de/series/mathematik-preprints/2002-2/PDF/2.pdf>.
- [67] Steve OUDOT. *Maillage de surfaces de genre zéro sans bord*. École Polytechnique. URL : <http://www.enseignement.polytechnique.fr/informatique/ARCHIVES/IF/projets/oudot/sujet.html>.
- [68] Grigory PANASENKO. « Cours d'homogénéisation ». In : *DEA Analyse numérique, EDP et calcul scientifique ; Univ Lyon 1 et St-Étienne, ECL, ENS Lyon* (1996).
- [69] Alexandre POPIER. *Espaces fonctionnels : un peu de topologie*. Université du Maine, Le Mans. URL : http://perso.univ-lemans.fr/~apopier/enseignement/M1_Maths_EDP/slides_espaces_fonct.pdf.
- [70] Jérôme POUSIN. « Cours estimateurs d'erreurs et remaillage ». In : *DEA Analyse numérique, EDP et calcul scientifique ; Univ Lyon 1 et St-Étienne, ECL, ENS Lyon* (1996).
- [71] Julien RÉTHORÉ. « Méthode éléments finis étendus en espace et en temps : application à la propagation dynamique des fissures ». Thèse de doctorat. INSA Lyon, 2005. URL : <http://tel.archives-ouvertes.fr/docs/00/12/11/72/PDF/these.pdf>.
- [72] François ROUVIÈRE. *Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation (2ème Ed)*. Cassini, 2003.
- [73] Patricia SAAD. « Modélisation et identification du comportement non linéaire des cales en caoutchouc ». Thèse de doctorat. École Centrale de Lyon, 2003. URL : <http://bibli.ec-lyon.fr/exl-doc/psaad.pdf>.
- [74] Ulf SANDBERG. « Low noise road surfaces. A state-of-the-art review ». In : *Eurosymposium Maîtrise du bruit routier en milieu urbain*. 1992. URL : http://www.asphaltrubber.org/ARTIC/International/RPA_A1177.pdf (cf. page 266).
- [75] Ulf SANDBERG. « Tyre/road noise : myths and realities ». In : *Plenary paper published in the Proceedings of The 2001 International Congress and Exhibition on Noise Control Engineering*. The Hague, The Netherlands, 2001, August 27-30. URL : <http://www.vti.se/sv/publikationer/pdf/tyreroad-noise--myths-and-realities.pdf> (cf. page 266).

- [76] Jean-François SCHEID. *méthodes numériques pour la dynamique des fluides*. Institut E.Cartan U.M.R. 7502, Univ. H. Poincaré Nancy 1. Cours de Master 2 IMOI option calculs scientifique 2011-2012. URL : <http://www.iecn.u-nancy.fr/~scheid/Enseignement/polyM2IMOI.pdf>.
- [77] Jean-François SCHEID. *Analyse numérique des équations de Navier-Stokes*. Institut E.Cartan U.M.R. 7502, Univ. H. Poincaré Nancy 1. Cours de Master 2 Mathématiques (Recherche) 2005-2006. URL : http://www.iecn.u-nancy.fr/~scheid/Enseignement/ns_main.pdf.
- [78] Laurent SCHWARTZ. *Analyse I : Théorie des ensembles et Topologie*. Hermann, 1992.
- [79] Laurent SCHWARTZ. *Analyse II : Calcul différentiel et équations différentielles*. Hermann, 1992.
- [80] Laurent SCHWARTZ. *Analyse III : Calcul intégral*. Hermann, 1992.
- [81] Christian SOIZE. « Modélisation probabiliste du flou structural en dynamique linéaire des système mécaniques complexes - Éléments théoriques. » In : *La Recherche Aéronautique* 5 (1986), pages 23–48 (cf. page 262).
- [82] Christian SOIZE. « Estimation of the fuzzy substructure model parameters using the mean power flow equation of the fuzzy structure ». In : *Journal of Vibration and Acoustics* 120 (1998), pages 279–286 (cf. page 262).
- [83] H. STEVEN. *Engine Torque and Speed Distribution of Commercial Vehicules in real traffic and its Consideration in Emission Measurement According to EC Directive 88/77/EEC resp. ECE regulation R49*. Rapport technique. Report no 104.05.316 - FiGE, 1994 (cf. page 266).
- [84] Soyibou SY et Cornel Marius MUREA. « Algorithm for solving fluid-structure interaction problem on a global moving mesh ». In : *Coupled Systems Mechanics* 1 (2012), pages 99–113. URL : http://www.edp.lmia.uha.fr/sy/ps/Sy_Murea_CSM_2012.pdf.
- [85] N.H. TAI et al. « Effects of thickness on the fatigue behavior of quasi-isotropic carbon/epoxy composites before and after low-energy impacts, » in : *Composites Science and Technologie* 59 (1999), pages 1753–1762.
- [86] T.K. TSOTSIS et S.M. LEE. « Long-term thermo-oxidative aging in composite materials : failure mechanisms ». In : *Composites Science and Technologie* 58 (1998), pages 355–368.
