



**HAL**  
open science

## Efficient C++ finite element computing with Rheolef

Pierre Saramito

► **To cite this version:**

Pierre Saramito. Efficient C++ finite element computing with Rheolef. DEA. Grenoble, France, France. 2018, pp.259. cel-00573970v14

**HAL Id: cel-00573970**

**<https://cel.hal.science/cel-00573970v14>**

Submitted on 21 Feb 2018 (v14), last revised 2 Jun 2022 (v16)

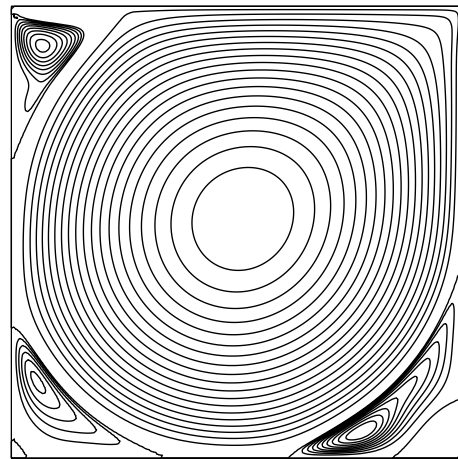
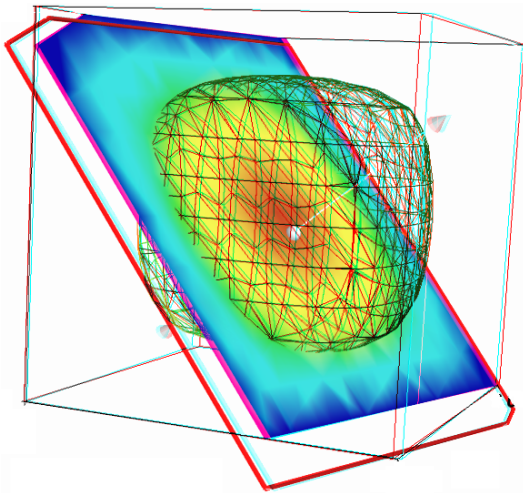
**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

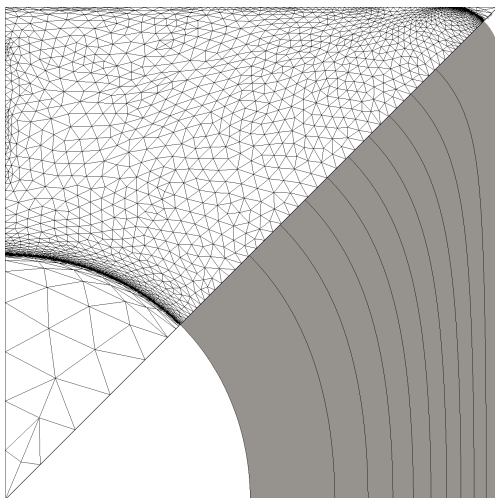
# Efficient C++ finite element computing with Rheolef

Pierre Saramito

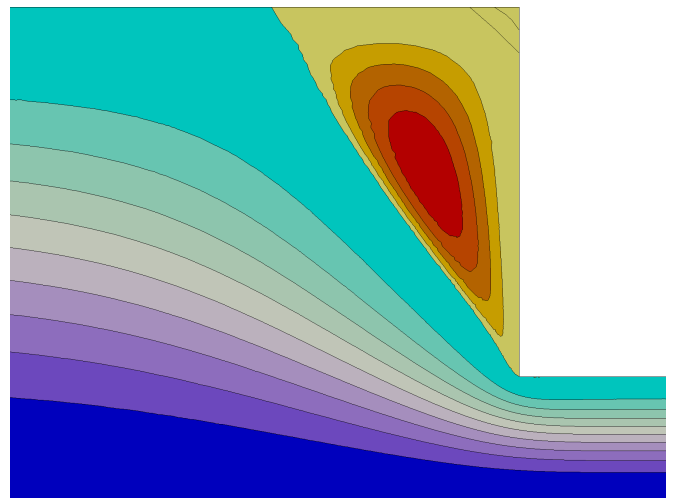
version 7.0 update 20 February 2018



$Re = 10000$



$Bi = 0.5$



$We = 0.7$

Copyright (c) 2003-2018 Pierre Saramito

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

## Introduction

Rheolef is a programming environment for finite element method computing. The reader is assumed to be familiar with (i) the c++ programming language and (ii) the finite element method. As a Lego game, the Rheolef bricks allow the user to solve most problems, from simple to complex multi-physics ones, in few lines of code. The concision and readability of codes written with Rheolef is certainly a major keypoint of this environment. Here is an example of a Rheolef code for solving the Poisson problem with homogeneous boundary conditions:

Example: find $u$ such that $-\Delta u = 1$ in $\Omega$ and $u = 0$ on $\partial\Omega$	
<pre>int main (int argc, char** argv) {   <b>environment</b> rheolef (argc, argv);   <b>geo</b> omega (argv[1]);   <b>space</b> Xh (omega, argv[2]);   Xh.block ("boundary");   <b>trial</b> u (Xh); <b>test</b> v (Xh);   <b>form</b> a = integrate (dot(grad(u),grad(v)));   <b>field</b> lh = integrate (v);   <b>field</b> uh (Xh);   uh ["boundary"] = 0;   <b>solver</b> sa (a.uu());   uh.u = sa.solve (lh.u());   dout &lt;&lt; uh; }</pre>	<pre>Let <math>\Omega \subset \mathbb{R}^N</math>, <math>N = 1, 2, 3</math> <math>X_h = \{v \in H^1(\Omega); v _K \in P_k, \forall K \in \mathcal{T}_h\}</math> <math>V_h = X_h \cap H_0^1(\Omega)</math> <math>a(u, v) = \int_{\Omega} \nabla u \cdot \nabla v \, dx</math> <math>l(v) = \int_{\Omega} v \, dx</math> (P) : find <math>u_h \in V_h</math> such that <math>a(u_h, v_h) = l(v_h), \forall v_h \in V_h</math></pre>

The right column shows the one-to-one line **correspondence between the code and the variational formulation**. Let us quote B. Stroustrup [106], the conceptor of the c++ language:

*"The time taken to write a program is at best roughly proportional to the **number of lines** written, and so is the number of errors in that code. It follows that a good way of writing correct programs is to write **short programs**. In other words, we need good libraries to allow us to write correct code that performs well. This in turn means that we need libraries to get our programs finished in a reasonable time. In many fields, such c++ libraries exist."*

Rheolef is an attempt to provide such a library in the field of finite element methods for partial differential equations. Rheolef provides both a c++ library and a set of unix commands for **shell** programming, providing **data structures** and **algorithms** [112].

- **Data structures** fit the variational formulation concept: **field**, bilinear **form** and functional **space**, are c++ types for variables. They can be combined in algebraic expressions, as you write it on the paper.
- **Algorithms** refer to the most up-to-date ones: direct an iterative **sparse matrix solvers** for linear systems. They supports efficient **distributed** memory and **parallel** computations. Non-linear c++ generic algorithms such as **fixed point**, **damped Newton** and **continuation** methods are also provided.

General *high order* piecewise polynomial finite element approximations are implemented, together with some mixed combinations for Stokes and incompressible elasticity. The *characteristic method* can be used for diffusion-convection problems while hyperbolic systems can be discretized by the discontinuous Galerkin method.

### Contacts

email [Pierre.Saramito@imag.fr](mailto:Pierre.Saramito@imag.fr)

home page <http://www-ljk.imag.fr/membres/Pierre.Saramito/rheolef>

Please send all patches, comments and bug reports by mail to

[rheolef@grenet.fr](mailto:rheolef@grenet.fr)

# Contents

Notations	8
<b>1 Getting started</b>	<b>11</b>
1.1 The model problem	11
1.1.1 Problem statement	12
1.1.2 Approximation	12
1.1.3 Comments	13
1.1.4 How to compile the code	14
1.1.5 How to run the program	15
1.1.6 Advanced and stereo visualization	15
1.1.7 High-order finite element methods	16
1.1.8 Tridimensional computations	17
1.1.9 Quadrangles, prisms and hexahedra	18
1.1.10 Direct versus iterative solvers	18
1.1.11 Distributed and parallel runs	20
1.1.12 Non-homogeneous Dirichlet conditions	22
1.2 Non-homogeneous Neumann boundary conditions for the Helmholtz operator	27
1.3 The Robin boundary conditions	29
1.4 Neumann boundary conditions for the Laplace operator	31
1.5 Non-constant coefficients and multi-regions	34
<b>2 Fluids and solids computations</b>	<b>41</b>
2.1 The linear elasticity and the Stokes problems	41
2.1.1 The linear elasticity problem	41
2.1.2 Computing the stress tensor	45
2.1.3 Mesh adaptation	47
2.1.4 The Stokes problem	51
2.1.5 Computing the vorticity	55
2.1.6 Computing the stream function	56
2.2 Nearly incompressible elasticity and the stabilized Stokes problems	58
2.2.1 The incompressible elasticity problem	58
2.2.2 The $P_1b - P_1$ element for the Stokes problem	60
2.2.3 Axisymmetric geometries	67
2.2.4 The axisymmetric stream function and stress tensor	67
2.3 Time-dependent problems	70

2.3.1	The heat equation . . . . .	70
2.3.2	The convection-diffusion problem . . . . .	73
2.4	The Navier-Stokes equations . . . . .	78
<b>3</b>	<b>Advanced and highly nonlinear problems</b>	<b>87</b>
3.1	Equation defined on a surface . . . . .	87
3.1.1	Approximation on an explicit surface mesh . . . . .	88
	The Helmholtz-Beltrami problem . . . . .	88
	The Laplace-Beltrami problem . . . . .	92
3.1.2	Building a surface mesh from a level set function . . . . .	95
3.1.3	The banded level set method . . . . .	98
3.1.4	Improving the banded level set method with a direct solver . . . . .	100
3.2	The highly nonlinear $p$ -laplacian problem . . . . .	105
3.2.1	Problem statement . . . . .	105
3.2.2	The fixed-point algorithm . . . . .	105
3.2.3	The Newton algorithm . . . . .	114
3.2.4	The damped Newton algorithm . . . . .	120
3.2.5	Error analysis . . . . .	123
3.3	[New] Continuation and bifurcation methods . . . . .	125
3.3.1	Problem statement and the Newton method . . . . .	126
3.3.2	Error analysis and multiplicity of solutions . . . . .	129
3.3.3	The Euler-Newton continuation algorithm . . . . .	132
3.3.4	Beyond the limit point : the Keller algorithm . . . . .	135
<b>4</b>	<b>Discontinuous Galerkin methods</b>	<b>141</b>
4.1	Scalar first-order problems . . . . .	141
4.1.1	The transport equation . . . . .	141
4.1.2	Nonlinear scalar hyperbolic problems . . . . .	144
4.1.3	Slope limiters . . . . .	146
4.1.4	Example: the Burgers equation . . . . .	148
4.2	Scalar second-order problems . . . . .	154
4.2.1	The Poisson problem with Dirichlet boundary conditions . . . . .	154
4.2.2	The Helmholtz problem with Neumann boundary conditions . . . . .	157
4.2.3	Nonlinear scalar hyperbolic problems with diffusion . . . . .	159
4.2.4	Example: the Burgers equation with diffusion . . . . .	159
4.3	Fluids and solids computations revisited . . . . .	165
4.3.1	The linear elasticity problem . . . . .	165
4.3.2	The Stokes problem . . . . .	167
4.4	The stationary Navier-Stokes equations . . . . .	169
4.4.1	Problem statement . . . . .	169
4.4.2	The discrete problem . . . . .	170
4.4.3	A conservative variant . . . . .	173
4.4.4	Newton solver . . . . .	176
4.4.5	Application to the driven cavity benchmark . . . . .	179
4.4.6	Upwinding . . . . .	180

---

<b>5</b>	<b>[New] Complex fluids</b>	<b>187</b>
5.1	[New] Yield slip at the wall . . . . .	187
5.1.1	Problem statement . . . . .	187
5.1.2	The augmented Lagrangian algorithm . . . . .	187
5.1.3	Newton algorithm . . . . .	192
5.1.4	Error analysis . . . . .	197
5.2	[New] Viscoplastic fluids . . . . .	200
5.2.1	Problem statement . . . . .	200
5.2.2	The augmented Lagrangian algorithm . . . . .	200
5.2.3	Mesh adaptation . . . . .	205
5.2.4	Error analysis . . . . .	207
5.2.5	Error analysis for the yield surface . . . . .	208
5.3	[New] Viscoelastic fluids . . . . .	213
5.3.1	A tensor transport equation . . . . .	213
5.3.2	The Oldroyd model . . . . .	216
5.3.3	The $\theta$ -scheme algorithm . . . . .	216
5.3.4	Flow in an abrupt ontraction . . . . .	221
<b>A</b>	<b>Technical appendices</b>	<b>229</b>
A.1	How to write a variational formulation ? . . . . .	229
A.1.1	The Green formula . . . . .	229
A.1.2	The vectorial Green formula . . . . .	229
A.1.3	The Green formula on a surface . . . . .	230
A.2	How to prepare a mesh ? . . . . .	230
A.2.1	Bidimensional mesh with <b>bamg</b> . . . . .	231
A.2.2	Unidimensional mesh with <b>gmsh</b> . . . . .	232
A.2.3	Bidimensional mesh with <b>gmsh</b> . . . . .	232
A.2.4	Tridimensional mesh with <b>gmsh</b> . . . . .	233
A.3	Migrating to <b>Rheolef</b> version 6.0 . . . . .	234
A.3.1	What is new in <b>Rheolef</b> 6.0 ? . . . . .	235
A.3.2	What should I have to change in my 5.x code ? . . . . .	235
A.3.3	New features in <b>Rheolef</b> 6.4 . . . . .	237
<b>B</b>	<b>GNU Free Documentation License</b>	<b>239</b>
	<b>List of example files</b>	<b>251</b>
	<b>List of commands</b>	<b>254</b>
	<b>Index</b>	<b>256</b>



## Notations

Rheolef	mathematics	description
d	$d \in \{1, 2, 3\}$	dimension of the physical space
dot(u,v)	$\mathbf{u} \cdot \mathbf{v} = \sum_{i=0}^{d-1} u_i v_i$	vector scalar product
ddot(sigma,tau)	$\sigma : \tau = \sum_{i,j=0}^{d-1} \sigma_{i,j} \tau_{i,j}$	tensor scalar product
tr(sigma)	$\text{tr}(\sigma) = \sum_{i=0}^{d-1} \sigma_{i,i}$	trace of a tensor
trans(sigma)	$\sigma^T$	tensor transposition
sqr(phi) norm2(phi)	$\phi^2$	square of a scalar
norm2(u)	$ \mathbf{u} ^2 = \sum_{i=0}^{d-1} u_i^2$	square of the vector norm
norm2(sigma)	$ \sigma ^2 = \sum_{i,j=0}^{d-1} \sigma_{i,j}^2$	square of the tensor norm
abs(phi) norm(phi)	$ \phi $	absolute value of a scalar
norm(u)	$ \mathbf{u}  = \left( \sum_{i=0}^{d-1} u_i^2 \right)^{1/2}$	vector norm
norm(sigma)	$ \sigma  = \left( \sum_{i,j=0}^{d-1} \sigma_{i,j}^2 \right)^{1/2}$	tensor norm
grad(phi)	$\nabla \phi = \left( \frac{\partial \phi}{\partial x_i} \right)_{0 \leq i < d}$	gradient of a scalar field in $\Omega \subset \mathbb{R}^d$
grad(u)	$\nabla \mathbf{u} = \left( \frac{\partial u_i}{\partial x_j} \right)_{0 \leq i,j < d}$	gradient of a vector field
div(u)	$\text{div}(\mathbf{u}) = \text{tr}(\nabla \mathbf{u}) = \sum_{i=0}^{d-1} \frac{\partial u_i}{\partial x_i}$	divergence of a vector field
D(u)	$D(\mathbf{u}) = (\nabla \mathbf{u} + \nabla \mathbf{u}^T) / 2$	symmetric part of the gradient of a vector field
curl(u)	$\text{curl}(\mathbf{u}) = \nabla \wedge \mathbf{u}$	curl of a vector field, when $d = 3$
curl(phi)	$\text{curl}(\phi) = \left( \frac{\partial \phi}{\partial x_1}, -\frac{\partial \phi}{\partial x_0} \right)$	curl of a scalar field, when $d = 2$
curl(u)	$\text{curl}(\mathbf{u}) = \frac{\partial u_1}{\partial x_0} - \frac{\partial u_0}{\partial x_1}$	curl of a vector field, when $d = 2$
grad_s(phi)	$\nabla_s \phi = P \nabla \phi$ where $P = I - \mathbf{n} \otimes \mathbf{n}$	tangential gradient of a scalar

<b>Rheolef</b>	mathematics	description
<code>grad_s(u)</code>	$\nabla_s \mathbf{u} = \nabla \mathbf{u} P$	tangential gradient of a vector
<code>Ds(u)</code>	$D_s(\mathbf{u}) = PD(\mathbf{u})P$	symmetrized tangential gradient
<code>div_s(u)</code>	$\text{div}_s(\mathbf{u}) = \text{tr}(D_s(\mathbf{u}))$	tangential divergence
<code>normal()</code>	$\mathbf{n}$	unit outward normal on $\Gamma = \partial\Omega$ or on an oriented surface $\Omega$ or on an internal oriented side $S$
<code>jump(phi)</code>	$[[\phi]] = \phi _{K_0} - \phi _{K_1}$	jump accross inter-element side $S = \partial K_0 \cap K_1$
<code>average(phi)</code>	$\{\{\phi\}\} = (\phi _{K_0} + \phi _{K_1})/2$	average accross $S$
<code>inner(phi)</code>	$\phi _{K_0}$	inner trace on $S$
<code>outer(phi)</code>	$\phi _{K_1}$	outer trace on $S$
<code>h_local()</code>	$h_K = \text{meas}(K)^{1/d}$	length scale on an element $K$
<code>penalty()</code>	$\varpi_s = \max\left(\frac{\text{meas}(\partial K_0)}{\text{meas}(K_0)}, \frac{\text{meas}(\partial K_1)}{\text{meas}(K_1)}\right)$	penalty coefficient on $S$
<code>grad_h(phi)</code>	$(\nabla_h \phi) _K = \nabla(\phi _K), \forall K \in \mathcal{T}_h$	broken gradient
<code>div_h(u)</code>	$(\text{div}_h \mathbf{u}) _K = \text{div}(\mathbf{u} _K), \forall K \in \mathcal{T}_h$	broken divergence of a vector field
<code>Dh(u)</code>	$(D_h(\mathbf{u})) _K = D(\mathbf{u} _K), \forall K \in \mathcal{T}_h$	broken symmetric part of the gradient of a vector field
<code>sin(phi)</code> <code>cos(phi)</code> <code>tan(phi)</code> <code>acos(phi)</code> <code>asin(phi)</code> <code>atan(phi)</code> <code>cosh(phi)</code> <code>sinh(phi)</code> <code>tanh(phi)</code> <code>exp(phi)</code> <code>log(phi)</code> <code>log10(phi)</code> <code>floor(phi)</code> <code>ceil(phi)</code> <code>min(phi,psi)</code>	$\sin(\phi)$ $\cos(\phi)$ $\tan(\phi)$ $\cos^{-1}(\phi)$ $\sin^{-1}(\phi)$ $\tan^{-1}(\phi)$ $\cosh(\phi)$ $\sinh(\phi)$ $\tanh(\phi)$ $\exp(\phi)$ $\log(\phi)$ $\log_{10}(\phi)$ $\lfloor \phi \rfloor$ $\lceil \phi \rceil$ $\min(\phi, \psi)$	standard mathematical functions extended to scalar fields  largest integral not greater than $\phi$ smallest integral not less than $\phi$

<b>Rheolef</b>	mathematics	description
<code>max(phi,psi)</code> <code>pow(phi,psi)</code> <code>atan2(phi,psi)</code> <code>fmod(phi,psi)</code>	$\max(\phi, \psi)$ $\phi^\psi$ $\tan^{-1}(\psi/\phi)$ $\phi - \lfloor \phi/\psi + 1/2 \rfloor \psi$	floating point remainder
<code>compose(f,phi)</code>	$f \circ \phi = f(\phi)$	applies an unary function $f$
<code>compose(f,phi1,...,phin)</code>	$f(\phi_1, \dots, \phi_n)$	applies a $n$ -ary function $f$ , $n \geq 1$
<code>compose(phi,X)</code>	$\phi \circ X$ , $X(x) = x + \mathbf{d}(x)$	composition with a characteristic

# Chapter 1

## Getting started

The first chapter of this book starts with the Dirichlet problem with homogeneous boundary condition: this example is declined with details in dimension 1, 2 and 3, as a starting point to **Rheolef**.

Next chapters present various boundary conditions: for completeness, we treat non-homogeneous Dirichlet, Neumann, and Robin boundary conditions for the model problem. The last two examples presents some special difficulties that appears in most problems: the first one introduce to problems with non-constant coefficients and the second one, a ill-posed problem where the solution is defined up to a constant.

This first chapter can be viewed as a pedagogic preparation for more advanced applications, such as Stokes and elasticity, that are treated in the second chapter of this book. Problem with non-constant coefficients are common as subproblems generated by various algorithms for non-linear problem.

### 1.1 The model problem

For obtaining and installing **Rheolef**, see the installation instructions on the **Rheolef** home page:

<http://www-ljk.imag.fr/membres/Pierre.Saramito/rheolef>

Before to run examples, please check your **Rheolef** installation with:

```
rheolef-config --check
```

The present book is available in the documentation directory of the **Rheolef** distribution. This documentation directory is given by the following unix command:

```
rheolef-config --docdir
```

All examples presented along the present book are also available in the **example/** directory of the **Rheolef** distribution. This directory is given by the following unix command:

```
rheolef-config --exampledir
```

This command returns you a path, something like `/usr/share/doc/rheolef-doc/examples` and you should make a copy of these files:

```
cp -a /usr/share/doc/rheolef-doc/examples .
cd examples
```

### 1.1.1 Problem statement

Let us consider the classical Poisson problem with homogeneous Dirichlet boundary conditions in a domain bounded  $\Omega \subset \mathbb{R}^d$ ,  $d = 1, 2, 3$ :

(P): find  $u$ , defined in  $\Omega$ , such that:

$$-\Delta u = 1 \text{ in } \Omega \quad (1.1)$$

$$u = 0 \text{ on } \partial\Omega \quad (1.2)$$

where  $\Delta$  denotes the Laplace operator. The variational formulation of this problem expresses (see appendix A.1.1 for details):

(VF): find  $u \in H_0^1(\Omega)$  such that:

$$a(u, v) = l(v), \quad \forall v \in H_0^1(\Omega) \quad (1.3)$$

where the bilinear form  $a(.,.)$  and the linear form  $l(.)$  are defined by

$$a(u, v) = \int_{\Omega} \nabla u \cdot \nabla v \, dx, \quad \forall u, v \in H_0^1(\Omega)$$

$$l(v) = \int_{\Omega} v \, dx, \quad \forall v \in L^2(\Omega)$$

The bilinear form  $a(.,.)$  defines a scalar product in  $H_0^1(\Omega)$  and is related to the *energy* form. This form is associated to the  $-\Delta$  operator.

### 1.1.2 Approximation

Let us introduce a mesh  $\mathcal{T}_h$  of  $\Omega$  and the finite dimensional space  $X_h$  of continuous piecewise polynomial functions.

$$X_h = \{v \in H^1(\Omega); v|_K \in P_k, \forall K \in \mathcal{T}_h\}$$

where  $k = 1$  or  $2$ . Let  $V_h = X_h \cap H_0^1(\Omega)$  be the functions of  $X_h$  that vanishes on the boundary of  $\Omega$ . The approximate problem expresses:

(VF)<sub>h</sub>: find  $u_h \in V_h$  such that:

$$a(u_h, v_h) = l(v_h), \quad \forall v_h \in V_h$$

By developing  $u_h$  on a basis of  $V_h$ , this problem reduces to a linear system. The following C++ code implement this problem in the **Rheolef** environment.

Example file 1.1: `dirichlet.cc`

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 int main(int argc, char**argv) {
5     environment rheolef (argc, argv);
6     geo omega (argv[1]);
7     space Xh (omega, argv[2]);
8     Xh.block ("boundary");
9     trial u (Xh); test v (Xh);
10    form a = integrate (dot(grad(u), grad(v)));
11    field lh = integrate (v);
12    field uh (Xh);
13    uh ["boundary"] = 0;
14    solver sa (a.uu());
15    uh.set_u() = sa.solve (lh.u() - a.ub()*uh.b());
16    dout << uh;
17 }

```

### 1.1.3 Comments

This code applies for both one, two or three dimensional meshes and for both piecewise linear or quadratic finite element approximations. Four major classes are involved, namely: `geo`, `space`, `form` and `field`.

Let us now comment the code, line by line.

```
#include "rheolef.h"
```

The first line includes the `Rheolef` header file `'rheolef.h'`.

```
using namespace rheolef;
using namespace std;
```

By default, in order to avoid possible name conflicts when using another library, all class and function names are prefixed by `rheolef::`, as in `rheolef::space`. This feature is called the name space. Here, since there is no possible conflict, and in order to simplify the syntax, we drop all the `rheolef::` prefixes, and do the same with the standard c++ library classes and variables, that are also prefixed by `std::`.

```
int main(int argc, char**argv) {
```

The entry function of the program is always called `main` and accepts arguments from the unix command line: `argc` is the counter of command line arguments and `argv` is the table of values. The character string `argv[0]` is the program name and `argv[i]`, for  $i = 1$  to `argc-1`, are the additional command line arguments.

```
environment rheolef (argc, argv);
```

These two command line parameters are immediately furnished to the distributed environment initializer of the `boost::mpi` library, that is a c++ library based on the usual message passing interface (MPI) library. Notice that this initialization is required, even when you run with only one processor.

```
geo omega (argv[1]);
```

This command get the first unix command-line argument `argv[1]` as a mesh file name and store the corresponding mesh in the variable `omega`.

```
space Xh (omega, argv[2]);
```

Build the finite element space `Xh` contains all the piecewise polynomial continuous functions. The polynomial type is the second command-line arguments `argv[2]`, and could be either `P1`, `P2` or any `Pk`, where  $k \geq 1$ .

```
Xh.block ("boundary");
```

The homogeneous Dirichlet conditions are declared on the boundary.

```
trial u (Xh); test v (Xh);
form a = integrate (dot(grad(u), grad(v)));
```

The bilinear form `a(.,.)` is the energy form: it is defined for all functions `u` and `v` in `Xh`.

```
field lh = integrate (v);
```

The linear form `lh(.)` is associated to the constant right-hand side  $f = 1$  of the problem. It is defined for all `v` in `Xh`.

```
field uh (Xh);
```

The field `uh` contains the the degrees of freedom.

```
uh ["boundary"] = 0;
```

Some degrees of freedom are prescribed as zero on the boundary.

Let  $(\varphi_i)_{0 \leq i < \dim(X_h)}$  be the basis of  $X_h$  associated to the Lagrange nodes, e.g. the vertices of the mesh for the  $P_1$  approximation and the vertices and the middle of the edges for the  $P_2$  approximation. The approximate solution  $u_h$  expresses as a linear combination of the continuous piecewise polynomial functions  $(\varphi_i)$ :

$$u_h = \sum_i u_i \varphi_i$$

Thus, the field  $u_h$  is completely represented by its coefficients  $(u_i)$ . The coefficients  $(u_i)$  of this combination are grouped into to sets: some have zero values, from the boundary condition and are related to *blocked* coefficients, and some others are *unknown*. Blocked coefficients are stored into the `uh.b` array while unknown one are stored into `uh.u`. Thus, the restriction of the bilinear form  $a(.,.)$  to  $X_h \times X_h$  can be conveniently represented by a block-matrix structure:

$$a(u_h, v_h) = \begin{pmatrix} \text{vh.u} & \text{vh.b} \end{pmatrix} \begin{pmatrix} \text{a.uu} & \text{a.ub} \\ \text{a.bu} & \text{a.bb} \end{pmatrix} \begin{pmatrix} \text{uh.u} \\ \text{uh.b} \end{pmatrix}$$

This representation also applies for the linear form  $l(.)$ :

$$l(v_h) = \begin{pmatrix} \text{vh.u} & \text{vh.b} \end{pmatrix} \begin{pmatrix} \text{lh.u} \\ \text{lh.b} \end{pmatrix}$$

Thus, the problem  $(VF)_h$  writes now:

$$\begin{pmatrix} \text{vh.u} & \text{vh.b} \end{pmatrix} \begin{pmatrix} \text{a.uu} & \text{a.ub} \\ \text{a.bu} & \text{a.bb} \end{pmatrix} \begin{pmatrix} \text{uh.u} \\ \text{uh.b} \end{pmatrix} = \begin{pmatrix} \text{vh.u} & \text{vh.b} \end{pmatrix} \begin{pmatrix} \text{lh.u} \\ \text{lh.b} \end{pmatrix}$$

for any `vh.u` and where `vh.b = 0`. After expansion, the problem reduces to *find* `uh.u` such that:

$$\text{a.uu} * \text{uh.u} = \text{lh.u} - \text{a.ub} * \text{uh.b}$$

The resolution of this linear system for the `a.uu` matrix is then performed. A preliminary step build the  $LDL^T$  factorization:

```
solver sa (a.uu());
```

Then, the second step solves the *unknown part*:

```
uh.set_u() = sa.solve (lh.u() - a.ub()*uh.b());
```

When  $d > 3$ , a faster iterative strategy is automatically preferred by the `solver` class for solving the linear system: in that case, the preliminary step build an incomplete Choleski factorization preconditioner, while the second step runs an iterative method: the preconditioned conjugate gradient algorithm. Finally, the field is printed to standard output:

```
dout << uh;
```

The `dout` stream is a specific variable defined in the **Rheolef** library: it is a distributed and parallel extension of the usual `cout` stream in C++

### 1.1.4 How to compile the code

First, create a file `'Makefile'` as follow:

```
include $(shell rheolef-config --libdir)/rheolef/rheolef.mk
CXXFLAGS = $(INCLUDES_RHEOLEF)
LDLIBS = $(LIBS_RHEOLEF)
default: dirichlet
```

Then, enter:

```
make dirichlet
```

Now, your program, linked with **Rheolef**, is ready to run on a mesh.

### 1.1.5 How to run the program

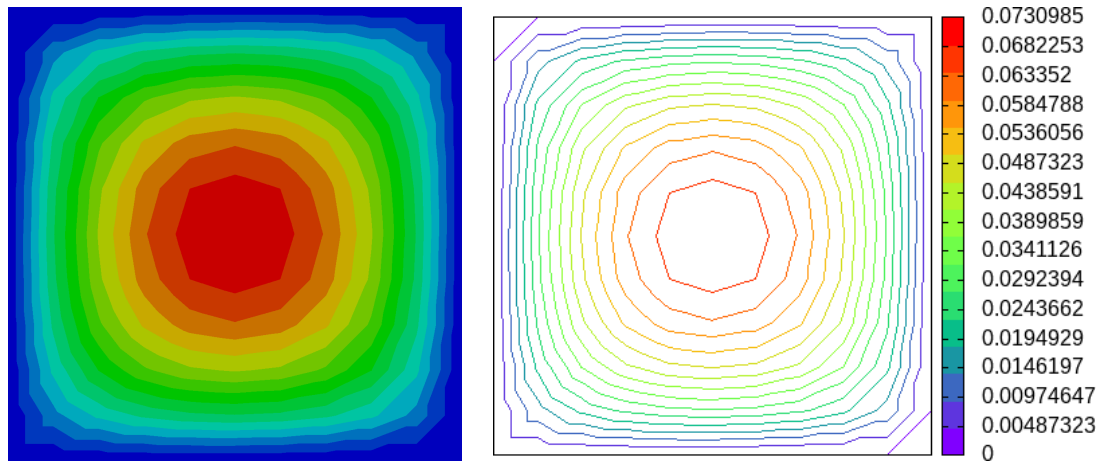


Figure 1.1: Solution of the model problem for  $d = 2$  with the  $P_1$  element: visualization (left) with `paraview` as filled isocontours ; (right) with `gnuplot` as unfilled isocontours.

Enter the commands:

```
mkgeo_grid -t 10 > square.geo
geo square.geo
```

The first command generates a simple 10x10 bidimensional mesh of  $\Omega = ]0, 1[^2$  and stores it in the file `square.geo`. The second command shows the mesh. It uses `gnuplot` visualization program by default.

The next commands perform the computation and visualization:

```
./dirichlet square.geo P1 > square.field
field square.field
```


The result is shown on Fig. 1.1. By default, the visualization appears in a `paraview` window. If you are in trouble with this software, you can switch to the simpler `gnuplot` visualization mode:

```
field square.field -gnuplot
```

### 1.1.6 Advanced and stereo visualization

We could explore some graphic rendering modes (see Fig. 1.2):

```
field square.field -bw
field square.field -gray
field square.field -elevation
field square.field -elevation -gray
field square.field -elevation -nofill -stereo
```

The last command shows the solution in elevation and in stereoscopic anaglyph mode (see Fig. 1.4, left). The anaglyph mode requires red-cyan glasses: red for the left eye and cyan for the right one, as shown on Fig. 1.3. In the book, stereo figures are indicated by the  logo in the right margin. See [http://en.wikipedia.org/wiki/Anaglyph\\_image](http://en.wikipedia.org/wiki/Anaglyph_image) for more and <http://www.alpes-stereo.com/lunettes.html> for how to find anaglyph red-cyan glasses. For simplicity, it would perhaps prefer to switch to the `gnuplot` render:



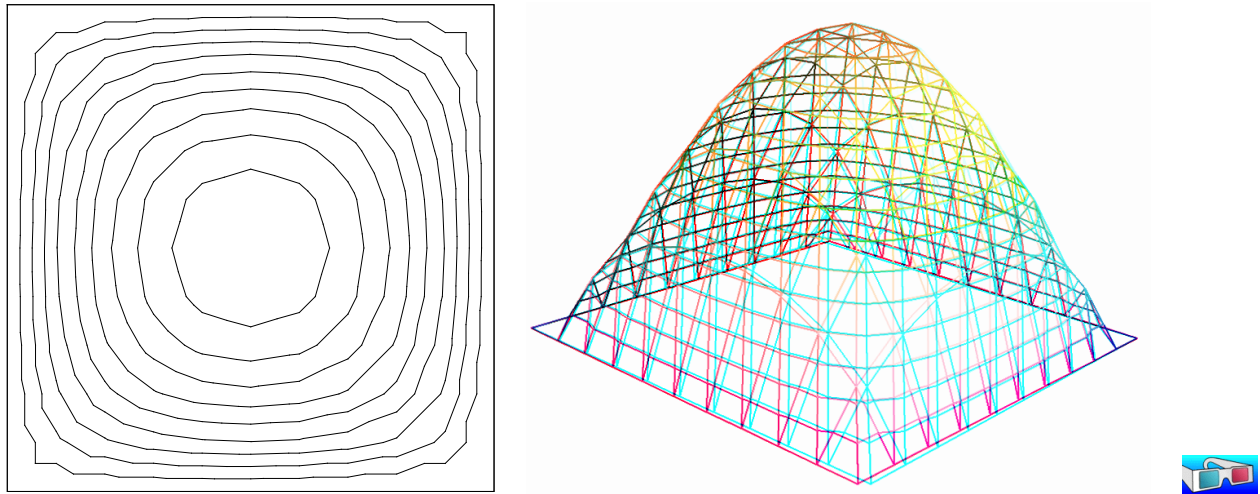


Figure 1.2: Alternative representations of the solution of the model problem ( $d = 2$  and the  $P_1$  element): (left) in black-and-white; (right) in elevation and stereoscopic anaglyph mode.

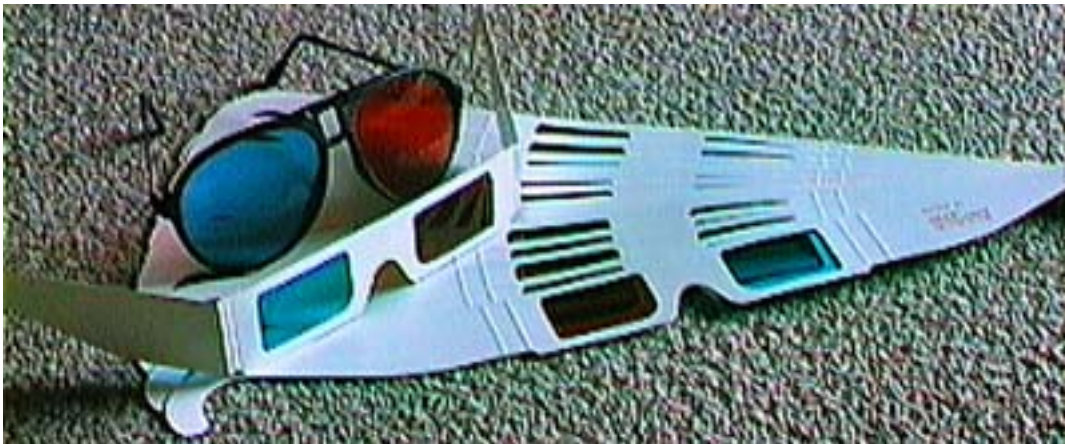


Figure 1.3: Red-cyan anaglyph glasses for the stereoscopic visualization.

```
field square.field -gnuplot
field square.field -gnuplot -bw
field square.field -gnuplot -gray
```

Please, consults the corresponding unix manual page for more on `field`, `geo` and `mkgeo_grid`:

```
man mkgeo_grid
man geo
man field
```

### 1.1.7 High-order finite element methods

Turning to the  $P_2$  or  $P_3$  approximations simply writes:

```
./dirichlet square.geo P2 > square-P2.field
field square-P2.field
```

You can replace the P2 command-line argument by any  $P_k$ , where  $k \geq 1$ . Now, let us consider a mono-dimensional problem  $\Omega = ]0, 1[$ :

```
mkgeo_grid -e 10 > line.geo
geo line.geo
./dirichlet line.geo P1 | field -
```

The first command generates a subdivision containing ten edge elements. The last two lines show the mesh and the solution via `gnuplot` visualization, respectively.

Conversely, the P2 case writes:

```
./dirichlet line.geo P2 | field -
```

### 1.1.8 Tridimensional computations

Let us consider a three-dimensional problem  $\Omega = ]0, 1[^3$ . First, let us generate a mesh:

```
mkgeo_grid -T 10 > cube.geo
geo cube.geo
geo cube.geo -fill
geo cube.geo -cut
geo cube.geo -shrink
geo cube.geo -shrink -cut
```

The 3D visualization bases on the `paraview` render. These commands present some cuts (`-cut`) inside the internal mesh structure: a simple click on the central arrow draws the cut plane normal vector or its origin, while the red square allows a translation. The following command draws the mesh with all internal edges (`-full`), together with the stereoscopic anaglyph (`-stereo`):

```
geo cube.geo -stereo -full
```

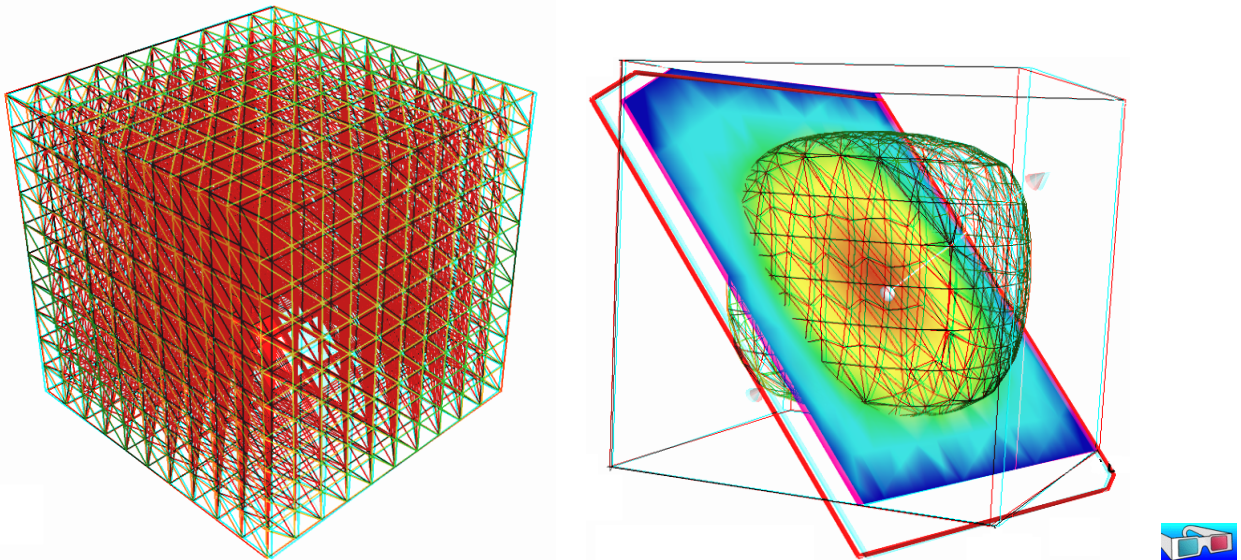


Figure 1.4: Solution of the model problem for  $d = 3$  and the  $P_1$  element : (left) mesh; (right) isovalue, cut planes and stereo anaglyph renderings.

Then, we perform the computation and the visualization:

```
./dirichlet cube.geo P1 > cube.field
field cube.field
```

The visualization presents an isosurface. Also here, you can interact with the cutting plane. On the **Properties** of the **paraview** window, select **Contour**, change the value of the isosurface and click on the green **Apply** button. Finally exit from the visualization and explore the stereoscopic anaglyph mode (see Fig. 1.4, right):

```
field cube.field -stereo
```

It is also possible to add a second isosurface (**Contour**) or a cutting plane (**Slice**) to this scene by using the corresponding **Properties** menu. Finally, the following command, with the **-volume** option, allows a 3D color light volume graphical rendering:

```
field cube.field -volume
```

After this exploration of the 3D visualization capacities of our environment, let us go back to the Dirichlet problem and perform the P2 approximation:

```
./dirichlet cube.geo P2 | field -
```

### 1.1.9 Quadrangles, prisms and hexahedra

Quadrangles and hexahedra are also supported in meshes:

```
mkgeo_grid -q 10 > square.geo
geo square.geo
mkgeo_grid -H 10 > cube.geo
geo cube.geo
```

Notices also that the one-dimensional exact solution writes:

$$u(x) = \frac{x(1-x)}{2}$$

while the two-and three dimensional ones support a Fourier expansion (see e.g. [96], annex).

### 1.1.10 Direct versus iterative solvers

In order to measure the performances of the solver, the `dirichlet.cc` (page 12) has been modified as:

```
double t0 = dis_wall_time();
solver_option sopt;
sopt.iterative = false; // or true
sopt.tol       = 1-5;   // when iterative
solver sa (a.uu(), sopt);
Float t_factorize = dis_wall_time() - t0;
uh.set_u() = sa.solve (lh.u() - a.ub()*uh.b());
double t_solve = dis_wall_time() - t_factorize;
derr << "time " << t_factorize << " " << t_solve << endl;
```

The `dis_wall_time` function returns the synchronized wall clock in seconds, while the `solver_option` enable to choose explicitly a direct or iterative solver method: by default **Rheolef** selects a direct method when  $d = 2$  and an iterative one when  $d = 3$ . For a large 3D mesh, the compilation and run writes:

```
make dirichlet
mkgeo_grid -T 60 > cube-60.geo
./dirichlet cube-60.geo P1 > cube-60.field
```

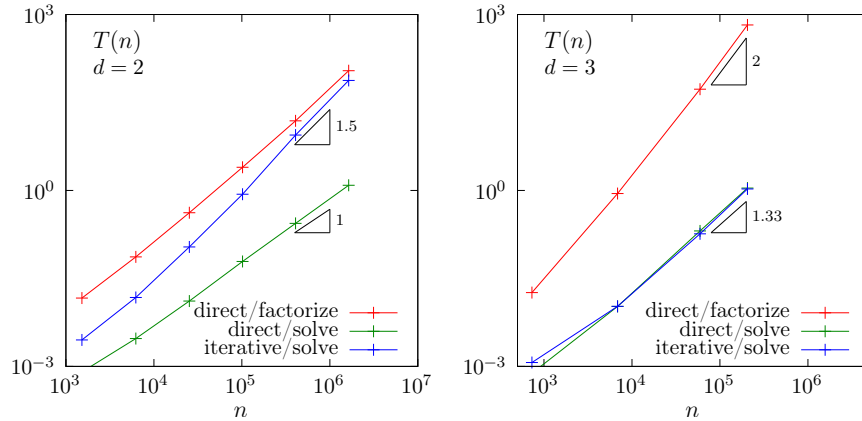


Figure 1.5: Compared performance between direct and iterative solvers: (left)  $d = 2$ ; (right)  $d = 3$ .

Fig. 1.5 plots the performances of the direct and iterative solvers used in **Rheolef**. The computing time  $T(n)$  is represented versus size  $n$  of the linear system, says  $Ax = b$ . Notice that for a `square- $k$ .geo` or `cube- $k$ .geo` mesh, we have  $n = (k - 1)^d$ . For the direct method, two times are represented: first, the time spend to *factorize*  $A = LDL^T$ , where  $L$  is lower triangular and  $D$  is diagonal, and second, the time used to *solve*  $LDL^T = x$  (in three steps: solving  $Lz = b$ , then  $Dy = z$  and finally  $L^T x = y$ ). For the iterative method, the conjugate gradient algorithm is considered, without building any preconditionner, so there is nothing to initialize, and only one time is represented. The tolerance on the residual term is set to  $10^{-5}$ .

In the bidimensional case, the iterative solver presents asymptotically, for large  $n$ , a computing time similar to the factorization time of the direct solver, roughly  $\mathcal{O}(n^{3/2})$  while the time to solve by the direct method is dramatically lower, roughly  $\mathcal{O}(n)$ . As the factorization can be done one time for all, the direct method is advantageous most of the time.

In the three dimensional case, the situation is different. The factorization time is very time consuming roughly  $\mathcal{O}(n^2)$ , while the time to solve for both direct and iterative methods behave as  $\mathcal{O}(n^{4/3})$ . Thus, the iterative method is clearly advantageous for threedimensionnal problems. Future works will improve the iterative approach by building preconditionners.

The asymptotic behaviors of direct methods strongly depends upon the ordering strategy used for the factorization. For the direct solver, **Rheolef** was configured with the `mumps` [4, 5] library and `mumps` was configured with the parallel `scotch` [72] ordering library. For a regular grid and in the bidimensional case, there exists a specific ordering called nested dissection [38, 48] that minimize the fillin of the sparse matrix during the factorization. For threedimensional case this ordering is called nested multi-section [9]. Asymptotic computing time for these regular grid are then explicitly known versus the grid size  $n$ :

$d$	direct/factorize	direct/solve	iterative
1	$n$	$n$	$n^2$
2	$n^{3/2}$	$n \log n$	$n^{3/2}$
3	$n^2$	$n^{4/3}$	$n^{4/3}$

The last column gives the asymptotic computing time for the conjugate gradient on a general mesh [89]. Remark that these theoretical results are consistent with numerical experiments presented on Fig. 1.5.

### 1.1.11 Distributed and parallel runs

For large meshes, a computation in a distributed and parallel environment is profitable:

```
mpirun -np 8 ./dirichlet cube-60.geo P1 > cube-60.field
mpirun -np 16 ./dirichlet cube-60.geo P1 > cube-60.field
```

The computing time  $T(n, p)$  depends now upon the linear system size  $n$  and the number of processes  $p$ . For a fixed system  $n$ , the speedup  $S(p)$  when using  $p$  processors is defined by the ratio of the time required by a sequential computation with the time used by a parallel one:  $S(p) = T(n, 1)/T(n, p)$ . The speedup is presented on Fig 1.6 for the two phases of the computation: the assembly phase and the solve one, and for  $d = 2$  (direct solver) and 3 (iterative solver). The ideal speedup  $S(p) = p$  and the null speedup  $S(p) = 1$  are represented by dotted lines. Observe on Fig 1.6 that for too small meshes, using too much processes is not profitable, as more time is spend by communications rather by computations, especially for the solve phase. Conversely, when the mesh size increases, using more processes leads to a remarkable speedup for both  $d = 2$  and 3. The largest mesh used here contains about three millions of elements. The speedup behavior is roughly linear up to a critical number of processor denotes as  $p_c$ . Then, there is a transition to a plateau (the Amdahl's law), where communications dominate. Notice that  $p_c$  increases with the mesh size: larger problems lead to a higher speedup. Also  $p_c$  increases also with the efficiency of communications.

Present computation times are measured on a BullX DLC supercomputer (Bull Newsca) composed of nodes having two intel sandy-bridge processors and connected to a FDR infiniband non-blocking low latency network. The assembly phase corresponds to `dirichlet.cc` (page 12) line 7 to 13 and the solve phase to lines 14 and 15.

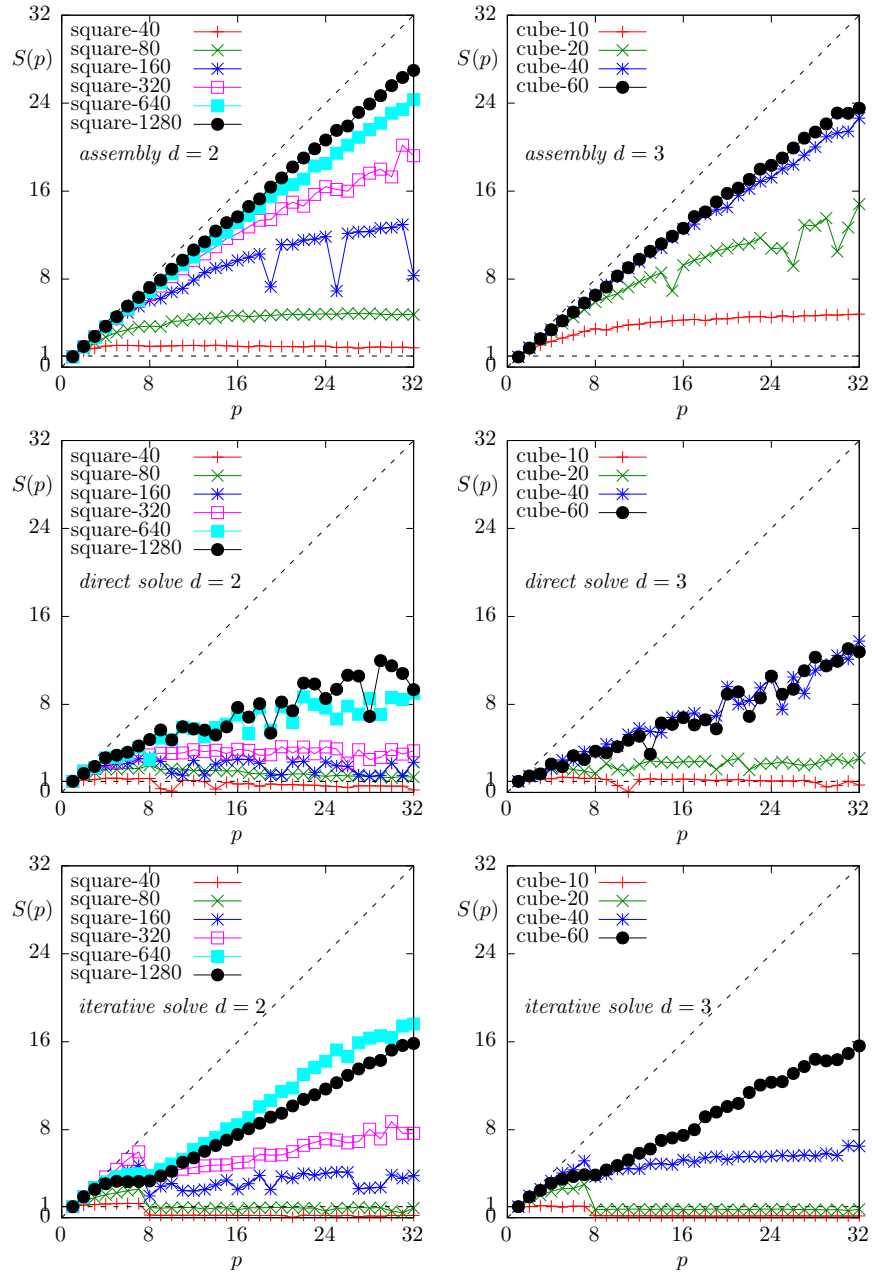


Figure 1.6: Distributed and massively parallel resolution of the model problem with  $P_1$  element: speedup  $S(p)$  versus the number of processors  $p$  during : (left-right) for  $d = 2$  and 3, respectively ; (top) the assembly phase ; (center-bottom) the solve phase, direct and iterative solvers, respectively.

### 1.1.12 Non-homogeneous Dirichlet conditions

#### Formulation

We turn now to the case of a non-homogeneous Dirichlet boundary conditions. Let  $f \in H^{-1}(\Omega)$  and  $g \in H^{\frac{1}{2}}(\partial\Omega)$ . The problem writes:

(P<sub>2</sub>) find  $u$ , defined in  $\Omega$  such that:

$$\begin{aligned} -\Delta u &= f \text{ in } \Omega \\ u &= g \text{ on } \partial\Omega \end{aligned}$$

The variational formulation of this problem expresses:

(VF<sub>2</sub>) find  $u \in V$  such that:

$$a(u, v) = l(v), \quad \forall v \in V_0$$

where

$$\begin{aligned} a(u, v) &= \int_{\Omega} \nabla u \cdot \nabla v \, dx \\ l(v) &= \int_{\Omega} f v \, dx \\ V &= \{v \in H^1(\Omega); v|_{\partial\Omega} = g\} \\ V_0 &= H_0^1(\Omega) \end{aligned}$$

#### Approximation

As usual, we introduce a mesh  $\mathcal{T}_h$  of  $\Omega$  and the finite dimensional space  $X_h$ :

$$X_h = \{v \in H^1(\Omega); v|_K \in P_k, \forall K \in \mathcal{T}_h\}$$

Then, we introduce:

$$\begin{aligned} V_h &= \{v \in X_h; v|_{\partial\Omega} = \pi_h(g)\} \\ V_{0,h} &= \{v \in X_h; v|_{\partial\Omega} = 0\} \end{aligned}$$

where  $\pi_h$  denotes the Lagrange interpolation operator. The approximate problem writes:  
(VF<sub>2</sub>)<sub>h</sub>: find  $u_h \in V_h$  such that:

$$a(u_h, v_h) = l(v_h), \quad \forall v_h \in V_{0,h}$$

The following C++ code implement this problem in the **Rheolef** environment.

Example file 1.2: dirichlet-nh.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "cosinusprod_laplace.icc"
5 int main(int argc, char**argv) {
6     environment rheolef(argc, argv);
7     geo omega (argv[1]);
8     size_t d = omega.dimension();
9     space Xh (omega, argv[2]);
10    Xh.block ("boundary");
11    trial u (Xh); test v (Xh);
12    form a = integrate (dot(grad(u), grad(v)));
13    field lh = integrate (f(d)*v);
14    field uh (Xh);
15    space Wh (omega["boundary"], argv[2]);
16    uh ["boundary"] = interpolate(Wh, g(d));
17    solver sa (a.uu());
18    uh.set_u() = sa.solve (lh.u() - a.ub()*uh.b());
19    dout << uh;
20 }

```

Let us choose  $\Omega \subset \mathbb{R}^d$ ,  $d = 1, 2, 3$  with

$$f(x) = d\pi^2 \prod_{i=0}^{d-1} \cos(\pi x_i) \quad \text{and} \quad g(x) = \prod_{i=0}^{d-1} \cos(\pi x_i) \quad (1.4)$$

Remarks the notation  $x = (x_0, \dots, x_{d-1})$  for the Cartesian coordinates in  $\mathbb{R}^d$ : since all arrays start at index zero in C++ codes, and in order to avoid any mistakes between the code and the mathematical formulation, we also adopt this convention here. This choice of  $f$  and  $g$  is convenient, since the exact solution is known:

$$u(x) = \prod_{i=0}^{d-1} \cos(\pi x_i)$$

The following C++ code implement this problem by using the concept of *function object*, also called *class-function* (see e.g. [61]). A convenient feature is the ability for function objects to store auxiliary parameters, such as the space dimension  $d$  for  $f$  here, or some constants, as  $\pi$  for  $f$  and  $g$ .

Example file 1.3: cosinusprod\_laplace.icc

```

1 struct f {
2     Float operator() (const point& x) const {
3         return d*pi*pi*cos(pi*x[0])*cos(pi*x[1])*cos(pi*x[2]); }
4     f(size_t d1) : d(d1), pi(acos(Float(-1))) {}
5     size_t d; const Float pi;
6 };
7 struct g {
8     Float operator() (const point& x) const {
9         return cos(pi*x[0])*cos(pi*x[1])*cos(pi*x[2]); }
10    g(size_t d1) : pi(acos(Float(-1))) {}
11    const Float pi;
12 };

```

## Comments

The class `point` describes the coordinates of a point  $(x_0, \dots, x_{d-1}) \in \mathbb{R}^d$  as a  $d$ -uplet of `Float`. The `Float` type is usually a `double`. This type depends upon the `Rheolef` configuration (see [88], installation instructions), and could also represent some high precision floating point class. The `dirichlet-nh.cc` code looks like the previous one `dirichlet.cc` related to homogeneous boundary conditions. Let us comments the changes. The dimension  $d$  comes from the geometry  $\Omega$ :



```
size_t d = omega.dimension();
```

The linear form  $l(\cdot)$  is associated to the right-hand side  $f$  and writes:

```
field lh = integrate (f(d)*v);
```

Notice that the function  $f$  that depends upon the dimension  $d$  parameter, is implemented by a *functor*, i.e. a C++ class that possesses the `operator()` member function. The space  $W_h$  of piecewise  $P_k$  functions defined on the boundary  $\partial\Omega$  is defined by:

```
space Wh (omega["boundary"], argv[2]);
```

where  $P_k$  is defined via the second command line argument `argv[2]`. This space is suitable for the Lagrange interpolation of  $g$  on the boundary:

```
uh ["boundary"] = interpolate(Wh, g(d));
```

The values of the degrees of freedom related to the boundary are stored into the field `uh.b`, where non-homogeneous Dirichlet conditions applies. The rest of the code is similar to the homogeneous Dirichlet case.

### How to run the program

First, compile the program:

```
make dirichlet-nh
```

Running the program is obtained from the homogeneous Dirichlet case, by replacing `dirichlet` by `dirichlet-nh`:

```
mkgeo_grid -e 10 > line.geo
./dirichlet-nh line.geo P1 > line.field
field line.field
```

for the bidimensional case:

```
mkgeo_grid -t 10 > square.geo
./dirichlet-nh square.geo P1 > square.field
field square.field
```

and for the tridimensional case:

```
mkgeo_grid -T 10 > box.geo
./dirichlet-nh box.geo P1 > box.field
field box.field -volume
```

The optional `-volume` allows a 3D color light volume graphical rendering. Here, the P1 approximation can be replaced by P2, P3, etc, by modifying the command-line argument.

### Error analysis

Since the solution  $u$  is regular, the following error estimates holds:

$$\begin{aligned} \|u - u_h\|_{0,2,\Omega} &= \mathcal{O}(h^{k+1}) \\ \|u - u_h\|_{0,\infty,\Omega} &= \mathcal{O}(h^{k+1}) \\ \|u - u_h\|_{1,2,\Omega} &= \mathcal{O}(h^k) \end{aligned}$$

providing the approximate solution  $u_h$  uses  $P_k$  continuous finite element method,  $k \geq 1$ . Here,  $\|\cdot\|_{0,2,\Omega}$ ,  $\|\cdot\|_{0,\infty,\Omega}$  and  $\|\cdot\|_{1,2,\Omega}$  denotes as usual the  $L^2(\Omega)$ ,  $L^\infty(\Omega)$  and  $H^1(\Omega)$  norms.

By denoting  $\pi_h$  the Lagrange interpolation operator, the triangular inequality leads to:

$$\|u - u_h\|_{0,2,\Omega} \leq \|(I - \pi_h)(u)\|_{0,2,\Omega} + \|u_h - \pi_h u\|_{0,2,\Omega}$$

From the fundamental properties of the Laplace interpolation  $\pi_h$ , and since  $u$  is smooth enough, we have  $\|(I - \pi_h)(u)\|_{0,2,\Omega} = O(h^{k+1})$ . Thus, we have just to check the  $\|u_h - \pi_h u\|_{0,2,\Omega}$  term. The following code implement the computation of the error.

Example file 1.4: `cosinusprod_error.cc`

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "cosinusprod.icc"
5 int main(int argc, char**argv) {
6     environment rheolef(argc, argv);
7     Float error_linf_expected = (argc > 1) ? atof(argv[1]) : 1e+38;
8     field uh; din >> uh;
9     space Xh = uh.get_space();
10    size_t d = Xh.get_geo().dimension();
11    field pi_h_u = interpolate(Xh, u_exact(d));
12    field eh = uh - pi_h_u;
13    trial u (Xh); test v (Xh);
14    form m = integrate (u*v);
15    form a = integrate (dot(grad(u), grad(v)));
16    dout << "error_l2 " << sqrt(m(eh,eh)) << endl
17         << "error_linf " << eh.max_abs() << endl
18         << "error_h1 " << sqrt(a(eh,eh)) << endl;
19    return (eh.max_abs() <= error_linf_expected) ? 0 : 1;
20 }

```

Example file 1.5: `cosinusprod.icc`

```

1 struct u_exact {
2     Float operator() (const point& x) const {
3         return cos(pi*x[0])*cos(pi*x[1])*cos(pi*x[2]); }
4     u_exact(size_t d1) : d(d1), pi(acos(Float(-1.0))) {}
5     size_t d; Float pi;
6 };

```

The  $m(.,.)$  is here the classical scalar product on  $L^2(\Omega)$ , and is related to the *mass* form.

## Running the program

```
make dirichlet-nh sinusprod_error
```

After compilation, run the code by using the command:

```
mkgeo_grid -t 10 > square.geo
./dirichlet-nh square.geo P1 | ./cosinusprod_error
```

The three  $L^2$ ,  $L^\infty$  and  $H^1$  errors are printed for a  $h = 1/10$  uniform mesh. Note that an unstructured quasi-uniform mesh can be simply generated by using the `mkgeo_ugrid` command:

```
mkgeo_ugrid -t 10 > square.geo
geo square.geo
```

Let  $n_{el}$  denotes the number of elements in the mesh. Since the mesh is quasi-uniform, we have  $h \approx n_{el}^{-\frac{1}{d}}$  where  $d$  is the physical space dimension. Here  $d = 2$  for our bidimensional mesh. Figure 1.7 plots in logarithmic scale the error versus  $n_{el}^{\frac{1}{2}}$  for both  $P_k$  approximations,  $k = 1, 2, 3$  and the various norms. Observe that the error behaves as predicted by the theory.

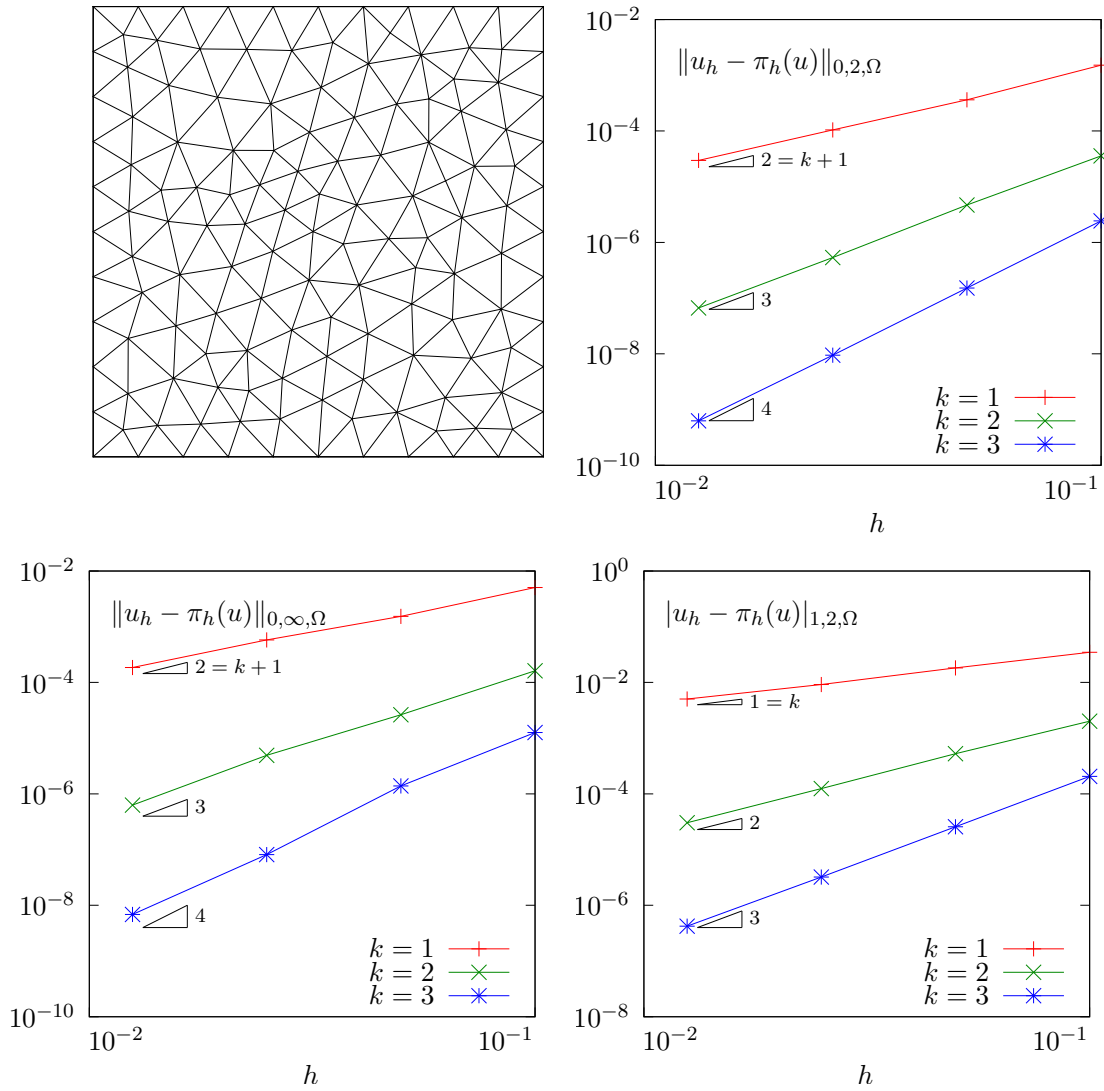


Figure 1.7: Strait geometry: error analysis in  $L^2$ ,  $L^\infty$  and  $H^1$  norms.

## Curved domains

The error analysis applies also for curved boundaries and high order approximations.

Example file 1.6: cosinusrad\_laplace.icc

```

1 struct f {
2   Float operator() (const point& x) const {
3     Float r = sqrt(sqr(x[0])+sqr(x[1])+sqr(x[2]));
4     Float sin_over_ar = (r == 0) ? 1 : sin(a*r)/(a*r);
5     return sqr(a)*((d-1)*sin_over_ar + cos(a*r)); }
6   f(size_t d1) : d(d1), a(acos(Float(-1.0))) {}
7   size_t d; Float a;
8 };
9 struct g {
10  Float operator() (const point& x) const {
11    return cos(a*sqrt(sqr(x[0])+sqr(x[1])+sqr(x[2]))); }
12  g(size_t=0) : a(acos(Float(-1.0))) {}
13  Float a;
14 };

```

Example file 1.7: cosinusrad.icc

```

1 struct u_exact {
2   Float operator() (const point& x) const {
3     Float r = sqrt(sqr(x[0])+sqr(x[1])+sqr(x[2]));
4     return cos(a*r); }
5   u_exact(size_t=0) : a(acos(Float(-1.0))) {}
6   Float a;
7 };

```

First, generate the test source file and compile it:

```

sed -e 's/sinusprod/sinusrad/' < dirichlet-nh.cc      > dirichlet_nh_ball.cc
sed -e 's/sinusprod/sinusrad/' < cosinusprod_error.cc > cosinusrad_error.cc
make dirichlet_nh_ball cosinusrad_error

```

Then, generates the mesh of a circle and run the test case:

```

mkgeo_ball -order 1 -t 10 > circle-P1.geo
geo circle-P1
./dirichlet_nh_ball circle-P1.geo P1 | ./cosinusrad_error

```

For a high order  $k = 3$  isoparametric approximation:

```

mkgeo_ball -order 3 -t 10 > circle-P3.geo
geo circle-P3
./dirichlet_nh_ball circle-P3.geo P3 | ./cosinusrad_error

```

Observe Fig. 1.8: for meshes based on triangles: the error behave as expected when  $k = 1, 2, 3, 4$ .

A similar result occurs for quadrangles, as shown on Fig. 1.9.

```

mkgeo_ball -order 3 -q 11 > circle-q-P3.geo
geo circle-q-P3
./dirichlet_nh_ball circle-q-P3.geo P3 | ./cosinusrad_error

```

These features are currently in development for arbitrarily  $P_k$  high order approximations and three-dimensional geometries.

## 1.2 Non-homogeneous Neumann boundary conditions for the Helmholtz operator

### Formulation

Let us show how to insert Neumann boundary conditions. Let  $f \in H^{-1}(\Omega)$  and  $g \in H^{-\frac{1}{2}}(\partial\Omega)$ . The problem writes:

(P<sub>3</sub>): find  $u$ , defined in  $\Omega$  such that:

$$\begin{aligned} u - \Delta u &= f \text{ in } \Omega \\ \frac{\partial u}{\partial n} &= g \text{ on } \partial\Omega \end{aligned}$$

The variational formulation of this problem expresses:

(VF<sub>3</sub>): find  $u \in H^1(\Omega)$  such that:

$$a(u, v) = l(v), \quad \forall v \in H^1(\Omega)$$

where

$$\begin{aligned} a(u, v) &= \int_{\Omega} (u v + \nabla u \cdot \nabla v) \, dx \\ l(v) &= \int_{\Omega} f v \, dx + \int_{\partial\Omega} g v \, ds \end{aligned}$$

### Approximation

As usual, we introduce a mesh  $\mathcal{T}_h$  of  $\Omega$  and the finite dimensional space  $X_h$ :

$$X_h = \{v \in H^1(\Omega); v|_K \in P_k, \forall K \in \mathcal{T}_h\}$$

The approximate problem writes:

(VF<sub>3</sub>)<sub>h</sub>: find  $u_h \in X_h$  such that:

$$a(u_h, v_h) = l(v_h), \quad \forall v_h \in X_h$$

Example file 1.8: neumann-nh.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "sinusprod_helmholtz.icc"
5 int main(int argc, char**argv) {
6     environment rheolef(argc, argv);
7     geo omega (argv[1]);
8     size_t d = omega.dimension();
9     space Xh (omega, argv[2]);
10    trial u (Xh); test v (Xh);
11    form a = integrate (u*v + dot(grad(u), grad(v)));
12    field lh = integrate (f(d)*v) + integrate ("boundary", g(d)*v);
13    field uh (Xh);
14    solver sa (a.uu());
15    uh.set_u() = sa.solve (lh.u() - a.ub()*uh.b());
16    dout << uh;
17 }

```

Let us choose  $\Omega \subset \mathbb{R}^d$ ,  $d = 1, 2, 3$  and

$$\begin{aligned} f(x) &= (1 + d\pi^2) \prod_{i=0}^{d-1} \sin(\pi x_i) \\ g(x) &= \begin{cases} -\pi & \text{when } d = 1 \\ -\pi \left( \sum_{i=0}^{d-1} \sin(\pi x_i) \right) & \text{when } d = 2 \\ -\pi \left( \sum_{i=0}^{d-1} \sin(\pi x_i) \sin(x_{(i+1) \bmod d}) \right) & \text{when } d = 3 \end{cases} \end{aligned}$$

This example is convenient, since the exact solution is known:

$$u(x) = \prod_{i=0}^{d-1} \sin(\pi x_i) \quad (1.5)$$

Example file 1.9: sinusprod\_helmholtz.icc

```

1 struct f {
2   Float operator() (const point& x) const {
3     switch (d) {
4       case 1: return (1+d*pi*pi)*sin(pi*x[0]);
5       case 2: return (1+d*pi*pi)*sin(pi*x[0])*sin(pi*x[1]);
6       default: return (1+d*pi*pi)*sin(pi*x[0])*sin(pi*x[1])*sin(pi*x[2]);
7     }
8   f(size_t d1) : d(d1), pi(acos(Float(-1.0))) {}
9   size_t d; const Float pi;
10 };
11 struct g {
12   Float operator() (const point& x) const {
13     switch (d) {
14       case 1: return -pi;
15       case 2: return -pi*(sin(pi*x[0]) + sin(pi*x[1]));
16       default: return -pi*( sin(pi*x[0])*sin(pi*x[1])
17                             + sin(pi*x[1])*sin(pi*x[2])
18                             + sin(pi*x[2])*sin(pi*x[0]));
19     }
20   g(size_t d1) : d(d1), pi(acos(Float(-1.0))) {}
21   size_t d; const Float pi;
22 };

```

### Comments

The `neumann-nh.cc` code looks like the previous one `dirichlet-nh.cc`. Let us comments only the changes.

```
form a = integrate (u*v + dot(grad(u),grad(v)));
```

The bilinear form  $a(\cdot, \cdot)$  is defined. Notes the flexibility of the `integrate` function that takes as argument an expression involving the trial and test functions. The right-hand side is computed as:

```
field lh = integrate (f(d)*v) + integrate ("boundary", g(d)*v);
```

The second integration is performed on  $\partial\Omega$ . The additional first argument of the `integrate` function is here the name of the integration domain.

### How to run the program

First, compile the program:

```
make neumann-nh
```

Running the program is obtained from the homogeneous Dirichlet case, by replacing `dirichlet` by `neumann-nh`.

## 1.3 The Robin boundary conditions

### Formulation

Let  $f \in H^{-1}(\Omega)$  and Let  $g \in H^{\frac{1}{2}}(\partial\Omega)$ . The problem writes:

(P<sub>4</sub>) find  $u$ , defined in  $\Omega$  such that:

$$\begin{aligned} -\Delta u &= f \text{ in } \Omega \\ \frac{\partial u}{\partial n} + u &= g \text{ on } \partial\Omega \end{aligned}$$

The variational formulation of this problem expresses:

(VF<sub>4</sub>): find  $u \in H^1(\Omega)$  such that:

$$a(u, v) = l(v), \quad \forall v \in H^1(\Omega)$$

where

$$\begin{aligned} a(u, v) &= \int_{\Omega} \nabla u \cdot \nabla v \, dx + \int_{\partial\Omega} uv \, ds \\ l(v) &= \int_{\Omega} f v \, dx + \int_{\partial\Omega} g v \, ds \end{aligned}$$

## Approximation

As usual, let

$$X_h = \{v \in H^1(\Omega); v|_K \in P_k, \forall K \in \mathcal{T}_h\}$$

The approximate problem writes:

(VF<sub>4</sub>)<sub>h</sub>: find  $u_h \in X_h$  such that:

$$a(u_h, v_h) = l(v_h), \quad \forall v_h \in X_h$$

Example file 1.10: robin.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "cosinusprod_laplace.icc"
5 int main(int argc, char**argv) {
6     environment rheolef(argc, argv);
7     geo omega (argv[1]);
8     size_t d = omega.dimension();
9     space Xh (omega, argv[2]);
10    trial u (Xh); test v (Xh);
11    form a = integrate (dot(grad(u), grad(v))) + integrate ("boundary", u*v);
12    field lh = integrate (f(d)*v) + integrate ("boundary", g(d)*v);
13    field uh (Xh);
14    solver sa (a.uu());
15    uh.set_u() = sa.solve (lh.u() - a.ub()*uh.b());
16    dout << uh;
17 }

```

## Comments

The code robin.cc looks like the previous one neumann-nh.cc. Let us comments the changes.