- [87] Cédric VILLANI. *Analyse II*. ENS Lyon. 2003-04. URL : <http://cedricvillani.org/wp-content/uploads/2013/03/ana2.pdf>.
- [88] Cédric VILLANI. *Intégration et Analyse de Fourier*. ENS Lyon. 2005-06. URL : <http://cedricvillani.org/wp-content/uploads/2013/03/IAF.pdf>.
- [89] Cédric VILLANI. *Intégration et Analyse de Fourier*. ENS Lyon. 2005-06. URL : <http://cedricvillani.org/wp-content/uploads/2013/03/IAF.pdf>.
- [90] Kyuichiro WASHIZU. *Variational methods in elasticity and plasticity*. Pergamon Press, 1982.
- [91] WIKIPÉDIA. *Calcul des variations*. URL : http://fr.wikipedia.org/wiki/Calcul_des_variations.
- [92] WIKIPÉDIA. *Courant–Friedrichs–Lewy condition*. URL : http://en.wikipedia.org/wiki/Courant%E2%80%93Friedrichs%E2%80%93Lewy_condition.
- [93] WIKIPÉDIA. *Diagramme de Voronoï*. URL : http://fr.wikipedia.org/wiki/Diagramme_de_Vorono%C3%AF.
- [94] WIKIPÉDIA. *Équations aux dérivées partielles*. URL : http://fr.wikipedia.org/wiki/%C3%A9quation_aux_d%C3%A9riv%C3%A9es_partielles.

- [95] WIKIPÉDIA. *Espace L^p* . URL : http://fr.wikipedia.org/wiki/Espace_Lp.
- [96] WIKIPÉDIA. *Formulation faible*. URL : http://fr.wikipedia.org/wiki/Formulation_faible.
- [97] WIKIPÉDIA. *Formulation variationnelle*. URL : http://fr.wikipedia.org/wiki/Formulation_variationnelle.
- [98] WIKIPÉDIA. *Formule de Newton-Cotes*. URL : https://fr.wikipedia.org/wiki/Formule_de_Newton-Cotes.
- [99] WIKIPÉDIA. *Histoire des mathématiques*. URL : http://fr.wikipedia.org/wiki/Histoire_des_math%C3%A9matiques.
- [100] WIKIPÉDIA. *Méthode de Monte-Carlo*. URL : http://fr.wikipedia.org/wiki/M%C3%A9thode_de_Monte-Carlo.
- [101] WIKIPÉDIA. *Méthode des éléments finis*. URL : http://fr.wikipedia.org/wiki/M%C3%A9thode_des_%C3%A9l%C3%A9ments_finis.
- [102] WIKIPÉDIA. *Méthodes de quadrature de Gauss*. URL : http://fr.wikipedia.org/wiki/M%C3%A9thodes_de_quadrature_de_Gauss.
- [103] WIKIPÉDIA. *Optimisation (mathématiques)*. URL : [http://fr.wikipedia.org/wiki/Optimisation_\(math%C3%A9matique\)](http://fr.wikipedia.org/wiki/Optimisation_(math%C3%A9matique)).
- [104] WIKIPÉDIA. *Polynôme homogène*. URL : https://fr.wikipedia.org/wiki/Polyn%C3%B4me_homog%C3%A8ne.
- [105] WIKIPÉDIA. *Résonance de Helmholtz*. URL : http://fr.wikipedia.org/wiki/R%C3%A9sonance_de_Helmholtz.
- [106] WIKIPÉDIA. *Théories des milieux effectifs*. URL : http://fr.wikipedia.org/wiki/Th%C3%A9ories_des_milieux_effectifs.
- [107] WIKIPÉDIA. *Toutes les notes bibliographiques et historiques de ce document*.
- [108] WIKIPÉDIA. *Triangulation de Delaunay*. URL : http://fr.wikipedia.org/wiki/Triangulation_de_Delaunay.
- [109] WIKIPÉDIA. *Zone absorbante parfaitement adaptée*. URL : http://fr.wikipedia.org/wiki/Zone_absorbante_parfaitement_adapt%C3%A9e.
- [110] Sebastian XHONNEUX. « Perception de l'optimisation en mathématiques et en économie au fil des siècles et l'enseignement du théorème de Lagrange ». In : *Journées Nationales de l'APMEP, La Rochelle*. 27 octobre 2008. URL : <http://www.apmep.asso.fr/IMG/pdf/Lu-33-Xhonneux-LaRochelle.pdf>.
- [111] Zheng YE, Victor L. BERDICHEVSKYB et Wenbin YU. « An Equivalent Classical Plate Model of Corrugated Structures ». In : *International Journal of Solids and Structures* (2014). URL : <http://dx.doi.org/10.1016/j.ijsolstr.2014.02.025> (cf. page 214).
- [112] Eberhard ZWICKER et Richard FELDTKELLER. *Psychoacoustique. L'oreille récepteur d'information*. Masson, 1981.