```
form a = integrate (dot(grad(u), grad(v))) + integrate ("boundary", u*v);
```

This statement reflects directly the definition of the bilinear form  $a(.,.)$ , as the sum of two integrals, the first one over  $\Omega$  and the second one over its boundary.

## How to run the program

First, compile the program:

make robin

Running the program is obtained from the homogeneous Dirichlet case, by replacing `dirichlet` by `robin`.

## 1.4 Neumann boundary conditions for the Laplace operator

In this chapter we study how to solve a ill-posed problem with a solution defined up to a constant.

### Formulation

Let  $\Omega$  be a bounded open and simply connected subset of  $\mathbb{R}^d$ ,  $d = 1, 2$  or  $3$ . Let  $f \in L^2(\Omega)$  and  $g \in H^{\frac{1}{2}}(\partial\Omega)$  satisfying the following compatibility condition:

$$\int_{\Omega} f \, dx + \int_{\partial\Omega} g \, ds = 0$$

The problem writes:

$(P_5)_h$ : find  $u$ , defined in  $\Omega$  such that:

$$\begin{aligned} -\Delta u &= f \text{ in } \Omega \\ \frac{\partial u}{\partial n} &= g \text{ on } \partial\Omega \end{aligned}$$

Since this problem only involves the derivatives of  $u$ , it is clear that its solution is never unique [41, p. 11]. A discrete version of this problem could be solved iteratively by the conjugate gradient or the MINRES algorithm [69]. In order to solve it by a direct method, we turn the difficulty by seeking  $u$  in the following space

$$V = \{v \in H^1(\Omega); \ b(v, 1) = 0\}$$

where

$$b(v, \mu) = \mu \int_{\Omega} v \, dx, \quad \forall v \in L^2(\Omega), \forall \mu \in \mathbb{R}$$

The variational formulation of this problem writes:

$(VF_5)$ : find  $u \in V$  such that:

$$a(u, v) = l(v), \quad \forall v \in V$$

where

$$\begin{aligned} a(u, v) &= \int_{\Omega} \nabla u \cdot \nabla v \, dx \\ l(v) &= m(f, v) + m_b(g, v) \\ m(f, v) &= \int_{\Omega} f v \, dx \\ m_b(g, v) &= \int_{\partial\Omega} g v \, ds \end{aligned}$$

Since the direct discretization of the space  $V$  is not an obvious task, the constraint  $b(u, 1) = 0$  is enforced by a Lagrange multiplier  $\lambda \in \mathbb{R}$ . Let us introduce the Lagrangian, defined for all  $v \in H^1(\Omega)$  and  $\mu \in \mathbb{R}$  by:

$$L(v, \mu) = \frac{1}{2}a(v, v) + b(v, \mu) - l(v)$$



The saddle point  $(u, \lambda) \in H^1(\Omega) \times \mathbb{R}$  of this Lagrangian is characterized as the unique solution of:

$$\begin{aligned} a(u, v) + b(v, \lambda) &= l(v), \quad \forall v \in H^1(\Omega) \\ b(u, \mu) &= 0, \quad \forall \mu \in \mathbb{R} \end{aligned}$$

It is clear that if  $(u, \lambda)$  is solution of this problem, then  $u \in V$  and  $u$  is a solution of  $(VF_5)$ . Conversely, let  $u \in V$  the solution of  $(VF_5)$ . Choosing  $v = v_0$  where  $v_0(x) = 1, \forall x \in \Omega$  leads to  $\lambda \text{meas}(\Omega) = l(v_0)$ . From the definition of  $l(\cdot)$  and the compatibility condition between the data  $f$  and  $g$ , we get  $\lambda = 0$ . Notice that the saddle point problem extends to the case when  $f$  and  $g$  does not satisfies the compatibility condition, and in that case  $\lambda = l(v_0)/\text{meas}(\Omega)$ .

### Approximation

As usual, we introduce a mesh  $\mathcal{T}_h$  of  $\Omega$  and the finite dimensional space  $X_h$ :

$$X_h = \{v \in H^1(\Omega); v|_K \in P_k, \forall K \in \mathcal{T}_h\}$$

The approximate problem writes:

$(VF_5)_h$ : find  $(u_h, \lambda_h) \in X_h \times \mathbb{R}$  such that:

$$\begin{aligned} a(u_h, v) + b(v, \lambda_h) &= l_h(v), \quad \forall v \in X_h \\ b(u_h, \mu) &= 0, \quad \forall \mu \in \mathbb{R} \end{aligned}$$

where

$$l_h(v) = m(\Pi_h f, v_h) + m_b(\pi_h g, v_h)$$

Example file 1.11: neumann-laplace.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 size_t d;
5 Float f (const point& x) { return 1; }
6 Float g (const point& x) { return -0.5/d; }
7 int main(int argc, char**argv) {
8     environment rheolef (argc, argv);
9     geo omega (argv[1]);
10    d = omega.dimension();
11    space Xh (omega, argv[2]);
12    trial u (Xh); test v (Xh);
13    form m = integrate (u*v);
14    form a = integrate (dot(grad(u), grad(v)));
15    field b = m*field(Xh, 1);
16    field lh = integrate (f*v) + integrate ("boundary", g*v);
17    csr<Float> A = {{ a.uu(), b.uu() },
18                  { trans(b.u()), 0 }};
19    vec<Float> B = { lh.u(), 0 };
20    A.set_symmetry(true);
21    solver sa = ldlt(A);
22    vec<Float> U = sa.solve (B);
23    field uh (Xh);
24    uh.set_u() = U [range(0, uh.u().size())];
25    Float lambda = (U.size() == uh.u().size()+1) ? U [uh.u().size()] : 0;
26 #ifdef _RHEOLEF_HAVE_MPI
27    mpi::broadcast (U.comm(), lambda, U.comm().size() - 1);
28 #endif // _RHEOLEF_HAVE_MPI
29    dout << uh
30          << "lambda" << lambda << endl;
31 }

```

### Comments

Let  $\Omega \subset \mathbb{R}^d$ ,  $d = 1, 2, 3$ . We choose  $f(x) = 1$  and  $g(x) = -1/(2d)$ . This example is convenient, since the exact solution is known:

$$u(x) = -\frac{1}{12} + \frac{1}{2d} \sum_{i=1}^d x_i(1 - x_i)$$

The code looks like the previous ones. Let us comment the changes. The discrete bilinear form  $b$  is computed as  $b_h \in X_h$  that interprets as a linear application from  $X_h$  to  $\mathbb{R}$ :  $b_h(v_h) = m(v_h, 1)$ . Thus  $b_h$  is computed as

```
field b = m*field(Xh,1.0);
```

where the discrete bilinear form  $m$  is identified to its matrix and `field(Xh,1.0)` is the constant vector equal to 1. Let

$$\mathcal{A} = \begin{pmatrix} \text{a.uu} & \text{trans(b.u)} \\ \text{b.u} & 0 \end{pmatrix}, \quad \mathcal{U} = \begin{pmatrix} \text{uh.u} \\ \text{lambda} \end{pmatrix}, \quad \mathcal{B} = \begin{pmatrix} \text{lh.u} \\ 0 \end{pmatrix}$$

The problem admits the following matrix form:

$$\mathcal{A} \mathcal{U} = \mathcal{B}$$

The matrix and right-hand side of the system are assembled by concatenation:

```
csr<Float> A = {{ a.uu, b.u},
               {trans(b.u), 0 }};
vec<Float> B = { lh.u, 0 };
```

where `csr` and `vec` are respectively the matrix and vector classes. The `csr` is the abbreviation of *compressed sparse row*, a sparse matrix compression standard format. Notice that the matrix  $\mathcal{A}$  is symmetric and non-singular, but indefinite : it admits eigenvalues that are either strictly positive or strictly negative. While the Choleski factorization is not possible, its variant the  $LDL^T$  one is performed, thanks to the `ldlt` function:

```
solver sa = ldlt(A);
```

Then, the `uh.u` vector is extracted from the `U` one:

```
uh.u = U [range(0,uh.u.size())];
```

The extraction of `lambda` from `U` is more technical in a distributed environment. In a sequential one, since it is stored after the `uh.u` values, it could be simply written as:

```
Float lambda = U [uh.u.size()];
```

In a distributed environment, `lambda` is stored in `U` on the last processor, identified by `U.comm().size()-1`. Here `U.comm()` denotes the `communicator`, from the `boost::mpi` library and `U.comm().size()` is the number of processors in use, e.g. as specified by the `mpirun` command. On this last processor, the array `U` has size equal to `uh.u.size()+1` and `lambda` is stored in `U[uh.u.size()]`. On the others processors, the array `U` has size equal to `uh.u.size()` and `lambda` is not available. The following statement extract `lambda` on the last processor and set it to zero on the others:

```
Float lambda = (U.size() == uh.u.size()+1) ? U [uh.u.size()] : 0;
```

Then, the value of `lambda` is broadcasted on the others processors:

```
mpi::broadcast (U.comm(), lambda, U.comm().size() - 1);
```

The preprocessing guards `#ifdef...#endif` assure that this example compile also when the library is not installed with the MPI support. Finally, the statement

```
dout << catchmark("u") << uh
      << catchmark("lambda") << lambda << endl;
```

writes the solution  $(u_h, \lambda)$ . The `catchmark` function writes marks together with the solution in the output stream. These marks are suitable for a future reading with the same format, as:

```
din  >> catchmark("u") >> uh
      >> catchmark("lambda") >> lambda;
```

This is useful for post-treatment, visualization and error analysis.

### How to run the program

As usual, enter:

```
make neumann-laplace
mkgeo_grid -t 10 > square.geo
./neumann-laplace square P1 | field -
```

## 1.5 Non-constant coefficients and multi-regions

This chapter is related to the so-called transmission problem. We introduce some new concepts: problems with non-constant coefficients, regions in the mesh, weighted forms and discontinuous approximations.

### Formulation

Let us consider a diffusion problem with a non-constant diffusion coefficient  $\eta$  in a domain bounded  $\Omega \subset \mathbb{R}^d$ ,  $d = 1, 2, 3$ :

(P): *find  $u$  defined in  $\Omega$  such that:*

$$-\operatorname{div}(\eta \nabla u) = f \text{ in } \Omega \quad (1.6)$$

$$u = 0 \text{ on } \Gamma_{\text{left}} \cup \Gamma_{\text{right}} \quad (1.7)$$

$$\frac{\partial u}{\partial n} = 0 \text{ on } \Gamma_{\text{top}} \cup \Gamma_{\text{bottom}} \text{ when } d \geq 2 \quad (1.8)$$

$$\frac{\partial u}{\partial n} = 0 \text{ on } \Gamma_{\text{front}} \cup \Gamma_{\text{back}} \text{ when } d = 3 \quad (1.9)$$

where  $f$  is a given source term. We consider here the important special case when  $\eta$  is piecewise constant:

$$\eta(x) = \begin{cases} \varepsilon & \text{when } x \in \Omega_{\text{west}} \\ 1 & \text{when } x \in \Omega_{\text{east}} \end{cases}$$

where  $(\Omega_{\text{west}}, \Omega_{\text{east}})$  is a partition of  $\Omega$  in two parts (see Fig. 1.10). This is the so-called **transmission** problem: the solution and the flux are continuous on the interface  $\Gamma_0 = \partial\Omega_{\text{east}} \cap \partial\Omega_{\text{west}}$  between the two domains where the problem reduce to a constant diffusion one:

$$\begin{aligned} u_{\Omega_{\text{west}}} &= u_{\Omega_{\text{east}}} \text{ on } \Gamma_0 \\ \varepsilon \frac{\partial u_{\Omega_{\text{west}}}}{\partial n} &= \frac{\partial u_{\Omega_{\text{east}}}}{\partial n} \text{ on } \Gamma_0 \end{aligned}$$

It expresses the transmission of the quantity  $u$  and its flux across the interface  $\Gamma_0$  between two regions that have different diffusion properties: Notice that the more classical problem, with constant diffusion  $\eta$  on  $\Omega$  is obtained by simply choosing when  $\varepsilon = 1$ .

The variational formulation of this problem expresses:

(VF): find  $u \in V$  such that:

$$a(u, v) = l(v), \quad \forall v \in V$$

where the bilinear forms  $a(.,.)$  and the linear one  $l(.)$  are defined by

$$\begin{aligned} a(u, v) &= \int_{\Omega} \eta \nabla u \cdot \nabla v \, dx, \quad \forall u, v \in H^1(\Omega) \\ l(v) &= \int_{\Omega} f v \, dx, \quad \forall v \in L^2(\Omega) \\ V &= \{v \in H^1(\Omega); v = 0 \text{ on } \Gamma_{\text{left}} \cup \Gamma_{\text{right}}\} \end{aligned}$$

The bilinear form  $a(.,.)$  defines a scalar product in  $V$  and is related to the *energy* form. This form is associated to the  $-\text{div } \eta \nabla$  operator.

The approximation of this problem could be performed by a standard Lagrange  $P_k$  continuous approximation.

Example file 1.12: transmission.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 int main(int argc, char**argv) {
5     environment rheolef (argc, argv);
6     const Float epsilon = 0.01;
7     geo omega (argv[1]);
8     space Xh (omega, argv[2]);
9     Xh.block ("left");
10    Xh.block ("right");
11    string eta_approx = "P" + itos(Xh.degree()-1) + "d";
12    space Qh (omega, eta_approx);
13    field eta_h (Qh);
14    eta_h ["east"] = 1;
15    eta_h ["west"] = epsilon;
16    trial u (Xh); test v (Xh);
17    form a = integrate (eta_h*dot(grad(u), grad(v)));
18    field lh = integrate (v);
19    field uh (Xh);
20    uh["left"] = uh["right"] = 0;
21    solver sa (a.uu());
22    uh.set_u() = sa.solve (lh.u() - a.ub()*uh.b());
23    dout << catchmark("epsilon") << epsilon << endl
24         << catchmark("u") << uh;
25 }

```

## Comments

This file is quite similar to those studied in the first chapters of this book. Let us comment only the changes. The Dirichlet boundary condition applies no more on the whole boundary  $\partial\Omega$  but on two parts  $\Gamma_{\text{left}}$  and  $\Gamma_{\text{right}}$ . On the other boundary parts, an homogeneous Neumann boundary condition is used: since these conditions does not produce any additional terms in the variational formulation, there are also nothing to write in the C++ code for these boundaries. We choose  $f = 1$ : this leads to a convenient test-problem, since the exact solution is known when  $\Omega = ]0, 1[^d$ :

$$u(x) = \begin{cases} \frac{x_0}{2\varepsilon} \left( \frac{1+3\varepsilon}{2(1+\varepsilon)} - x_0 \right) & \text{when } x_0 < 1/2 \\ \frac{1-x_0}{2} \left( x_0 + \frac{1-\varepsilon}{2(1+\varepsilon)} \right) & \text{otherwise} \end{cases}$$

The field  $\eta$  belongs to a discontinuous finite element  $P_{k-1}$  space denoted by  $Q_h$ :

```
string eta_approx = "P" + itos(Xh.degree()-1) + "d";
space Qh (omega, eta_approx);
field eta (Qh);
```

For instance, when `argv[2]` contains "P2", i.e.  $k = 2$ , then the string `eta_approx` takes value "P1d". Then  $\eta$  is initialized by:

```
eta["east"] = 1;
eta["west"] = epsilon;
```

The energy form  $a$  is then constructed with  $\eta$  as additional parameter that acts as a integration *weight*:

```
form a = integrate (eta_h*dot(grad(u), grad(v)));
```

Such forms with a additional *weight* function are called *weighted forms* in **Rheolef**.

## How to run the program ?

Build the program as usual: `make transmission`. Then, creates a one-dimensional geometry with two regions:

```
mkgeo_grid -e 100 -region > line.geo
geo line.geo
```

The trivial mesh generator `mkgeo_grid`, defines two regions `east` and `west` when used with the `-region` option. This correspond to the situation:

$$\Omega = [0, 1]^d, \quad \Omega_{\text{west}} = \Omega \cap \{x_0 < 1/2\} \quad \text{and} \quad \Omega_{\text{east}} = \Omega \cap \{x_0 > 1/2\}.$$

In order to avoid mistakes with the C++ style indexes, we denote by  $(x_0, \dots, x_{d-1})$  the Cartesian coordinate system in  $\mathbb{R}^d$ .

Finally, run the program and look at the solution:

```
make transmission
./transmission line.geo P1 > line.field
field line.field
```

Since the exact solution is a piecewise second order polynomial and the change in the diffusion coefficient value fits the element boundaries, we obtain the exact solution for all the degrees of freedom of any  $P_k$  approximation,  $k \geq 1$ , as shown on Fig. 1.11 when  $k = 1$ . Moreover, when  $k \geq 2$  then  $u_h = u$  since  $X_h$  contains the exact solution  $u$ . The two dimensional case corresponds to the commands:

```
mkgeo_grid -t 10 -region > square.geo
geo square.geo
./transmission square.geo P1 > square.field
field square.field -elevation
```

while the tridimensional to

```
mkgeo_grid -T 10 -region > cube.geo
./transmission cube.geo P1 > cube.field
field cube.field
```

As for all the others examples, you can replace `P1` by higher-order approximations, change elements shapes, such as `q`, `H` or `P`, and run distributed computations computations with `mpirun`.

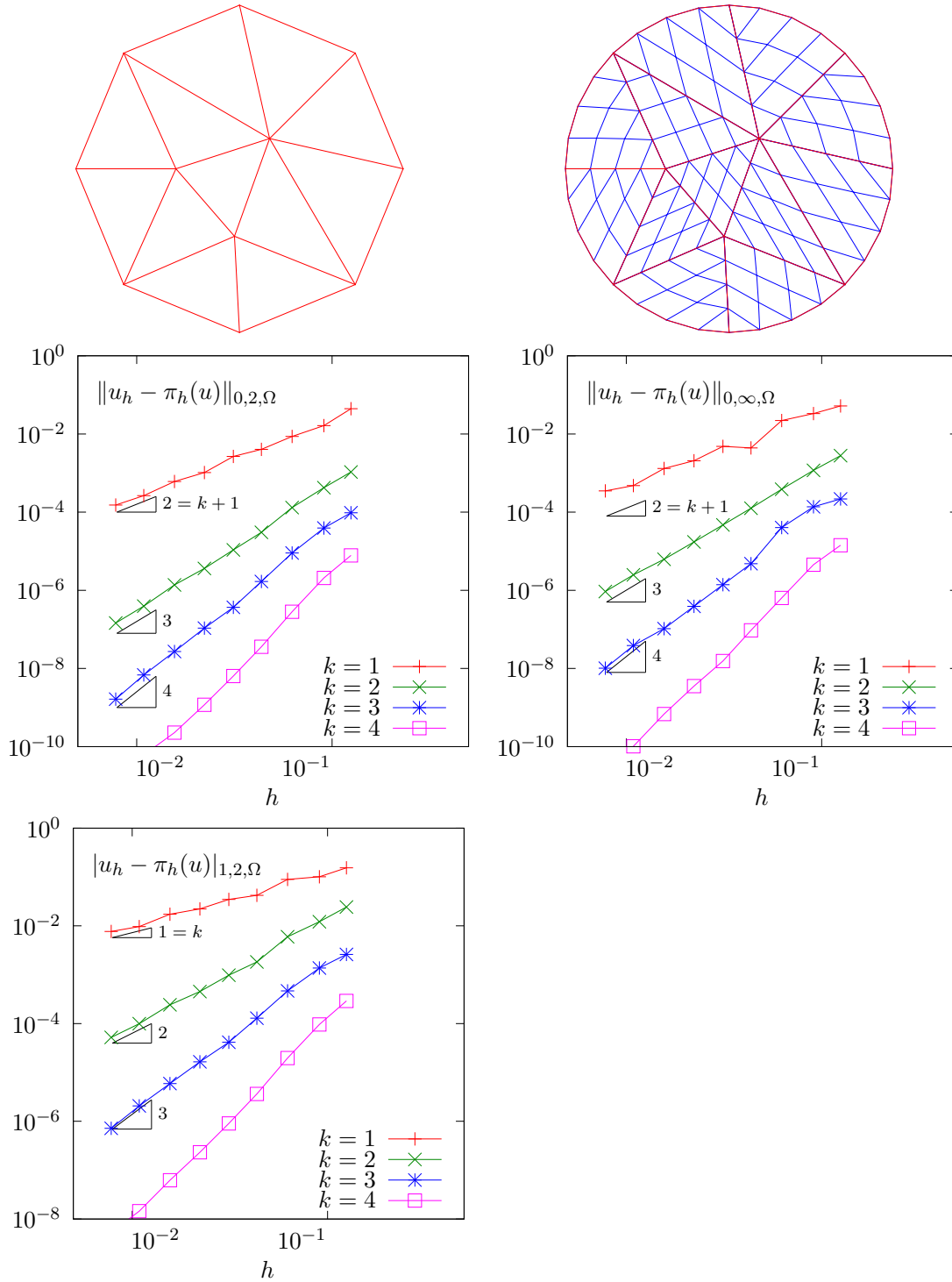
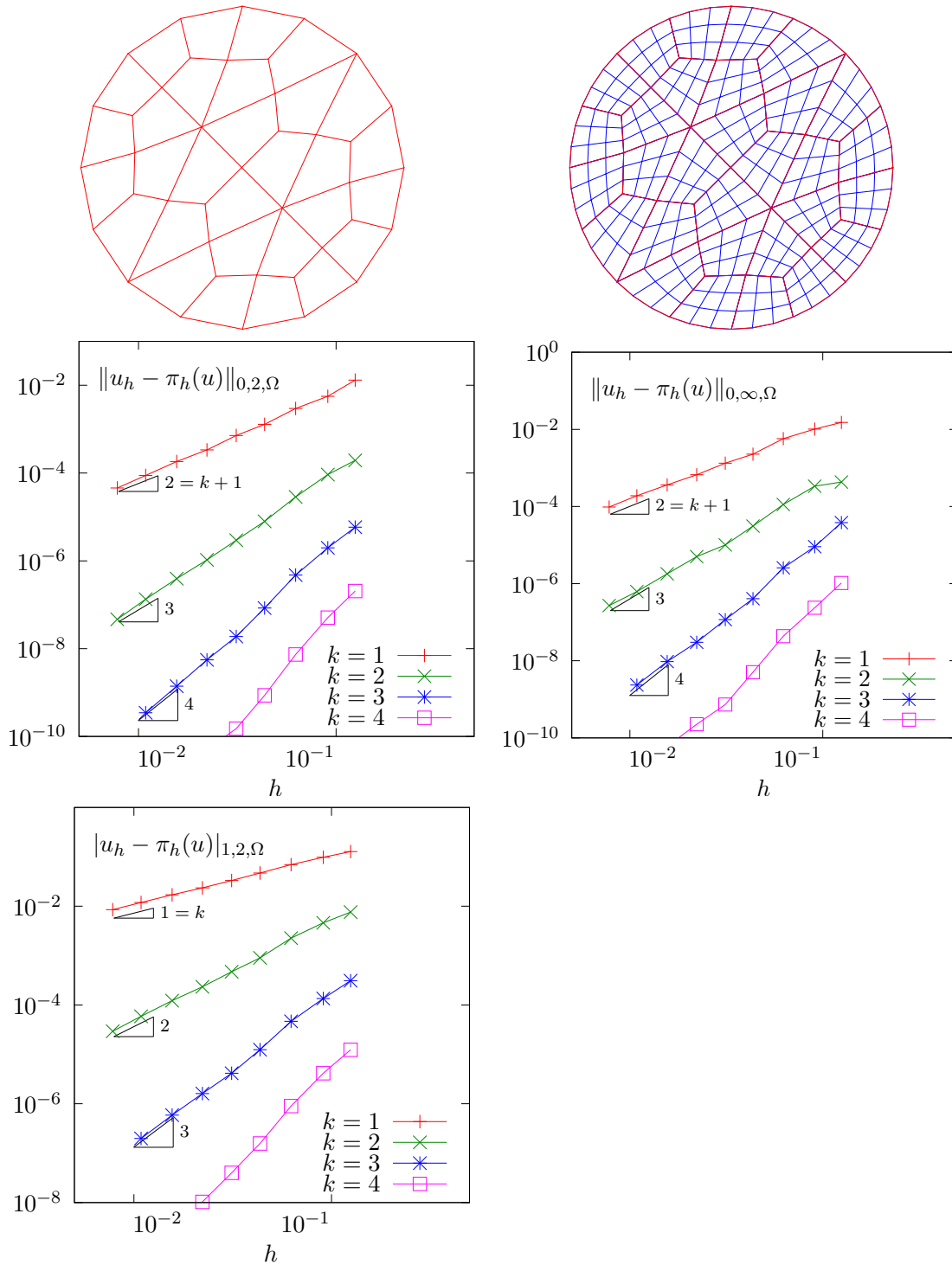


Figure 1.8: Curved domains (triangles): error analysis in  $L^2$ ,  $L^\infty$  and  $H^1$  norms.

Figure 1.9: Curved domains (quadrangles): error analysis in  $L^2$ ,  $L^\infty$  and  $H^1$  norms.

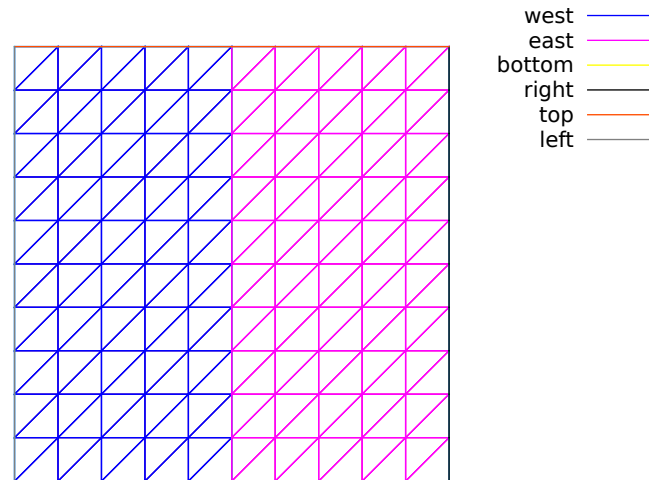


Figure 1.10: Transmission problem: the domain  $\Omega$  partition: ( $\Omega_{\text{west}}$  and  $\Omega_{\text{east}}$ ).

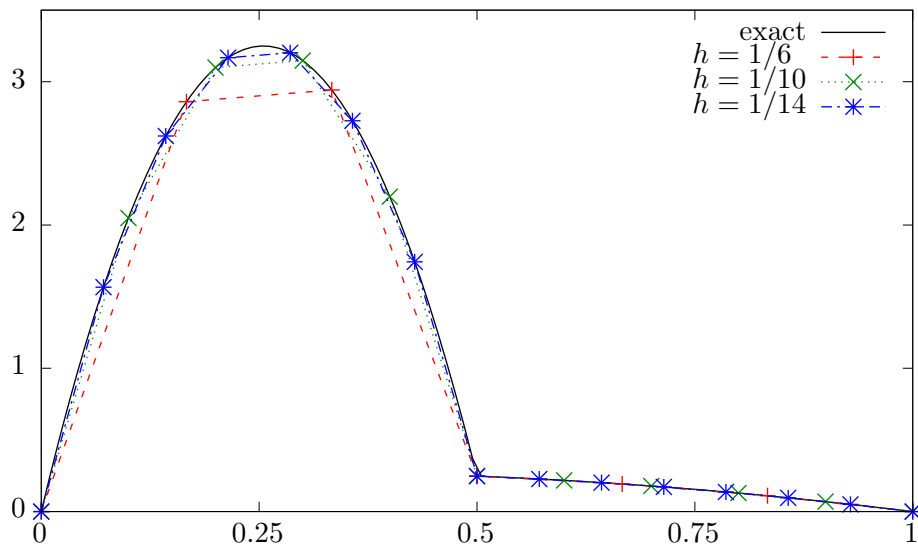


Figure 1.11: Transmission problem:  $u_h = \pi_h(u)$  ( $\varepsilon = 10^{-2}$ ,  $d = 1$ ,  $P_1$  approximation).





## Chapter 2

# Fluids and solids computations

### 2.1 The linear elasticity and the Stokes problems

#### 2.1.1 The linear elasticity problem

##### Formulation

The total Cauchy stress tensor expresses:

$$\sigma(\mathbf{u}) = \lambda \operatorname{div}(\mathbf{u}).I + 2\mu D(\mathbf{u}) \quad (2.1)$$

where  $\lambda$  and  $\mu$  are the Lamé coefficients. Here,  $D(\mathbf{u})$  denotes the symmetric part of the gradient operator and  $\operatorname{div}$  is the divergence operator. Let us consider the elasticity problem for the *embankment*, in  $\Omega = ]0, 1[^d$ ,  $d = 2, 3$ . The problem writes:

(P): find  $\mathbf{u} = (u_0, \dots, u_{d-1})$ , defined in  $\Omega$ , such that:

$$\begin{aligned} -\operatorname{div} \sigma(\mathbf{u}) &= \mathbf{f} \text{ in } \Omega, \\ \frac{\partial \mathbf{u}}{\partial \mathbf{n}} &= 0 \text{ on } \Gamma_{\text{top}} \cup \Gamma_{\text{right}}, \\ \mathbf{u} &= 0 \text{ on } \Gamma_{\text{left}} \cup \Gamma_{\text{bottom}}, \\ \mathbf{u} &= 0 \text{ on } \Gamma_{\text{front}} \cup \Gamma_{\text{back}}, \text{ when } d = 3 \end{aligned} \quad (2.2)$$

where  $\mathbf{f} = (0, -1)$  when  $d = 2$  and  $\mathbf{f} = (0, 0, -1)$  when  $d = 3$ . The Lamé coefficients are assumed to satisfy  $\mu > 0$  and  $\lambda + \mu > 0$ . Since the problem is linear, we can suppose that  $\mu = 1$  without any loss of generality. recall that, in order to avoid mistakes with the C++ style indexes, we denote by  $(x_0, \dots, x_{d-1})$  the Cartesian coordinate system in  $\mathbb{R}^d$ .

For  $d = 2$  we define the boundaries:

$$\begin{aligned} \Gamma_{\text{left}} &= \{0\} \times ]0, 1[, & \Gamma_{\text{right}} &= \{1\} \times ]0, 1[ \\ \Gamma_{\text{bottom}} &= ]0, 1[ \times \{0\}, & \Gamma_{\text{top}} &= ]0, 1[ \times \{1\} \end{aligned}$$

and for  $d = 3$ :

$$\begin{aligned} \Gamma_{\text{back}} &= \{0\} \times ]0, 1[^2, & \Gamma_{\text{front}} &= \{1\} \times ]0, 1[^2 \\ \Gamma_{\text{left}} &= ]0, 1[ \times \{0\} \times ]0, 1[, & \Gamma_{\text{right}} &= ]0, 1[ \times \{1\} \times ]0, 1[ \\ \Gamma_{\text{bottom}} &= ]0, 1[^2 \times \{0\}, & \Gamma_{\text{top}} &= ]0, 1[^2 \times \{1\} \end{aligned}$$

These boundaries are represented on Fig. 2.1.

The variational formulation of this problem expresses:

(VF): find  $\mathbf{u} \in \mathbf{V}$  such that:

$$a(\mathbf{u}, \mathbf{v}) = l(\mathbf{v}), \quad \forall \mathbf{v} \in \mathbf{V}, \quad (2.3)$$

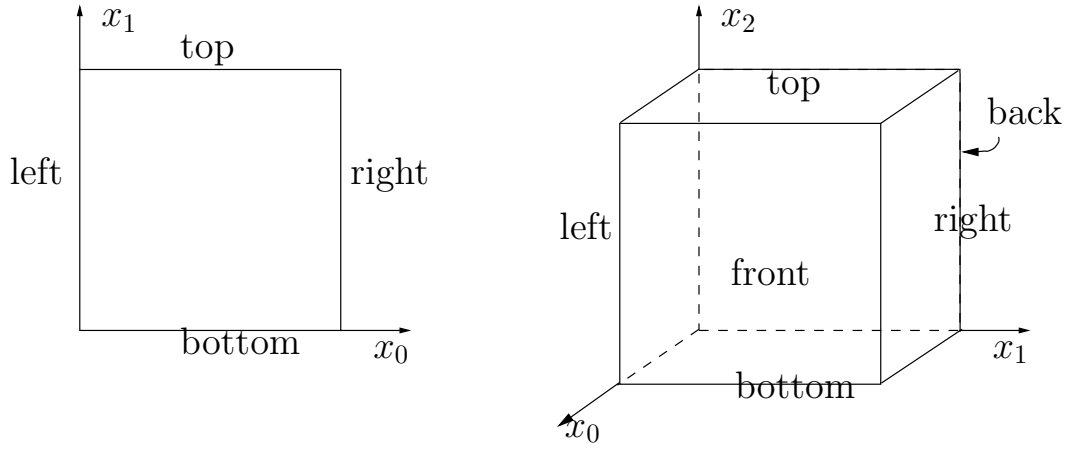


Figure 2.1: The boundary domains for the square and the cube.

where

$$\begin{aligned}
 a(\mathbf{u}, \mathbf{v}) &= \int_{\Omega} (\lambda \operatorname{div} \mathbf{u} \operatorname{div} \mathbf{v} + 2D(\mathbf{u}) : D(\mathbf{v})) \, dx, \\
 l(\mathbf{v}) &= \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, dx, \\
 \mathbf{V} &= \{\mathbf{v} \in (H^1(\Omega))^2; \mathbf{v} = 0 \text{ on } \Gamma_{\text{left}} \cup \Gamma_{\text{bottom}}\}, \text{ when } d = 2 \\
 \mathbf{V} &= \{\mathbf{v} \in (H^1(\Omega))^3; \mathbf{v} = 0 \text{ on } \Gamma_{\text{left}} \cup \Gamma_{\text{bottom}} \cup \Gamma_{\text{right}} \cup \Gamma_{\text{back}}\}, \text{ when } d = 3
 \end{aligned}$$

### Approximation

We introduce a mesh  $\mathcal{T}_h$  of  $\Omega$  and for  $k \geq 1$ , the following finite dimensional spaces:

$$\begin{aligned}
 \mathbf{X}_h &= \{\mathbf{v}_h \in (H^1(\Omega))^d; \mathbf{v}_{h/K} \in (P_k)^d, \forall K \in \mathcal{T}_h\}, \\
 \mathbf{V}_h &= \mathbf{X}_h \cap \mathbf{V}
 \end{aligned}$$

The approximate problem writes:

$(VF)_h$ : find  $\mathbf{u}_h \in \mathbf{V}_h$  such that:

$$a(\mathbf{u}_h, \mathbf{v}_h) = l(\mathbf{v}_h), \quad \forall \mathbf{v}_h \in \mathbf{V}_h$$

Example file 2.1: embankment.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "embankment.icc"
5 int main(int argc, char**argv) {
6     environment rheolef(argc, argv);
7     geo omega (argv[1]);
8     space Xh = embankment_space (omega, argv[2]);
9     Float lambda = (argc > 3) ? atof(argv[3]) : 1;
10    size_t d = omega.dimension();
11    point f (0,0,0);
12    f[d-1] = -1;
13    trial u (Xh); test v (Xh);
14    form a = integrate (lambda*div(u)*div(v) + 2*ddot(D(u),D(v)));
15    field lh = integrate (dot(f,v));
16    solver sa (a.uu());
17    field uh (Xh, 0);
18    uh.set_u() = sa.solve (lh.u() - a.ub()*uh.b());
19    dout << catchmark("inv_lambda") << 1/lambda << endl
20         << catchmark("u") << uh;
21 }

```

Example file 2.2: embankment.icc

```

1 space embankment_space (const geo& omega, string approx) {
2     space Xh (omega, approx, "vector");
3     Xh.block("left");
4     if (omega.dimension() >= 2)
5         Xh.block("bottom");
6     if (omega.dimension() == 3) {
7         Xh.block("right");
8         Xh.block("back");
9     }
10    return Xh;
11 }

```

## Comments

The space is defined in a separate file ‘`embankment.icc`’, since it will be reused in others examples along this chapter:

```
space Vh (omega, "P2", "vector");
```

Note here the multi-component specification “`vector`” as a supplementary argument to the `space` constructor. The boundary condition contain a special cases for bi- and tridimensional cases. The right-hand-side  $f_h$  represents the dimensionless gravity forces, oriented on the vertical axis: the last component of  $f_h$  is set to  $-1$  as:

```
f_h [d-1] = -1;
```

The code for the bilinear form  $a(.,.)$  and the linear one  $l(.)$  are closed to their mathematical definitions:

```
form a = integrate (lambda*div(u)*div(v) + 2*ddot(D(u),D(v)));
field lh = integrate (dot(f,v));
```

Finally, the  $1/\lambda$  parameter and the multi-field result are printed, using mark labels, thanks to the `catchmark` stream manipulator. Labels are convenient for post-processing purpose, as we will see in the next paragraph.

## How to run the program

Compile the program as usual (see page 14):

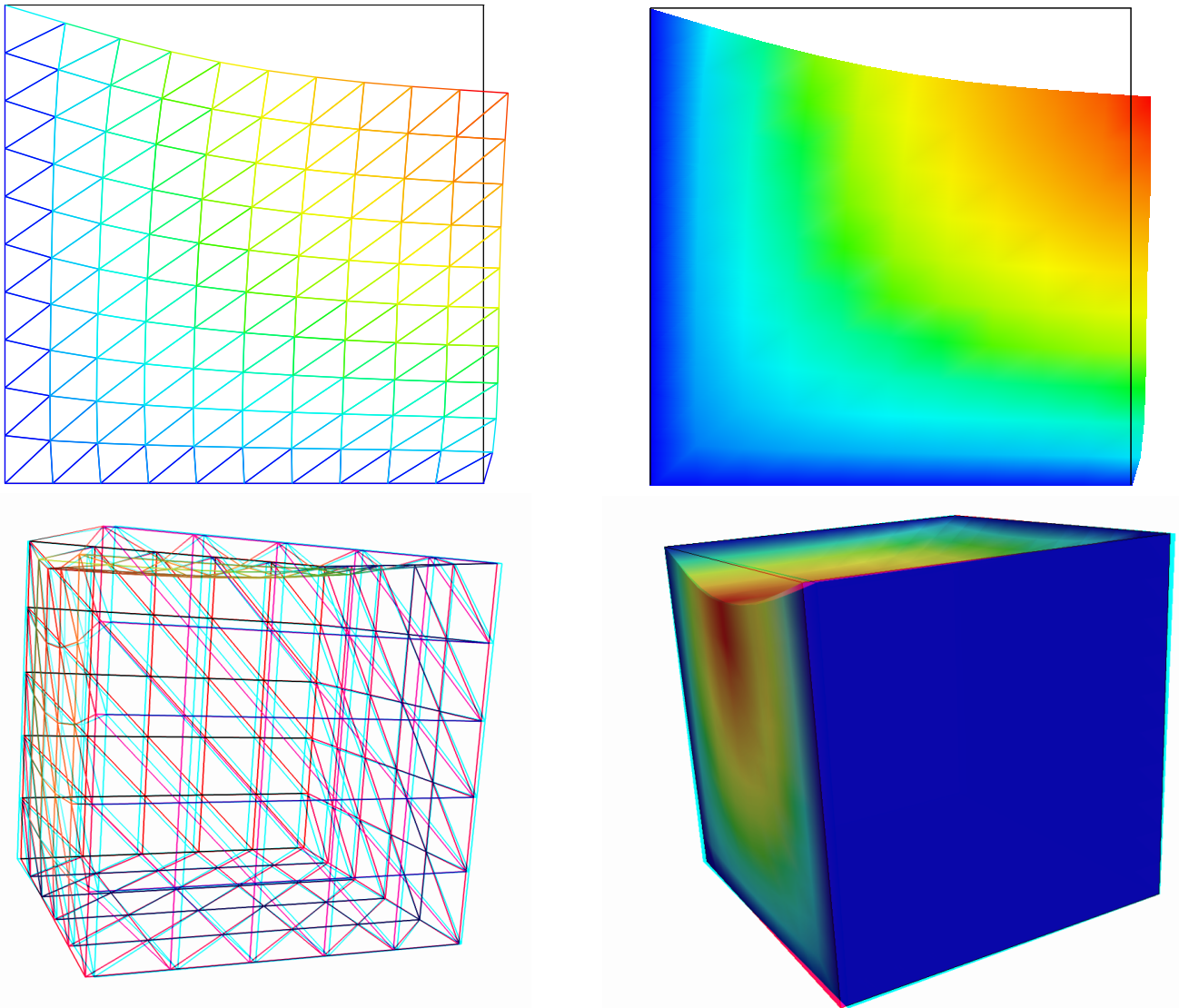


Figure 2.2: The linear elasticity for  $\lambda = 1$  and  $d = 2$  and  $d = 3$ : both wireframe and filled surfaces ; stereoscopic anaglyph mode for 3D solutions.

```
make embankment
```

and enter the commands:

```
mkgeo_grid -t 10 > square.geo
geo square.geo
```

The triangular mesh has four boundary domains, named `left`, `right`, `top` and `bottom`. Then, enter:

```
./embankment square.geo P1 > square-P1.field
```

The previous command solves the problem for the corresponding mesh and writes the multi-component solution in the `.field` file format. Run the deformation vector field visualization:

```
field square-P1.field
field square-P1.field -nofill
```

It bases on the default `paraview` render for 2D and 3D geometries. The view is shown on Fig. 2.2. If you are in trouble with `paraview`, you can switch to the simpler `gnuplot` render:

```
field square-P1.field -nofill -gnuplot
```

The unix manual for the `field` command is available as:

```
man field
```

A specific field component can be also selected for a scalar visualization:

```
field -comp 0 square-P1.field
field -comp 1 square-P1.field
```

Next, perform a  $P_2$  approximation of the solution:

```
./embankment square.geo P2 > square-P2.field
field square-P2.field -nofill
```

Finally, let us consider the three dimensional case

```
mkgeo_grid -T 10 > cube.geo
./embankment cube.geo P1 > cube-P1.field
field cube-P1.field -stereo
field cube-P1.field -stereo -fill
```

The two last commands show the solution in 3D stereoscopic anaglyph mode. The graphic is represented on Fig. 2.2. The  $P_2$  approximation writes:

```
./embankment cube.geo P2 > cube-P2.field
field cube-P2.field
```

## 2.1.2 Computing the stress tensor

### Formulation and approximation

The following code computes the total Cauchy stress tensor, reading the Lamé coefficient  $\lambda$  and the deformation field  $\mathbf{u}_h$  from a `.field` file. Let us introduce:

$$T_h = \{\tau_h \in (L^2(\Omega))^{d \times d}; \tau_h = \tau_h^T \text{ and } \tau_{h;ij/K} \in P_{k-1}, \forall K \in \mathcal{T}_h, 1 \leq i, j \leq d\}$$

This computation expresses:

*find*  $\sigma_h$  such that:

$$m(\sigma_h, \tau) = b(\tau, \mathbf{u}_h), \forall \tau \in T_h$$

where

$$\begin{aligned} m(\sigma, \tau) &= \int_{\Omega} \sigma : \tau \, dx, \\ b(\tau, \mathbf{u}) &= \int_{\Omega} (2D(\mathbf{u}) : \tau \, dx + \lambda \operatorname{div}(\mathbf{u}) \operatorname{tr}(\tau)) \, dx, \end{aligned}$$

where  $\operatorname{tr}(\tau) = \sum_{i=1}^d \tau_{ii}$  is the trace of the tensor  $\tau$ .

Example file 2.3: stress.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 int main(int argc, char** argv) {
5     environment rheolef (argc,argv);
6     Float inv_lambda;
7     field uh;
8     din >> catchmark("inv_lambda") >> inv_lambda
9         >> catchmark("u") >> uh;
10    const geo& omega = uh.get_geo();
11    const space& Xh = uh.get_space();
12    string grad_approx = "P" + itos(Xh.degree()-1) + "d";
13    space Th (omega, grad_approx, "tensor");
14    size_t d = omega.dimension();
15    tensor I = tensor::eye (d);
16    field sigma_h = (inv_lambda == 0) ?
17        interpolate (Th, 2*D(uh)) :
18        interpolate (Th, 2*D(uh) + (1/inv_lambda)*div(uh)*I);
19    dout << catchmark("s") << sigma_h;
20 }

```

### Comments

In order to our code `stress.cc` to apply also for the forthcoming incompressible case  $\lambda = +\infty$ , the Lamé coefficient is introduced as  $1/\lambda$ . Its value is zero in the incompressible case. By this way, the previous code applies for any deformation field, and is not restricted to our *embankment* problem. The stress tensor is obtained by a direct interpolation of the  $u_h$  first derivatives. As  $u_h$  is continuous and piecewise polynomial  $P_k$ , its derivatives are also piecewise polynomials with degree  $k - 1$ , but *discontinuous* at inter-element boundaries : this approximation is denoted as  $P_{k-1,d}$ . Thus, the stress tensor belongs to the space  $T_h$  with the  $P_{k-1,d}$  element.

### How to run the program

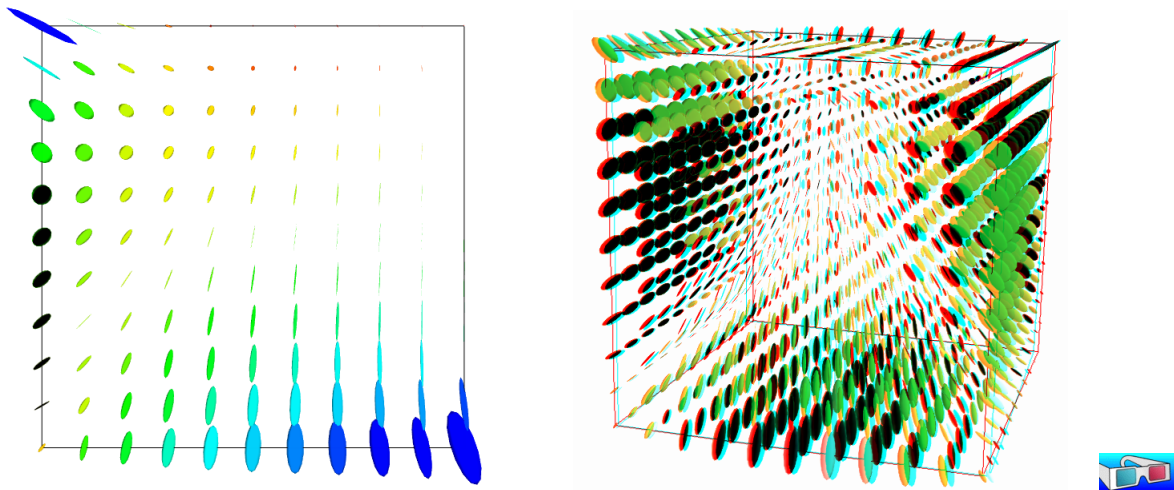


Figure 2.3: The stress tensor visualization (linear elasticity  $\lambda = 1$ ).

First, compile the program:

```
make stress
```

The visualization for the stress tensor as ellipses writes:

```
./stress < square-P1.field > square-stress-P1.field
./stress < square-P2.field > square-stress-P2.field
field square-stress-P1.field
field square-stress-P2.field
```

The visualization based on `paraview` requires the `TensorGlyph` plugin<sup>1</sup>If this plugin is not available on our installation, turns to the `mayavi`<sup>2</sup>render:

```
field square-stress-P1.field -proj -mayavi
field square-stress-P2.field -proj -mayavi
```

Recall that the stress, as a derivative of the deformation, is P0 (resp. P1d) and discontinuous when the deformation is P1 (resp. P2) and continuous. The approximate stress tensor field is projected on a continuous piecewise linear space, using the `-proj` option. Conversely, the 3D visualization bases on ellipsoids:

```
./stress < cube-P1.field > cube-stress-P1.field
field cube-stress-P1.field -stereo
```

Also, if the `TensorGlyph` plugin is not available in your `paraview` installation, and the `-mayavi` option in the previous command. You can observe a discontinuous constant or piecewise linear representation of the approximate stress component  $\sigma_{01}$  (see Fig. 2.4):

```
field square-stress-P1.field -comp 01
field square-stress-P2.field -comp 01 -elevation
field square-stress-P2.field -comp 01 -elevation -stereo
```

Notice that the `-stereo` implies the `paraview` render: this feature available with `paraview` and `mayavi` renders. The approximate stress field can be also projected on a continuous piecewise space:

```
field square-stress-P2.field -comp 01 -elevation -proj
```

The tridimensional case writes simply (see Fig. 2.4):

```
./stress < cube-P1.field > cube-stress-P1.field
./stress < cube-P2.field > cube-stress-P2.field
field cube-stress-P1.field -comp 01 -stereo
field cube-stress-P2.field -comp 01 -stereo
```

and also the P1-projected versions write:

```
field cube-stress-P1.field -comp 01 -stereo -proj
field cube-stress-P2.field -comp 01 -stereo -proj
```

These operations can be repeated for each  $\sigma_{ij}$  components and for both P1 and P2 approximation of the deformation field.

### 2.1.3 Mesh adaptation

The main principle of the auto-adaptive mesh writes [13, 21, 46, 79, 108]:

<sup>1</sup> [http://paraview.org/Wiki/ParaView/User\\_Created\\_Plugins](http://paraview.org/Wiki/ParaView/User_Created_Plugins) The tensor glyph paraview plugin is still not part of the paraview distribution and its installation requires a compilation from paraview source code.

<sup>2</sup> On some Debian and Ubuntu distributions, the `mayavi2` package installation conflicts with `paraview` one: it requires to delete the `paraview-python` package and it is not yet possible to have these both tools installed at the same time.



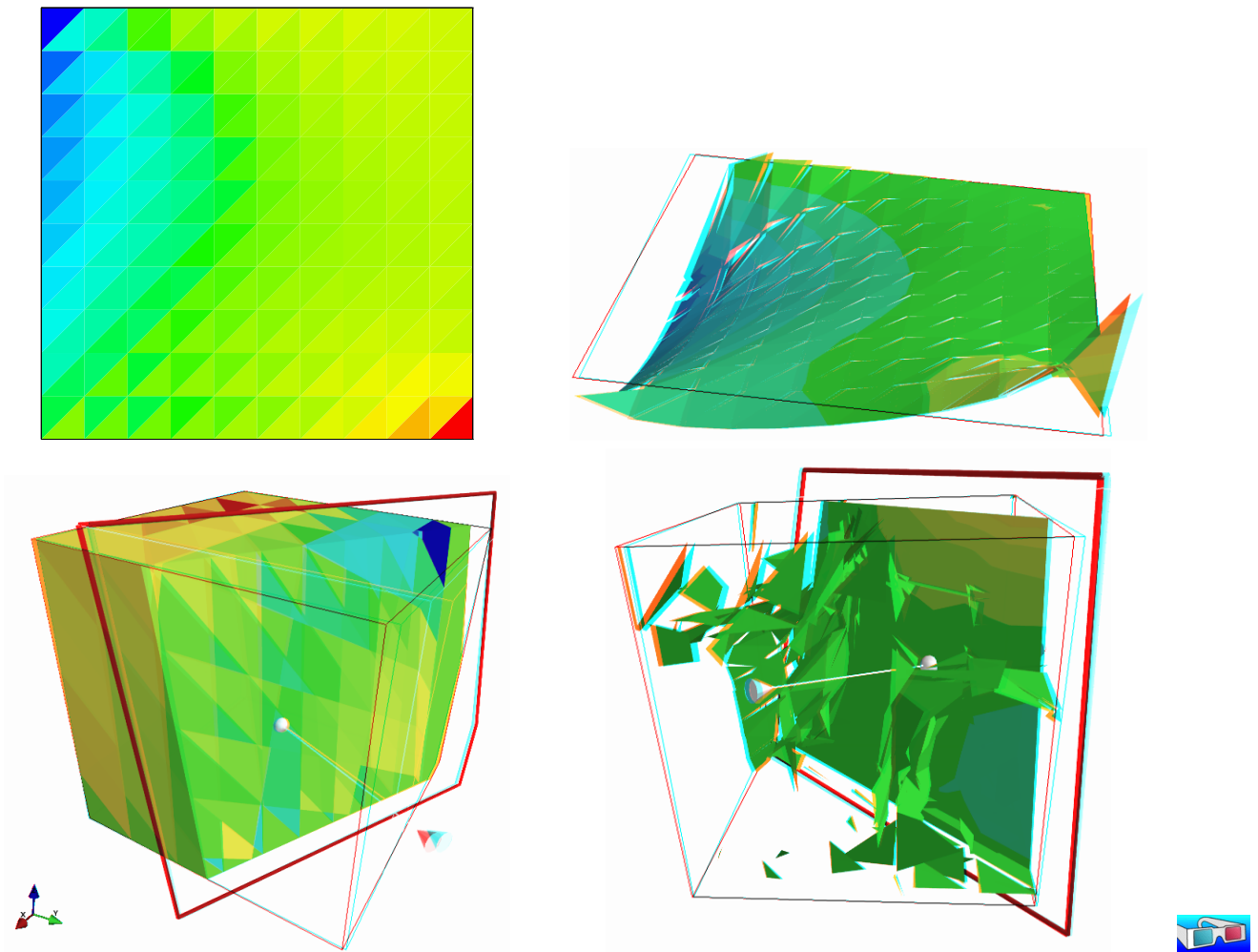


Figure 2.4: The  $\sigma_{01}$  stress component (linear elasticity  $\lambda = 1$ ):  $d = 2$  (top) and  $d = 3$  (bottom) ;  $P_0$  (left) and  $P_1$  discontinuous approximation (right).

```

cin >> omega;
uh = solve(omega);
for (unsigned int i = 0; i < n; i++) {
    ch = criterion(uh);
    omega = adapt(ch);
    uh = solve(omega);
}

```

The initial mesh is used to compute a first solution. The adaptive loop compute an *adaptive criterion*, denoted by  $ch$ , that depends upon the problem under consideration and the polynomial approximation used. Then, a new mesh is generated, based on this criterion. A second solution on an adapted mesh can be constructed. The adaptation loop converges generally in roughly 5 to 20 iterations.

Let us apply this principle to the elasticity problem.

Example file 2.4: embankment\_adapt.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "elasticity_solve.icc"
5 #include "elasticity_criterion.icc"
6 #include "embankment.icc"
7 int main(int argc, char**argv) {
8     environment rheolef (argc, argv);
9     const Float lambda = 1;
10    geo omega (argv[1]);
11    adapt_option options;
12    string approx = (argc > 2) ? argv[2] : "P1";
13    options.err = (argc > 3) ? atof(argv[3]) : 5e-3;
14    size_t n_adapt = (argc > 4) ? atoi(argv[4]) : 5;
15    options.hmin = 0.004;
16    for (size_t i = 0; true; i++) {
17        space Xh = embankment_space (omega, approx);
18        field uh = elasticity_solve (Xh, lambda);
19        odistream of (omega.name(), "field");
20        of << catchmark("lambda") << lambda << endl
21         << catchmark("u") << uh;
22        if (i == n_adapt) break;
23        field ch = elasticity_criterion (lambda,uh);
24        omega = adapt(ch, options);
25        odistream og (omega.name(), "geo");
26        og << omega;
27    }
28 }

```

Example file 2.5: elasticity\_solve.icc

```

1 field elasticity_solve (const space& Xh, Float lambda) {
2     size_t d = Xh.get_geo().dimension();
3     point f (0,0,0);
4     f[d-1] = -1;
5     trial u (Xh); test v (Xh);
6     field lh = integrate (dot(f,v));
7     form a = integrate (lambda*div(u)*div(v) + 2*ddot(D(u),D(v)));
8     solver sa (a.uu());
9     field uh (Xh, 0);
10    uh.set_u() = sa.solve (lh.u() - a.ub()*uh.b());
11    return uh;
12 }

```

Example file 2.6: elasticity\_criterion.icc

```

1 field elasticity_criterion (Float lambda, const field& uh) {
2     string grad_approx = "P" + itos(uh.get_space().degree()-1) + "d";
3     space Xh (uh.get_geo(), grad_approx);
4     if (grad_approx == "P0d") return interpolate (Xh, norm(uh));
5     space T0h (uh.get_geo(), grad_approx);
6     size_t d = uh.get_geo().dimension();
7     tensor I = tensor::eye (d);
8     return interpolate (T0h, sqrt(2*norm2(D(uh)) + lambda*sqr(div(uh))));
9 }

```

### Comments

The criterion is here:

$$c_h = \begin{cases} |\mathbf{u}_h| & \text{when using } P_1 \\ (\sigma(\mathbf{u}_h) : D(\mathbf{u}_h))^{1/2} & \text{when using } P_2 \end{cases}$$

The `elasticity_criterion` function compute it as

```
return interpolate (Xh, norm(uh));
```

when using  $P_1$ , and as

```
return interpolate (T0h, sqrt(2*norm2(D(uh)) + lambda*sqr(div(uh))));
```

when using  $P_2$ . The `sqr` function returns the square of a scalar. Conversely, the `norm2` function returns the square of the norm. In the `min` program, the result of the `elasticity_criterion` function is sent to the `adapt` function:

```
field ch = elasticity_criterion (lambda, uh);
omega = adapt (ch, options);
```

The `adapt_option` declaration is used by **Rheolef** to send options to the mesh generator. The `err` parameter controls the error via the edge length of the mesh: the smaller it is, the smaller the edges of the mesh are. In our example, is set by default to one. Conversely, the `hmin` parameter controls minimal edge length.

### How to run the program

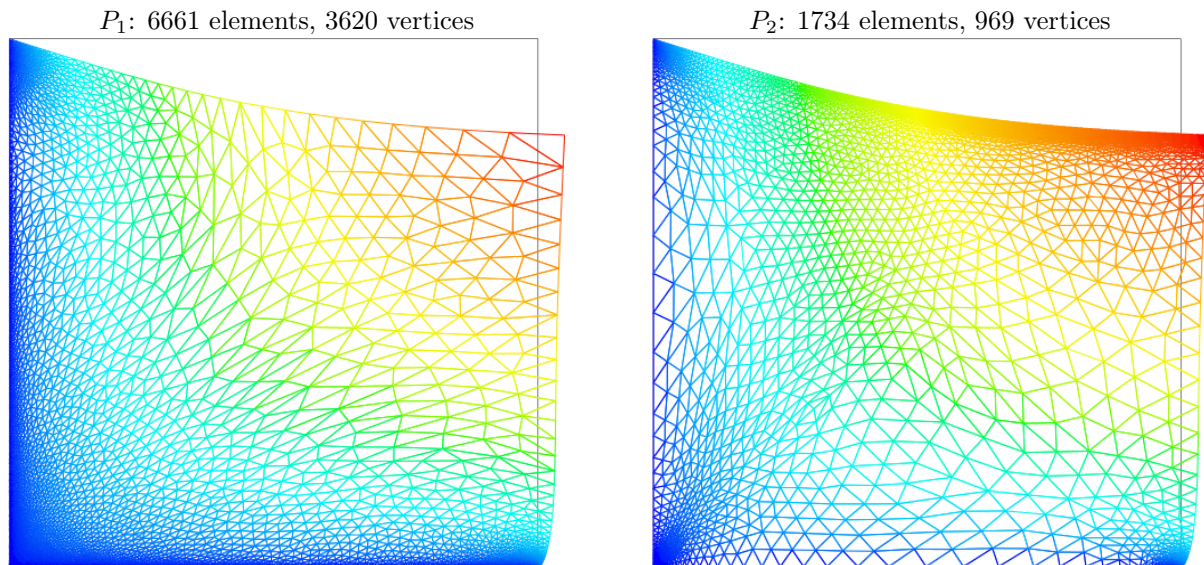


Figure 2.5: Adapted meshes: the deformation visualization for  $P_1$  and  $P_2$  approximations.

The compilation command writes:

```
make embankment_adapt
```

The mesh loop adaptation is initiated from a `bamg` mesh (see also appendix A.2.1).

```
bamg -g square.bamgcad -o square.bamg
bamg2geo square.bamg square.dmn > square.geo
./embankment_adapt square P1 2e-2
```

The last command line argument is the target error. The code performs a loop of five mesh adaptations: the corresponding meshes are stored in files, from `square-001.geo.gz` to `square-005.geo.gz`, and the associated solutions in files, from `square-001.field.gz` to `square-005.field.gz`. The additional `.gz` suffix expresses that the files are compressed using `gzip`.

```
geo square-005.geo
field square-005.field -nofill
```

Note that the ‘.gz’ suffix is automatically assumed by the `geo` and the `field` commands.  
For a piecewise quadratic approximation:

```
./embankment_adapt square P2 5e-3
```

Then, the visualization writes:

```
geo square-005.geo
field square-005.field -nofill
```

A one-dimensional mesh adaptive procedure is also possible:

```
gmsh -1 line.mshcad -o line.msh
msh2geo line.msh > line.geo
geo line.geo
./embankment_adapt line P2
geo line-005.geo
field line-005.field -comp 0 -elevation
```

The three-dimensional extension of this mesh adaptive procedure is in development.

## 2.1.4 The Stokes problem

### Formulation

Let us consider the Stokes problem for the driven cavity in  $\Omega = ]0, 1[^d$ ,  $d = 2, 3$ . The problem writes:

(S) find  $\mathbf{u} = (u_0, \dots, u_{d-1})$  and  $p$  defined in  $\Omega$  such that:

$$\begin{aligned} -\operatorname{div}(2D(\mathbf{u})) + \nabla p &= \mathbf{0} \text{ in } \Omega, \\ -\operatorname{div} \mathbf{u} &= 0 \text{ in } \Omega, \\ \mathbf{u} &= (1, 0) \text{ on } \Gamma_{\text{top}}, \\ \mathbf{u} &= \mathbf{0} \text{ on } \Gamma_{\text{left}} \cup \Gamma_{\text{right}} \cup \Gamma_{\text{bottom}}, \\ \frac{\partial u_0}{\partial \mathbf{n}} &= \frac{\partial u_1}{\partial \mathbf{n}} = u_2 = 0 \text{ on } \Gamma_{\text{back}} \cup \Gamma_{\text{front}} \text{ when } d = 3, \end{aligned}$$

where  $D(\mathbf{u}) = (\nabla \mathbf{u} + \nabla \mathbf{u}^T)/2$ . The boundaries are represented on Fig. 2.1, page 42.

The variational formulation of this problem expresses:

(VFS) find  $\mathbf{u} \in \mathbf{V}(1)$  and  $p \in L_0^2(\Omega)$  such that:

$$\begin{aligned} a(\mathbf{u}, \mathbf{v}) + b(\mathbf{v}, p) &= 0, \quad \forall \mathbf{v} \in \mathbf{V}(0), \\ b(\mathbf{u}, q) &= 0, \quad \forall q \in L_0^2(\Omega), \end{aligned}$$

where

$$a(\mathbf{u}, \mathbf{v}) = \int_{\Omega} 2D(\mathbf{u}) : D(\mathbf{v}) \, dx,$$

$$b(\mathbf{v}, q) = - \int_{\Omega} \operatorname{div}(\mathbf{v}) q \, dx.$$

$$\mathbf{V}(\alpha) = \{\mathbf{v} \in (H^1(\Omega))^2; \mathbf{v} = \mathbf{0} \text{ on } \Gamma_{\text{left}} \cup \Gamma_{\text{right}} \cup \Gamma_{\text{bottom}} \text{ and } \mathbf{v} = (\alpha, 0) \text{ on } \Gamma_{\text{top}}\}, \text{ when } d = 2,$$

$$\mathbf{V}(\alpha) = \{\mathbf{v} \in (H^1(\Omega))^3; \mathbf{v} = \mathbf{0} \text{ on } \Gamma_{\text{left}} \cup \Gamma_{\text{right}} \cup \Gamma_{\text{bottom}},$$

$$\mathbf{v} = (\alpha, 0, 0) \text{ on } \Gamma_{\text{top}} \text{ and } v_2 = 0 \text{ on } \Gamma_{\text{back}} \cup \Gamma_{\text{front}}\}, \text{ when } d = 3,$$

$$L_0^2(\Omega) = \{q \in L^2(\Omega); \int_{\Omega} q \, dx = 0\}.$$

## Approximation

The Taylor-Hood [49] finite element approximation of the Stokes problem is considered. We introduce a mesh  $\mathcal{T}_h$  of  $\Omega$  and the following finite dimensional spaces:

$$\begin{aligned}\mathbf{X}_h &= \{\mathbf{v} \in (H^1(\Omega))^d; \mathbf{v}_{/K} \in (P_2)^d, \forall K \in \mathcal{T}_h\}, \\ \mathbf{V}_h(\alpha) &= \mathbf{X}_h \cap \mathbf{V}(\alpha), \\ Q_h &= \{q \in L^2(\Omega) \cap C^0(\bar{\Omega}); q_{/K} \in P_1, \forall K \in \mathcal{T}_h\},\end{aligned}$$

The approximate problem writes:

$(VFS)_h$  find  $\mathbf{u}_h \in \mathbf{V}_h(1)$  and  $p \in Q_h$  such that:

$$\begin{aligned}a(\mathbf{u}_h, \mathbf{v}) + b(\mathbf{v}, p_h) &= 0, \forall \mathbf{v} \in \mathbf{V}_h(0), \\ b(\mathbf{u}_h, q) &= 0, \forall q \in Q_h.\end{aligned}\tag{2.4}$$

Example file 2.7: cavity.icc

```

1 struct cavity {
2   static space velocity_space (const geo& omega, string approx) {
3     space Xh (omega, approx, "vector");
4     Xh.block("top"); Xh.block("bottom");
5     if (omega.dimension() == 3) {
6       Xh.block("back"); Xh.block("front");
7       Xh[1].block("left"); Xh[1].block("right");
8     } else {
9       Xh.block("left"); Xh.block("right");
10    }
11    return Xh;
12  }
13  static field velocity_field (const space& Xh, Float alpha=1) {
14    field uh (Xh, 0.);
15    uh[0]["top"] = alpha;
16    return uh;
17  }
18  static space streamf_space (geo omega, string approx) {
19    string valued = (omega.dimension() == 3) ? "vector" : "scalar";
20    space Ph (omega, approx, valued);
21    Ph.block("top"); Ph.block("bottom");
22    if (omega.dimension() == 3) {
23      Ph.block("back"); Ph.block("front");
24    } else {
25      Ph.block("left"); Ph.block("right");
26    }
27    return Ph;
28  }
29  static field streamf_field (space Ph) {
30    return field(Ph, 0);
31  }
32 };

```

Example file 2.8: stokes\_cavity.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "cavity.icc"
5 int main(int argc, char**argv) {
6     environment rheolef (argc, argv);
7     geo omega (argv[1]);
8     space Xh = cavity::velocity_space (omega, "P2");
9     space Qh (omega, "P1");
10    trial u (Xh), p (Qh); test v (Xh), q (Qh);
11    form a = integrate (2*ddot(D(u),D(v)));
12    form b = integrate (-div(u)*q);
13    form mp = integrate (p*q);
14    field uh = cavity::velocity_field (Xh, 1);
15    field ph (Qh, 0.);
16    solver_abtb stokes (a.uu(), b.uu(), mp.uu());
17    stokes.solve (-(a.ub()*uh.b()), -(b.ub()*uh.b()),
18                uh.set_u(), ph.set_u());
19    dout << catchmark("u") << uh
20          << catchmark("p") << ph;
21 }

```

## Comments

The spaces and boundary conditions and grouped in specific functions, defined in file ‘cavity.icc’. This file is suitable for a future re-usage. Next, forms are defined as usual, in file ‘stokes\_cavity.cc’.

The problem admits the following matrix form:

$$\begin{pmatrix} a.uu & \text{trans}(b.uu) \\ b.uu & 0 \end{pmatrix} \begin{pmatrix} uh.u \\ ph.u \end{pmatrix} = \begin{pmatrix} -a.ub * uh.b \\ -b.ub * uh.b \end{pmatrix}$$

An initial value for the pressure field is provided:

```
field ph (Qh, 0);
```

The main Stokes solver call writes:

```
solver_abtb stokes (a.uu(), b.uu(), mp.uu());
stokes.solve (-(a.ub()*uh.b()), -(b.ub()*uh.b()),
             uh.set_u(), ph.set_u());
```

For tridimensional geometries ( $d = 3$ ), this system is solved by the preconditioned conjugate gradient algorithm. the preconditioner is here the mass matrix `mp.uu` for the pressure: as showed in [53], the number of iterations need by the conjugate gradient algorithm to reach a given precision is then independent of the mesh size. For more details, see the Rheolef reference manual related to mixed solvers, available e.g. via the unix command:

```
man solver_abtb
```

When  $d = 2$ , it is interesting to turn to direct methods and factorize the whole matrix of the linear system. As the pressure is defined up to a constant, the whole matrix is singular. By adding a Lagrange multiplier that impose a null average pressure value, the system becomes regular and the modified matrix can be inverted. Such a technique has already been presented in section 1.4 for the Neumann-Laplace problem. Finally, the choice between iterative and direct algorithm for the Stokes solver is automatically done, regarding the geometry dimension.

## How to run the program

We assume that the previous code is contained in the file ‘stokes\_cavity.cc’. Then, compile the program as usual (see page 14):

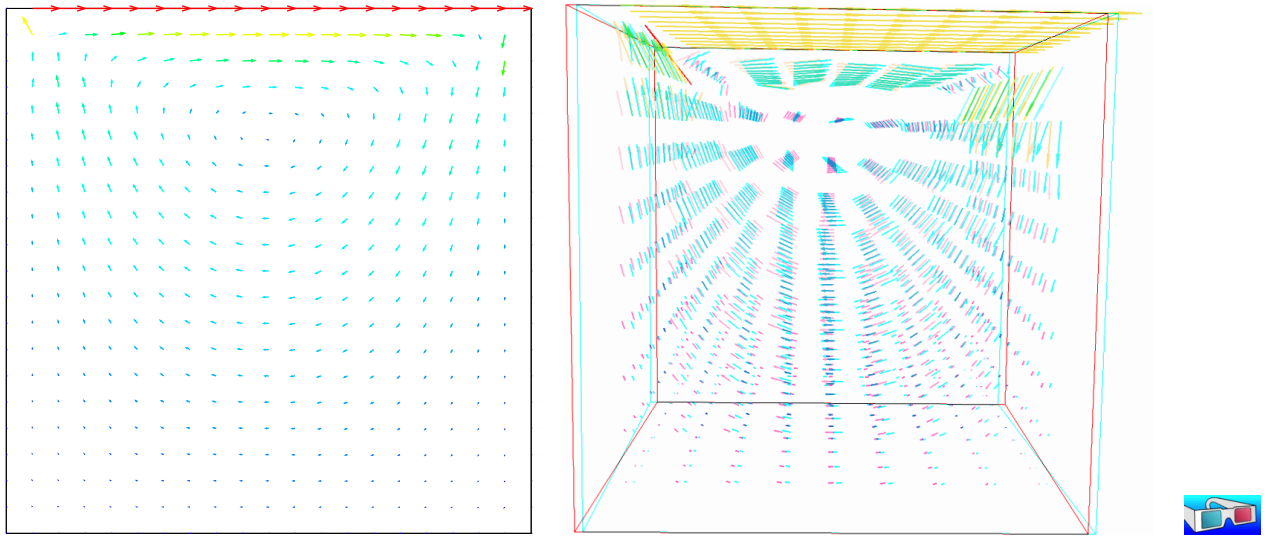


Figure 2.6: The velocity visualization for  $d = 2$  and  $d = 3$  with stereo anaglyph.

```
make stokes_cavity
```

and enter the commands:

```
mkgeo_grid -t 10 > square.geo
./stokes_cavity square > square.field
```

The previous command solves the problem for the corresponding mesh and writes the solution in a '.field' file. Run the velocity vector visualization :

```
field square.field -velocity
```

Run also some scalar visualizations:

```
field square.field -comp 0
field square.field -comp 1
field square.field -mark p
```

Note the `-mark` option to the `field` command: the file reader jumps to the label and then starts to read the selected field. Next, perform another computation on a finer mesh:

```
mkgeo_grid -t 20 > square-20.geo
./stokes_cavity square-20.geo > square-20.field
```

and observe the convergence.

Finally, let us consider the three dimensional case:

```
mkgeo_grid -T 5 > cube.geo
./stokes_cavity cube.geo > cube.field
```

and the corresponding visualization:

```
field cube.field -velocity -scale 3
field cube.field -comp 0
field cube.field -comp 1
field cube.field -comp 2
field cube.field -mark p
```

### 2.1.5 Computing the vorticity

#### Formulation and approximation

When  $d = 2$ , we define [41, page 30] for any distributions  $\phi$  and  $\mathbf{v}$ :

$$\begin{aligned}\mathbf{curl} \phi &= \left( \frac{\partial \phi}{\partial x_1}, -\frac{\partial \phi}{\partial x_0} \right), \\ \mathbf{curl} \mathbf{v} &= \frac{\partial v_1}{\partial x_0} - \frac{\partial v_0}{\partial x_1},\end{aligned}$$

and when  $d = 3$ :

$$\mathbf{curl} \mathbf{v} = \left( \frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2}, \frac{\partial v_0}{\partial x_2} - \frac{\partial v_2}{\partial x_0}, \frac{\partial v_1}{\partial x_0} - \frac{\partial v_0}{\partial x_1} \right)$$

Let  $\mathbf{u}$  be the solution of the Stokes problem ( $S$ ). The vorticity is defined by:

$$\begin{aligned}\omega &= \mathbf{curl} \mathbf{u} \quad \text{when } d = 2, \\ \boldsymbol{\omega} &= \mathbf{curl} \mathbf{u} \quad \text{when } d = 3.\end{aligned}$$

Since the approximation of the velocity is piecewise quadratic, we are looking for a discontinuous piecewise linear vorticity field that belongs to:

$$\begin{aligned}Y_h &= \{ \xi \in L^2(\Omega); \xi_{/K} \in P_1, \forall K \in \mathcal{T}_h \}, \quad \text{when } d = 2 \\ \mathbf{Y}_h &= \{ \boldsymbol{\xi} \in (L^2(\Omega))^3; \xi_{i/K} \in P_1, \forall K \in \mathcal{T}_h \}, \quad \text{when } d = 3\end{aligned}$$

The approximate variational formulation writes:

$$\begin{aligned}\omega_h \in Y_h, \quad \int_{\Omega} \omega_h \xi \, dx &= \int_{\Omega} \mathbf{curl} \mathbf{u}_h \xi \, dx, \quad \forall \xi \in Y_h \quad \text{when } d = 2, \\ \boldsymbol{\omega} \in \mathbf{Y}_h, \quad \int_{\Omega} \boldsymbol{\omega}_h \cdot \boldsymbol{\xi} \, dx &= \int_{\Omega} \mathbf{curl} \mathbf{u}_h \cdot \boldsymbol{\xi} \, dx, \quad \forall \boldsymbol{\xi} \in \mathbf{Y}_h \quad \text{when } d = 3.\end{aligned}$$

Example file 2.9: vorticity.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 int main(int argc, char** argv) {
5     environment rheolef (argc, argv);
6     field uh;
7     din >> uh;
8     const space& Xh = uh.get_space();
9     string grad_approx = "P" + itos(Xh.degree()-1) + "d";
10    string valued = (uh.size() == 3) ? "vector" : "scalar";
11    space Lh (uh.get_geo(), grad_approx, valued);
12    field curl_uh = interpolate (Lh, curl(uh));
13    dout << catchmark("w") << curl_uh;
14 }

```

#### Comments

As for the stress tensor (see `stress.cc`, page 46), the vorticity is obtained by a direct interpolation of the  $u_h$  first derivatives and its approximation is *discontinuous* at inter-element boundaries.

#### How to run the program

For  $d = 2$ , just enter:

```

make vorticity
./vorticity < square.field | field -elevation -stereo -

```



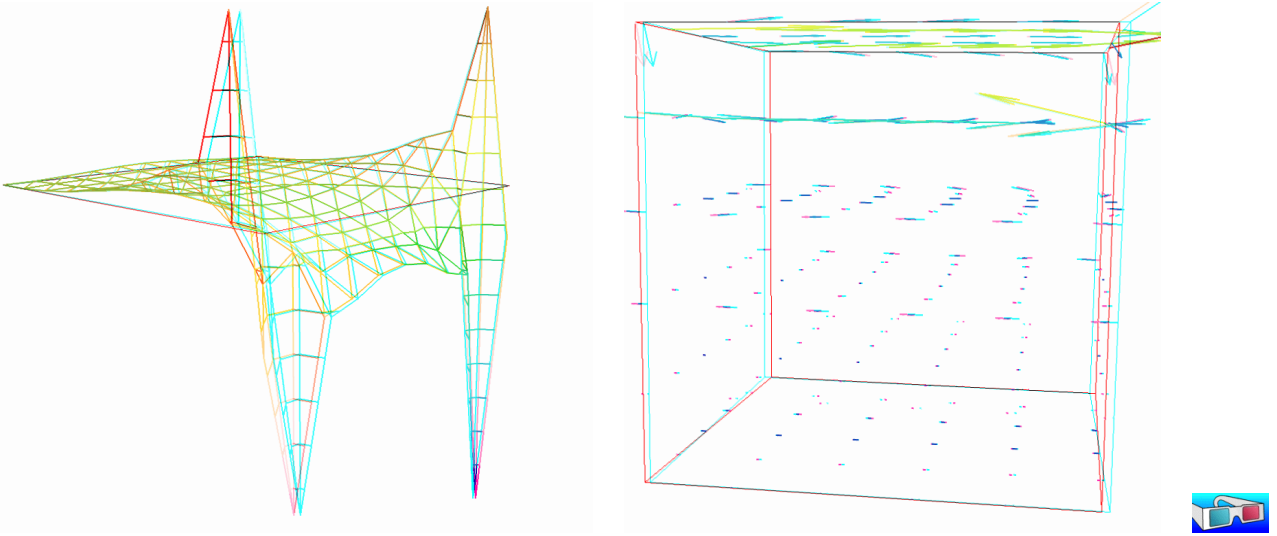


Figure 2.7: The vorticity: elevation view for  $d = 2$  and vector representation for  $d = 3$  (with anaglyph).

and you observe a discontinuous piecewise linear representation of the approximate vorticity. Also, the vorticity presents two sharp peaks at the upper corners of the driven cavity: the vorticity is unbounded and the peaks will increase with mesh refinements. This singularity of the solution is due to the boundary condition for the first component of the velocity  $u_0$  that jumps from zero to one at the corners. The approximate vorticity field can also be projected on a continuous piecewise linear space, using the `-proj` option (See Fig. 2.7 left):

```
./vorticity < square.field | field -elevation -stereo -nofill -
./vorticity < square.field | field -elevation -stereo -proj -
```

For  $d = 3$ , the whole vorticity vector can also be visualized (See Fig. 2.7 right):

```
./vorticity < cube.field | field -proj -velocity -stereo -
```

In the previous command, the `-proj` option has been used: since the 3D render has no support for discontinuous piecewise linear fields, the P1-discontinuous field is transformed into a P1-continuous one, thanks to a  $L^2$  projection. P1 The following command shows the second component of the vorticity vector, roughly similar to the bidimensional case.

```
./vorticity < cube.field | field -comp 1 -
./vorticity < cube.field | field -comp 1 -proj -
```

## 2.1.6 Computing the stream function

### Formulation and approximation

When  $d = 3$ , the stream function is a vector-valued field  $\psi$  that satisfies [41, page 90]:  $\mathbf{curl} \psi = \mathbf{u}$  and  $\text{div} \psi = 0$ . From the identity:

$$\mathbf{curl} \mathbf{curl} \psi = -\Delta \psi + \nabla(\text{div} \psi)$$

we obtain the following characterization of  $\psi$  :

$$\begin{aligned} -\Delta \psi &= \mathbf{curl} \mathbf{u} && \text{in } \Omega, \\ \psi &= 0 && \text{on } \Gamma_{\text{back}} \cup \Gamma_{\text{front}} \cup \Gamma_{\text{top}} \cup \Gamma_{\text{bottom}}, \\ \frac{\partial \psi}{\partial n} &= 0 && \text{on } \Gamma_{\text{left}} \cup \Gamma_{\text{right}}. \end{aligned}$$

When  $d = 2$ , the stream function  $\psi$  is a scalar-valued field the solution of the following problem [41, page 88]:

$$\begin{aligned} -\Delta \psi &= \operatorname{curl} \mathbf{u} && \text{in } \Omega, \\ \psi &= 0 && \text{on } \partial\Omega. \end{aligned}$$

Example file 2.10: streamf\_cavity.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "cavity.icc"
5 int main (int argc, char** argv) {
6     environment rheolef (argc, argv);
7     field uh;
8     din >> uh;
9     space Ph = cavity::streamf_space (uh.get_geo(), uh.get_approx());
10    space Xh = uh.get_space();
11    size_t d = uh.get_geo().dimension();
12    trial u (Xh), psi (Ph); test phi (Ph);
13    form a = (d == 3) ? integrate (ddot(grad(psi), grad(phi)))
14                  : integrate (dot(grad(psi), grad(phi)));
15    form b = (d==3) ? integrate (dot(curl(u), phi))
16                  : integrate (curl(u)*phi);
17    field psi_h = cavity::streamf_field (Ph);
18    field lh = b*uh;
19    solver sa (a.uu());
20    psi_h.set_u() = sa.solve (lh.u() - a.ub()*psi_h.b());
21    dout << catchmark("psi") << psi_h;
22 }

```

### How to run the program

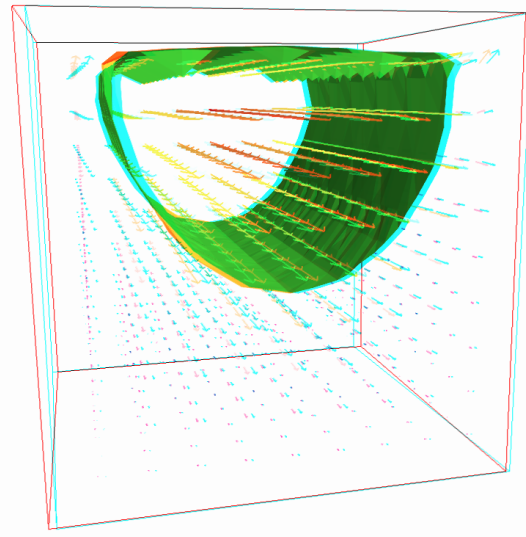
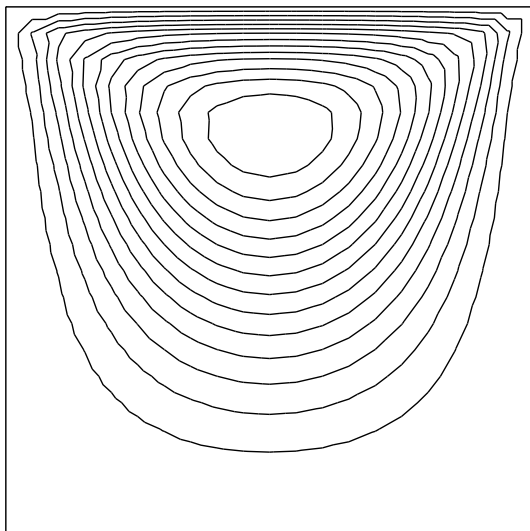


Figure 2.8: The stream function visualization: isolines for  $d = 2$ , and combined vectors and isonorm surface for  $d = 3$ .

For  $d = 2$ , just enter (see Fig. 2.8 left):

```

make streamf_cavity
./streamf_cavity < square.field | field -bw -

```

For  $d = 3$ , the whole stream function vector can be visualized:

```
./streamf_cavity < cube.field | field -velocity -scale 10 -
```

The second component of the stream function is showed by:

```
./streamf_cavity < cube.field | field -comp 1 -
```

The combined representation of Fig. 2.8.right has been obtained in two steps. First, enter:

```
./streamf_cavity < cube.field | field -comp 1 -noclean -noexecute -
mv output.vtk psi1.vtk
./streamf_cavity < cube.field | field -velocity -
```

The `-noclean -noexecute` options cause the creation of the `.vtk` file for the second component, without running the `paraview` visualization. Next, in the `paraview` window associated to the whole stream function, select the `File->Open` menu and load `'psi1.vtk'` and click on the green button `Apply`. Finally, select the `Filters/Common/Contours` menu: the isosurface appears. Observe that the 3D stream function is mainly represented by its second component.

## 2.2 Nearly incompressible elasticity and the stabilized Stokes problems

### 2.2.1 The incompressible elasticity problem

#### Formulation

Let us go back to the linear elasticity problem.

When  $\lambda$  becomes large, this problem is related to the *incompressible elasticity* and cannot be solved as it was previously done. To overcome this difficulty, the pressure is introduced :

$$p = -\lambda \operatorname{div} \mathbf{u}$$

and the problem becomes:

(E) find  $\mathbf{u}$  and  $p$  defined in  $\Omega$  such that:

$$\begin{aligned} -\operatorname{div}(2D(\mathbf{u})) + \nabla p &= \mathbf{f} \text{ in } \Omega, \\ -\operatorname{div} \mathbf{u} - \frac{1}{\lambda} p &= 0 \text{ in } \Omega, \\ +B.C. \end{aligned}$$

The variational formulation of this problem expresses:

(VFE) find  $\mathbf{u} \in V(1)$  and  $p \in L^2(\Omega)$  such that:

$$\begin{aligned} a(\mathbf{u}, \mathbf{v}) + b(\mathbf{v}, p) &= m(\mathbf{f}, \mathbf{v}), \forall \mathbf{v} \in V(0), \\ b(\mathbf{u}, q) - c(p, q) &= 0, \forall q \in L_0^2(\Omega), \end{aligned}$$

where

$$\begin{aligned} m(\mathbf{u}, \mathbf{v}) &= \int_{\Omega} \mathbf{u} \cdot \mathbf{v} \, dx, \\ a(\mathbf{u}, \mathbf{v}) &= \int_{\Omega} D(\mathbf{u}) : D(\mathbf{v}) \, dx, \\ b(\mathbf{v}, q) &= - \int_{\Omega} \operatorname{div}(\mathbf{v}) q \, dx. \\ c(p, q) &= \frac{1}{\lambda} \int_{\Omega} p q \, dx. \\ V &= \{ \mathbf{v} \in (H^1(\Omega))^2; \mathbf{v} = 0 \text{ on } \Gamma_{left} \cup \Gamma_{bottom} \} \end{aligned}$$

When  $\lambda$  becomes large, we obtain the incompressible elasticity problem, that coincides with the Stokes problem.

### Approximation

As for the Stokes problem, the Talar-Hood [49] finite element approximation is considered. We introduce a mesh  $\mathcal{T}_h$  of  $\Omega$  and the following finite dimensional spaces:

$$\begin{aligned} X_h &= \{ \mathbf{v} \in (H^1(\Omega)); \mathbf{v}_{/K} \in (P_2)^2, \forall K \in \mathcal{T}_h \}, \\ V_h(\alpha) &= X_h \cap V, \\ Q_h &= \{ q \in L^2(\Omega) \cap C^0(\bar{\Omega}); q_{/K} \in P_1, \forall K \in \mathcal{T}_h \}, \end{aligned}$$

The approximate problem writes:

$(VFE)_h$  find  $\mathbf{u}_h \in V_h(1)$  and  $p \in Q_h$  such that:

$$\begin{aligned} a(\mathbf{u}_h, \mathbf{v}) + b(\mathbf{v}, p_h) &= 0, \forall \mathbf{v} \in V_h(0), \\ b(\mathbf{u}_h, q) - c(p, q) &= 0, \forall q \in Q_h. \end{aligned}$$

Example file 2.11: incompressible-elasticity.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "embankment.icc"
5 int main(int argc, char**argv) {
6     environment rheolef (argc, argv);
7     geo omega (argv[1]);
8     Float inv_lambda = (argc > 2 ? atof(argv[2]) : 0);
9     size_t d = omega.dimension();
10    space Xh = embankment_space(omega, "P2");
11    space Qh (omega, "P1");
12    point f (0,0,0);
13    f [d-1] = -1;
14    trial u (Xh), p (Qh); test v (Xh), q (Qh);
15    field lh = integrate (dot(f,v));
16    form a = integrate (2*ddot(D(u),D(v)));
17    form b = integrate (-div(u)*q);
18    form mp = integrate (p*q);
19    form c = inv_lambda*mp;
20    field uh (Xh, 0), ph (Qh, 0);
21    solver_abtb elasticity (a.uu(), b.uu(), c.uu(), mp.uu());
22    elasticity.solve (lh.u() - a.ub()*uh.b(), -(b.ub()*uh.b()),
23                    uh.set_u(), ph.set_u());
24    dout << catchmark("inv_lambda") << inv_lambda << endl
25          << catchmark("u") << uh
26          << catchmark("p") << ph;
27 }

```

### Comments

The problem admits the following matrix form:

$$\begin{pmatrix} a.uu & \text{trans}(b.uu) \\ b.uu & -c.uu \end{pmatrix} \begin{pmatrix} uh.u \\ ph.u \end{pmatrix} = \begin{pmatrix} lh.u - a.ub * uh.b \\ -b.ub * uh.b \end{pmatrix}$$

The problem is similar to the Stokes one (see page 53). This system is solved by:

```

solver_abtb elasticity (a.uu(), b.uu(), c.uu(), mp.uu());
elasticity.solve (lh.u() - a.ub()*uh.b(), -(b.ub()*uh.b()),
                  uh.set_u(), ph.set_u());

```

For two-dimensional problems, a direct solver is used by default. In the three-dimensional case, an iterative algorithm is the default: the preconditioned conjugate gradient. The preconditioner is here the mass matrix `mp.uu` for the pressure. As showed in [53], the number of iterations need by the conjugate gradient algorithm to reach a given precision is then independent of the mesh size and is uniformly bounded when  $\lambda$  becomes small, i.e. in the incompressible case.

### How to run the program

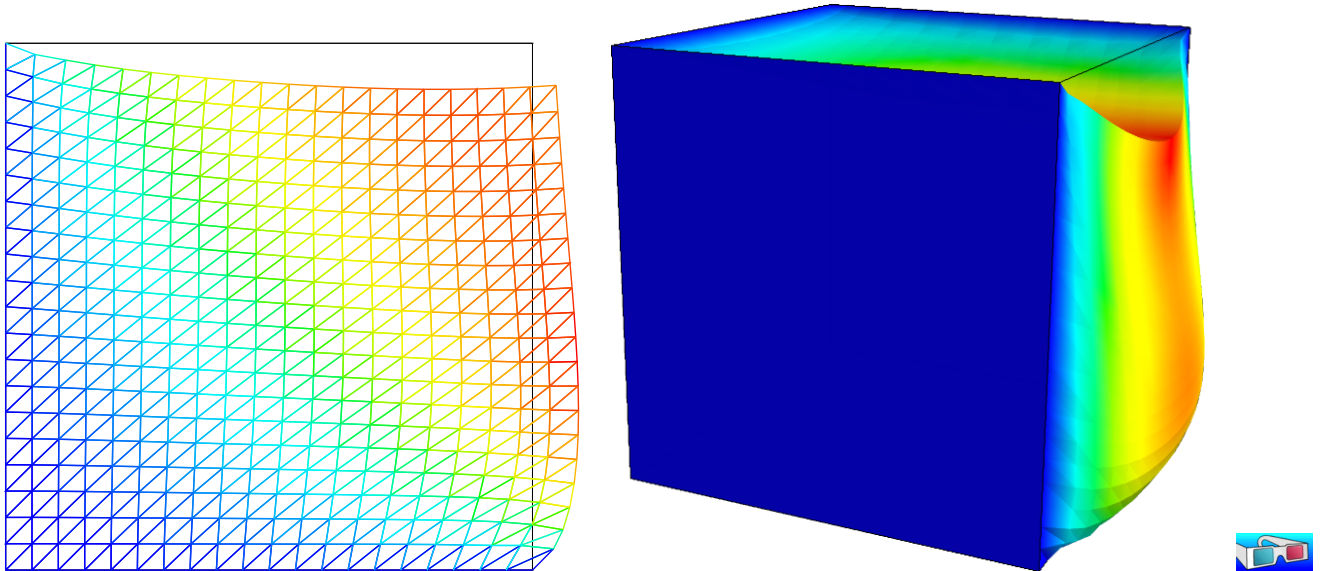


Figure 2.9: The incompressible linear elasticity ( $\lambda = +\infty$ ) for  $N = 2$  and  $N = 3$ .

We assume that the previous code is contained in the file ‘`incompressible-elasticity.cc`’. Compile the program as usual (see page 14):

```
make incompressible-elasticity
```

and enter the commands:

```
mkgeo_grid -t 10 > square.geo
./incompressible-elasticity square.geo 0 > square.field
field square.field -nofill
```

```
mkgeo_grid -T 10 > cube.geo
./incompressible-elasticity cube.geo 0 > cube.field
field cube.field -fill -scale 2
```

The visualization is performed as usual: see section 2.1.1, page 43. Compare the results on FIG. 2.9, obtained for  $\lambda = +\infty$  with those of FIG. 2.2, page 44, obtained for  $\lambda = 1$ .

Finally, the stress computation and the mesh adaptation loop is left as an exercise to the reader.

## 2.2.2 The $P_1b - P_1$ element for the Stokes problem

### Formulation and approximation

Let us go back to the Stokes problem. In section 2.1.4, page 51, the Taylor-Hood finite element was considered. Here, we turn to the mini-element [6] proposed by Arnold, Brezzi and Fortin, also

well-known as the *P1-bubble* element. This element is generally less precise than the Taylor-Hood one, but becomes popular, mostly because it is easy to implement in two and three dimensions and furnishes a  $P_1$  approximation of the velocity field. Moreover, this problem develops some links with stabilization technique and will presents some new **Rheolef** features.

We consider a mesh  $\mathcal{T}_h$  of  $\Omega \subset \mathbb{R}^d$ ,  $d = 2, 3$  composed only of simplicial elements: triangles when  $d = 2$  and tetrahedra when  $d = 3$ . The following finite dimensional spaces are introduced:

$$\begin{aligned}\mathbf{X}_h^{(1)} &= \{\mathbf{v} \in (H^1(\Omega))^d; \mathbf{v}_{/K} \in (P_1)^d, \forall K \in \mathcal{T}_h\}, \\ \mathbf{B}_h &= \{\boldsymbol{\beta} \in (C^0(\bar{\Omega}))^d; \boldsymbol{\beta}_{/K} \in B(K)^d, \forall K \in \mathcal{T}_h\} \\ \mathbf{X}_h &= \mathbf{X}_h^{(1)} \oplus \mathbf{B}_h \\ \mathbf{V}_h(\alpha) &= X_h \cap \mathbf{V}(\alpha), \\ Q_h &= \{q \in L^2(\Omega) \cap C^0(\bar{\Omega}); q_{/K} \in P_1, \forall K \in \mathcal{T}_h\},\end{aligned}$$

where  $B(K) = \text{vect}(\lambda_1 \times \dots \times \lambda_{d+1})$  and  $\lambda_i$  are the barycentric coordinates of the simplex  $K$ . The  $B(K)$  space is related to the *bubble* local space. The approximate problem is similar to (2.4), page 52, up to the choice of finite dimensional spaces.

Remark that the velocity field splits in two terms:  $\mathbf{u}_h = \mathbf{u}_h^{(1)} + \mathbf{u}_h^{(b)}$ , where  $\mathbf{u}_h^{(1)} \in \mathbf{X}_h^{(1)}$  is continuous and piecewise linear, and  $\mathbf{u}_h^{(b)} \in \mathbf{B}_h$  is the bubble term.

We consider the abrupt contraction geometry:

$$\Omega = ]-L_u, 0[ \times ]0, c[ \cup [0, L_d[ \times ]0, 1[$$

where  $c \geq 1$  stands for the contraction ratio, and  $L_u, L_d > 0$ , are the upstream and downstream tube lengths. The boundary conditions on  $\mathbf{u} = (u_0, u_1)$  for this test case are:

$$\begin{aligned}u_0 &= u_{\text{poiseuille}} \quad \text{and} \quad u_1 = 0 \quad \text{on} \quad \Gamma_{\text{upstream}} \\ \mathbf{u} &= 0 \quad \text{on} \quad \Gamma_{\text{wall}} \\ \frac{\partial u_0}{\partial x_1} &= 0 \quad \text{and} \quad u_1 = 0 \quad \text{on} \quad \Gamma_{\text{axis}} \\ \frac{\partial \mathbf{u}}{\partial n} &= 0 \quad \text{on} \quad \Gamma_{\text{downstream}}\end{aligned}$$

where

$$\begin{aligned}\Gamma_{\text{upstream}} &= \{-L_u\} \times ]0, c[ \\ \Gamma_{\text{downstream}} &= \{L_d\} \times ]0, 1[ \\ \Gamma_{\text{axis}} &= ]-L_u, L_d[ \times \{0\} \\ \Gamma_{\text{wall}} &= ]-L_u, 0[ \times \{c\} \cup \{0\} \times ]1, c[ \cup [0, L_d[ \times \{1\}\end{aligned}$$

The matrix structure is similar to those of the Taylor-Hood element (see section 2.1.4, page 51). Since  $\mathbf{X}_h = \mathbf{X}_h^{(1)} \oplus \mathbf{B}_h$ , any element  $u_h \in X_h$  can be written as a sum  $u_h = u_{1,h} + u_{b,h}$  where  $u_{1,h} \in \mathbf{X}_h^{(1)}$  and  $u_{b,h} \in \mathbf{B}_h$ . Remark that

$$a(u_{1,h}, v_{b,h}) = 0, \quad \forall u_{1,h} \in \mathbf{X}_h^{(1)}, \quad \forall v_{b,h} \in \mathbf{B}_h.$$

Thus, the form  $a(\cdot, \cdot)$  defined over  $\mathbf{X}_h \times \mathbf{X}_h$  writes simply as the sum of the forms  $a_1(\cdot, \cdot)$  and  $a_b(\cdot, \cdot)$ , defined over  $\mathbf{X}_h^{(1)} \times \mathbf{X}_h^{(1)}$  and  $\mathbf{B}_h \times \mathbf{B}_h$  respectively. Finally, the form  $b(\cdot, \cdot)$  defined over  $\mathbf{X}_h \times Q_h$  writes as the sum of the forms  $b_1(\cdot, \cdot)$  and  $b_b(\cdot, \cdot)$ , defined over  $\mathbf{X}_h^{(1)} \times Q_h$  and  $\mathbf{B}_h \times Q_h$  respectively. Then, the linear system admits the following block structure :

$$\begin{pmatrix} A_1 & 0 & B_1^T \\ 0 & A_b & B_b^T \\ B_1 & B_b & 0 \end{pmatrix} \begin{pmatrix} U_1 \\ U_b \\ P \end{pmatrix} = \begin{pmatrix} L_1 \\ L_b \\ L_p \end{pmatrix}$$

An alternative and popular implementation of this element eliminates the unknowns related to the bubble components (see e.g. [1], page 24). Remark that, on any element  $K \in \mathcal{T}_h$ , any bubble function  $v_K$  that belongs to  $B(K)$  vanishes on the boundary of  $K$  and have a compact support in  $K$ . Thus, the  $A_b$  matrix is block-diagonal. Moreover,  $A_b$  is invertible and  $U_b$  writes :

$$U_b = A_b^{-1}(B_b^T p - L_b)$$

As  $A_b$  is block-diagonal, its inverse can be efficiently inverted at the element level during the assembly process. Then,  $U_b$  can be easily eliminated from the system that reduces to:

$$\begin{pmatrix} A_1 & B_1^T \\ B_1 & -C \end{pmatrix} \begin{pmatrix} U_1 \\ P \end{pmatrix} = \begin{pmatrix} L_1 \\ \tilde{L}_p \end{pmatrix}$$

where  $\tilde{L}_p = L_p - A_b^{-1}L_p$  and  $C = B_b A_b^{-1} B_b^T$ . Remarks that the matrix structure is similar to those of the nearly incompressible elasticity (see 2.2.1, page 2.2.1). This reduced matrix formulation of the  $P_1b - P_1$  element is similar to the direct  $P_1 - P_1$  stabilized element, proposed by Brezzi and Pitkäranta [17].

Example file 2.12: stokes\_contraction\_bubble.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "contraction.icc"
5 int main(int argc, char**argv) {
6     environment rheolef (argc, argv);
7     geo omega (argv[1]);
8     space X1h = contraction::velocity_space (omega, "P1");
9     space Bh (omega, "bubble", "vector");
10    space Qh (omega, "P1");
11    trial u1 (X1h), ub (Bh), p (Qh);
12    test v1 (X1h), vb (Bh), q (Qh);
13    form mp = integrate (p*q);
14    form b1 = integrate (-div(u1)*q);
15    form bb = integrate (-div(ub)*q);
16    form a1 = integrate (2*ddot(D(u1),D(v1)));
17    integrate_option fopt;
18    fopt.invert = true;
19    form inv_ab = integrate (2*ddot(D(ub),D(vb)), fopt);
20    form c = bb*inv_ab*trans(bb);
21    field u1h = contraction::velocity_field (X1h);
22    field ph (Qh, 0);
23    solver_abtb stokes (a1.uu(), b1.uu(), c.uu(), mp.uu());
24    stokes.solve (-(a1.ub()*u1h.b()), -(b1.ub()*u1h.b()),
25                u1h.set_u(), ph.set_u());
26    dout << catchmark("inv_lambda") << 0 << endl
27          << catchmark("u") << u1h
28          << catchmark("p") << ph;
29 }

```

## Comments

First,  $A_b^{-1}$  is computed as:

```

integrate_option fopt;
fopt.invert = true;
form inv_ab = integrate (2*ddot(D(ub),D(vb)), fopt);

```

Notice the usage of the optional parameter `fopt` to the `integrate` function. As the form is bloc-diagonal, its inverse is computed element-by-element during the assembly process. Next, the  $C = B_b A_b^{-1} B_b^T$  form is simply computed as:

```

form c = bb*inv_ab*trans(bb);

```

The file 'contraction.icc' contains code for the velocity and stream function boundary conditions.

Example file 2.13: contraction.icc

```

1 struct contraction {
2   struct base {
3     base (geo omega) {
4       c = omega.xmax()[1];
5       string sys_coord = omega.coordinate_system_name();
6       cartesian = (sys_coord == "cartesian");
7       umax = cartesian ? 3/(2*c) : 4/sqr(c);
8     }
9     Float c, umax;
10    bool cartesian;
11  };
12  struct u_upstream: base {
13    u_upstream (geo omega) : base(omega) {}
14    Float operator() (const point& x) const {
15      return base::umax*(1-sqr(x[1]/base::c)); }
16  };
17  static space velocity_space (geo omega, string approx) {
18    space Xh (omega, approx, "vector");
19    Xh.block ("wall");
20    Xh.block ("upstream");
21    Xh[1].block ("axis");
22    Xh[1].block ("downstream");
23    return Xh;
24  }
25  static field velocity_field (space Xh) {
26    geo omega = Xh.get_geo();
27    space Wh (omega["upstream"], Xh.get_approx());
28    field uh (Xh, 0.);
29    uh[0]["upstream"] = interpolate (Wh, u_upstream(omega));
30    return uh;
31  }
32  static space streamf_space (geo omega, string approx) {
33    space Ph (omega, approx);
34    Ph.block ("upstream");
35    Ph.block ("wall");
36    Ph.block ("axis");
37    return Ph;
38  }
39  struct psi_upstream: base {
40    psi_upstream (geo omega) : base(omega) {}
41    Float operator() (const point& x) const {
42      Float y = (x[1]/base::c);
43      if (base::cartesian) {
44        return (base::umax*base::c)*(y*(1-sqr(y)/3) - 2./3);
45      } else {
46        return 0.5*(base::umax*sqr(base::c))*(sqr(y)*(1-sqr(y)/2) - 0.5);
47      }
48    }
49  };
50  static field streamf_field (space Ph) {
51    geo omega = Ph.get_geo();
52    space Wh (omega["upstream"], Ph.get_approx());
53    field psi_h (Ph, 0);
54    psi_upstream psi_up (omega);
55    psi_h["upstream"] = interpolate (Wh, psi_up);
56    psi_h["wall"] = 0;
57    psi_h["axis"] = -1;
58    return psi_h;
59  }
60 };

```

Without loss of generality, we assume that the half width of the downstream channel is assumed to be equal to one. The Poiseuille velocity upstream boundary condition `u_upstream` is then scaled such that the downstream average velocity is equal to one. By this way, the flow rate in the half upstream and downstream channel are also equal to one. The stream function is defined up to a



constant: we assume that it is equal to  $-1$  on the axis of symmetry: by this way, it is equal to zero on the wall.

The file ‘contraction.icc’ also contains a treatment of the axisymmetric variant of the geometry: this case will be presented in the next paragraph. Notice also the automatic computation of the geometric coordinate system and contraction ratio  $c$  from the input mesh, as:

```
c = omega.xmax()[1];
string sys_coord = omega.coordinate_system_name();
```

These parameters are transmitted via a base class to the class-function that computes the Poiseuille upstream flow boundary condition.

### How to run the program

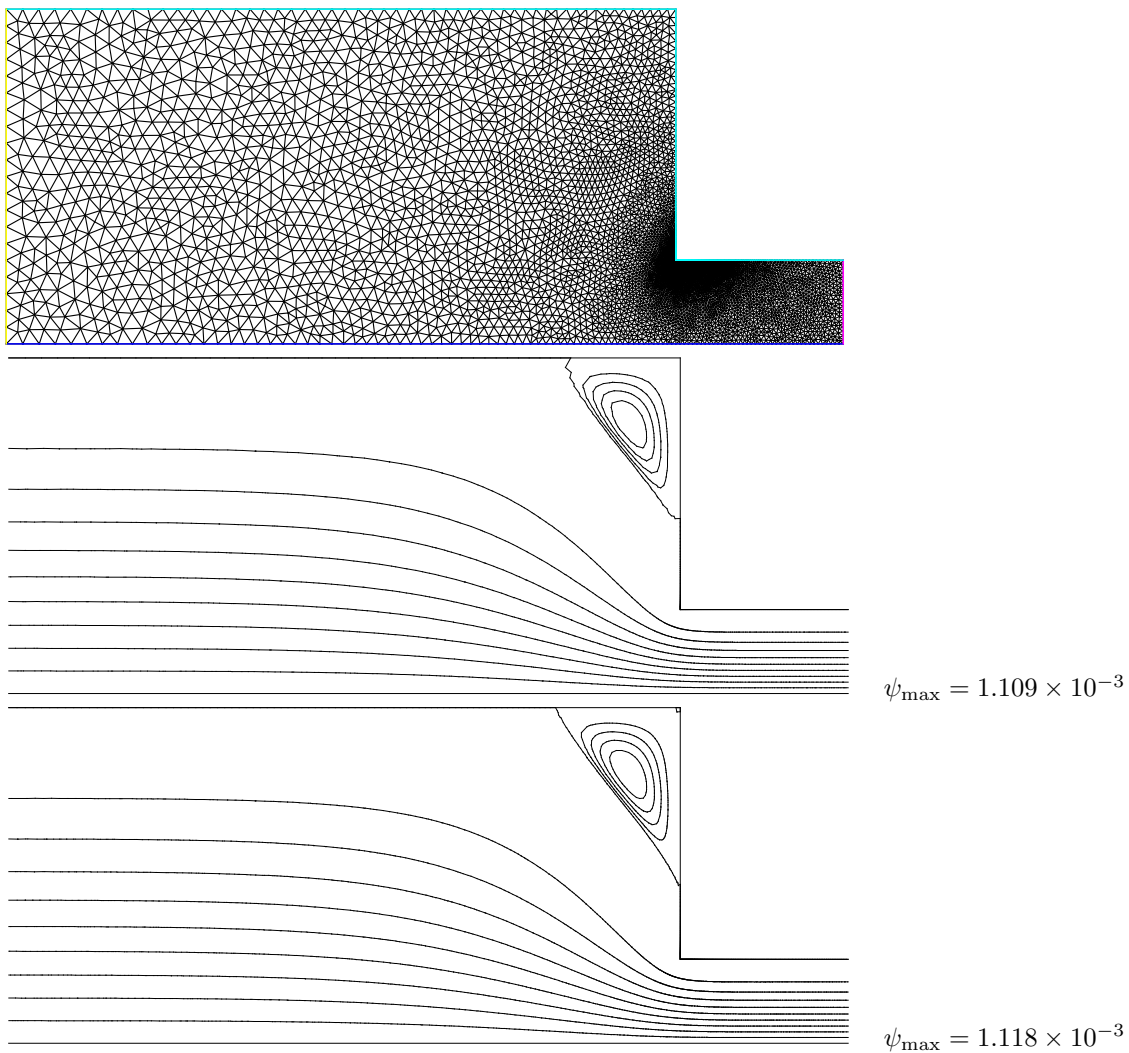


Figure 2.10: Solution of the Stokes problem in the abrupt contraction: (top) the mesh; (center) the  $P_1b - P_1$  stream function associated to the  $P_1b - P_1$  element; (bottom) the  $P_2 - P_1$  Taylor-Hood element.

The boundary conditions in this example are related to an abrupt contraction geometry with a free surface. The corresponding mesh ‘contraction.geo’ can be easily builded from the geometry

description file ‘[contraction.mshcad](#)’, which is provided in the example directory of the **Rheolef** distribution. The building mesh procedure is presented with details in [appendix A.2](#), page [A.2](#).

```
gmsht -2 contraction.mshcad -o contraction.msh
msh2geo contraction.msh > contraction.geo
geo contraction.geo
```

The mesh is represented on [Fig. 2.10.top](#). Then, the computation and the visualization writes:

```
make stokes_contraction_bubble
./stokes_contraction_bubble contraction.geo > contraction-P1.field
field contraction-P1.field -velocity
```

The visualization of the velocity field brings few informations about the properties of the flow. The stream function is more relevant for stationary flow visualization.

Example file 2.14: streamf\_contraction.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "contraction.icc"
5 int main (int argc, char** argv) {
6     environment rheolef (argc, argv);
7     field uh;
8     din >> uh;
9     const geo& omega = uh.get_geo();
10    size_t d = omega.dimension();
11    string sys_coord = omega.coordinate_system_name();
12    Float c = omega.xmax()[1];
13    string approx = "P" + itos(uh.get_space().degree());
14    space Ph = contraction::streamf_space (omega, approx);
15    field psi_h = contraction::streamf_field (Ph);
16    integrate_option fopt;
17    fopt.ignore_sys_coord = true;
18    const space& Xh = uh.get_space();
19    trial psi (Ph), u (Xh);
20    test xi (Ph), v (Xh);
21    form a = (d == 3) ? integrate (ddot(grad(psi), grad(xi)))
22                        : integrate (dot(grad(psi), grad(xi)), fopt);
23    field lh = integrate (dot(uh, bcurl(xi)));
24    solver sa (a.uu());
25    psi_h.set_u() = sa.solve (lh.u() - a.ub()*psi_h.b());
26    dout << catchmark("psi") << psi_h;
27 }

```

Notice the usage of the optional parameter `fopt` to the `integrate` function.

```
fopt.ignore_sys_coord = true;
```

In the axisymmetric coordinate system, there is a specific definition of the stream function, together with the use of a variant of the `curl` operator, denoted as `bcurl` in **Rheolef**.

```
field lh = integrate (dot(uh, bcurl(xi)));
```

The axisymmetric case will be presented in the next section. By this way, our code is able to deal with both cartesian and axisymmetric geometries.

The stream function  $\psi$  (see also section 2.1.6) is computed and visualized as:

```

make streamf_contraction
./streamf_contraction < contraction-P1.field > contraction-P1-psi.field
field contraction-P1-psi.field
field contraction-P1-psi.field -n-iso 15 -n-iso-negative 10 -bw

```

The  $P_1$  stream function is represented on Fig. 2.10.center. The stream function is zero along the wall and the line separating the main flow and the vortex located in the outer corner of the contraction. Thus, the isoline associated to the zero value separates the main flow from the vortex. In order to observe this vortex, an extra `-n-iso-negative 10` option is added: ten isolines are drawn for negatives values of  $\psi$ , associated to the main flow, and `n_iso-10` for the positives values, associated to the vortex.

A similar computation based on the Taylor-Hood  $P_2 - P_1$  element is implemented in `stokes_contraction.cc`. The code is similar, up to the boundary conditions, to `stokes_cavity.cc` (see page 52): thus it is not listed here but is available in the **Rheolef** example directory.

```

make stokes_contraction
./stokes_contraction contraction.geo > contraction-P2.field
field contraction-P2.field -velocity
./streamf_contraction < contraction-P2.field > contraction-P2-psi.field
field contraction-P2-psi.field -n-iso-negative 10 -bw

```

The associated  $P_2$  stream function is represented on Fig. 2.10.bottom. Observe that the two solutions are similar and that the vortex activity, defined as  $\psi_{\max}$ , is accurately computed with the two methods (see also [84], Fig. 5.11.a, page 143).

```
field contraction-P1-psi.field -max
field contraction-P2-psi.field -max
```

Recall that the stream function is negative in the main flow and positive in the vortex located in the outer corner of the contraction. Nevertheless, the Taylor-Hood based solution is more accurate : this is perceptible on the graphic, in the region where the upstream vortex reaches the boundary.

### 2.2.3 Axisymmetric geometries

Axisymmetric geometries are fully supported in **Rheolef**: the coordinate system is associated to the geometry description, stored together with the mesh in the ‘.geo’ and this information is propagated in spaces, forms and fields without any change in the code. Thus, a code that works in plane a 2D plane geometry is able to support a 3D axisymmetric one without changes. A simple axisymmetric geometry writes:

```
mkgeo_grid -t 10 -zr > square-zr.geo
more square-zr.geo
```

Remark the additional line in the header:

```
coordinate_system zr
```

The axis of symmetry is denoted as  $z$  while the polar coordinates are  $(r, \theta)$ . By symmetry, the problem is supposed to be independent upon  $\theta$  and the computational domain is described by  $(x_0, x_1) = (z, r)$ . Conversely, in some cases, it could be convenient to swap the order of the coordinates and use  $(r, z)$ : this feature is obtained by the `-rz` option:

```
mkgeo_grid -t 10 -rz > square-rz.geo
more square-rz.geo
```

Axisymmetric problems uses  $L^2$  functional space equipped with the following weighted scalar product

$$(f, g) = \int_{\Omega} f(z, r) g(z, r) r \, dr dz$$

and all usual bilinear forms support this weight. Thus, the **coordinate system can be chosen at run time** and we can expect an efficient source code reduction.

### 2.2.4 The axisymmetric stream function and stress tensor

In the axisymmetric case, the velocity field  $\mathbf{u} = (u_z, u_r)$  can be expressed in terms of the Stokes stream function  $\psi$  by (see Batchelor [11, p.453] and [111]):

$$\mathbf{u} = (u_z, u_r) = \left( \frac{1}{r} \frac{\partial \psi}{\partial r}, -\frac{1}{r} \frac{\partial \psi}{\partial z} \right) \quad (2.5)$$

Recall that in the axisymmetric case:

$$\mathbf{curl} \psi = \left( \frac{1}{r} \frac{\partial(r\psi)}{\partial r}, -\frac{\partial \psi}{\partial z} \right)$$

Thus, from this definition, in axisymmetric geometries  $\mathbf{u} \neq \mathbf{curl} \psi$  and the definition of  $\psi$  differs from the 2D plane or 3D cases (see section 2.1.6, page 56).

Let us turn to a variational formulation in order to compute  $\psi$  from  $\mathbf{u}$ . For any  $\xi \in H^1(\Omega)$ , let us multiply (2.5) by  $\mathbf{v} = (\partial_r \xi, -\partial_z \xi)$  and then integrate over  $\Omega$  with the  $r dr dz$  weight. For any known  $\mathbf{u}$  velocity field, the problem writes:

(P): find  $\psi \in \Psi(\psi_\Gamma)$  such that

$$a(\psi, \xi) = l(\xi), \quad \forall \xi \in \Psi(0)$$

where we have introduced the following bilinear forms:

$$\begin{aligned} a(\psi, \xi) &= \int_{\Omega} \left( \frac{\partial \psi}{\partial r} \frac{\partial \xi}{\partial r} + \frac{\partial \psi}{\partial z} \frac{\partial \xi}{\partial z} \right) dr dz \\ l(\xi) &= \int_{\Omega} \left( \frac{\partial \xi}{\partial r} u_z - \frac{\partial \xi}{\partial z} u_r \right) r dr dz \end{aligned}$$

These forms are defined in ‘`streamf_contraction.cc`’ as:

```
integrate_option fopt;
fopt.ignore_sys_coord = true;
form a = integrate (dot(grad(psi), grad(xi)), fopt);
```

and

```
field lh = integrate (dot(uh, bcurl(xi)));
```

The `fopt.ignore_sys_coord` allows us to drop the  $r$  integration weight, i.e. replace  $r dr dz$  by  $dr dz$  when computing the  $a(\cdot, \cdot)$  form. Conversely,  $l$  involves the `bcurl` operator defined as:

$$\mathbf{bcurl} \xi = \left( \frac{\partial \xi}{\partial r}, -\frac{\partial \xi}{\partial z} \right)$$

It is closely related but differs from the standard `curl` operator:

$$\mathbf{curl} \xi = \left( \frac{1}{r} \frac{\partial(r\xi)}{\partial r}, -\frac{\partial \xi}{\partial z} \right)$$

The `bcurl` operator is a specific notation introduced in **Rheolef**: it coincides with the usual `curl` operator except for axisymmetric geometries. In that case, it refers to the Batchelor trick, suitable for the computation of the stream function.

As an example, let us reconsider the contraction geometry (see section 2.2.2, page 60), extended in the axisymmetric case. In that case, the functional space is defined by:

$$\Psi(\psi_\Gamma) = \{ \varphi \in H^1(\Omega); \varphi = \psi_\Gamma \text{ on } \Gamma_{\text{upstream}} \cup \Gamma_{\text{wall}} \cup \Gamma_{\text{axis}} \}$$

with

$$\psi_\Gamma = \begin{cases} \psi_{\text{poiseuille}} & \text{on } \Gamma_{\text{upstream}} \\ 0 & \text{on } \Gamma_{\text{wall}} \\ -1 & \text{on } \Gamma_{\text{axis}} \end{cases}$$

This space corresponds to the imposition of Dirichlet boundary conditions on  $\Gamma_{\text{upstream}}$ ,  $\Gamma_{\text{wall}}$  and  $\Gamma_{\text{axis}}$  and a Neumann boundary condition on  $\Gamma_{\text{downstream}}$ .

The following unix commands generate the axisymmetric geometry:

```
gmsht -2 contraction.mshcad -o contraction.msh
msh2geo -zr contraction.msh > contraction-zr.geo
more contraction-zr.geo
geo contraction-zr.geo
```

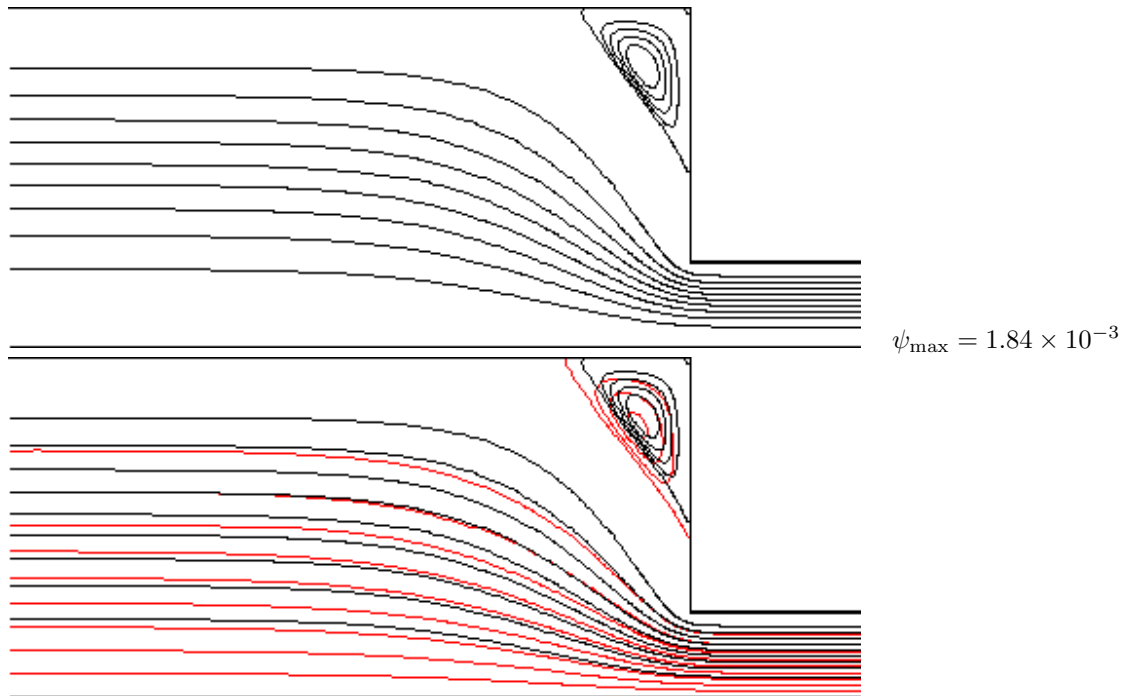


Figure 2.11: Solution of the axisymmetric Stokes problem in the abrupt contraction: (top) the  $P_2$  stream function associated to the  $P_2 - P_1$  element; (bottom) comparison with the 2D Cartesian solution (in red).

The previous code `stokes_contraction.cc` and `streamf_contraction.cc` are both reused as:

```
./stokes_contraction contraction-zr.geo > contraction-zr-P2.field
./streamf_contraction < contraction-zr-P2.field > contraction-zr-P2-psi.field
field contraction-zr-P2-psi.field -n-iso-negative 10 -bw
```

The solution is represented on Fig. 2.11: it slightly differs from the 2D Cartesian solution, as computed in the previous section (see Fig. 2.10). The vortex size is smaller but its intensity  $\psi_{\max} = 1.84 \times 10^{-3}$  is higher. Despite the stream functions looks like similar, the plane solutions are really different, as we can observe from a cut of the first component of the velocity along the axis (see Fig. 2.12):

```
field contraction-P2.field -comp 0 -cut -normal 0 1 -origin 0 1e-15 -gnuplot
field contraction-zr-P2.field -comp 0 -cut -normal 0 1 -origin 0 1e-15 -gnuplot
```

The `1e-15` argument replace the zero value, as the mesh intersection cannot yet be done exactly on the boundary. Notice that the `stokes_contraction_bubble.cc` can be also reused in a similar way:

```
./stokes_contraction_bubble contraction-zr.geo > contraction-zr-P1.field
./streamf_contraction < contraction-zr-P1.field > contraction-zr-P1-psi.field
field contraction-zr-P1-psi.field -n-iso-negative 10 -bw
```

There is another major difference with axisymmetric problems: the rate of deformation tensor writes:

$$\tau = 2D(\mathbf{u}) = \begin{pmatrix} \tau_{zz} & \tau_{rz} & 0 \\ \tau_{rz} & \tau_{rr} & 0 \\ 0 & 0 & \tau_{\theta\theta} \end{pmatrix}$$

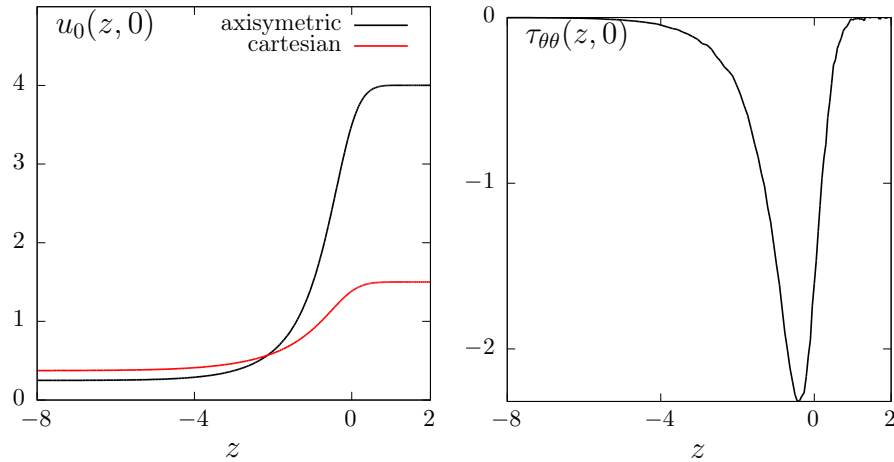


Figure 2.12: Solution of the plane and axisymmetric Stokes problem in the abrupt contraction: cut along the axis of symmetry: (left):  $u_0$ ; (right)  $\tau_{\theta\theta}$ .

Thus, there is an additional non-zero component  $\tau_{\theta\theta}$  that is automatically integrated into the computations in **Rheolef**. The incompressibility relation leads to  $\text{tr}(\tau) = \tau_{zz} + \tau_{rr} + \tau_{\theta\theta} = 0$ . Here  $\sigma_{\text{tot}} = -p.I + \tau$  is the total Cauchy stress tensor (by a dimensionless procedure, the viscosity can be taken as one). By reusing the `stress.cc` code (see page 46) we are able to compute the tensor components:

```
make stress
./stress < contraction-zr-P1.field > contraction-zr-P1-tau.field
```

The visualization along the axis of symmetry for the  $\tau_{\theta\theta}$  component is obtained by (see Fig. 2.12):

```
field contraction-zr-P1-tau.field -comp 22 -proj -cut -normal 0 1 -origin 0 1e-15 -gnuplot
```

Recall that the  $\tau_{zz}$  and  $\tau_{rr}$  components are obtained by the `-comp 00` and `-comp 11` options, respectively. The non-zero values along the axis of symmetry expresses the elongational effects in the entry region of the abrupt contraction.

## 2.3 Time-dependent problems

### 2.3.1 The heat equation

#### Formulation

Let  $T > 0$ ,  $\Omega \subset \mathbb{R}^d$ ,  $d = 1, 2, 3$  and  $f$  defined in  $\Omega$ . The heat problem writes:

(P): find  $u$ , defined in  $\Omega \times ]0, T[$ , such that

$$\begin{aligned} \frac{\partial u}{\partial t} - \Delta u &= f \text{ in } \Omega \times ]0, T[, \\ u(0) &= 0 \text{ in } \Omega, \\ u(t) &= 0 \text{ on } \partial\Omega \times ]0, T[. \end{aligned}$$

where  $f$  is a known function. In the present example, we consider  $f = 1$ .

### Approximation

Let  $\Delta t > 0$  and  $t_n = n\Delta t$ ,  $n \geq 0$ . The problem is approximated with respect to time by the following first-order implicit Euler scheme:

$$\frac{u^{n+1} - u^n}{\Delta t} - \Delta u^{n+1} = f(t_{n+1}) \quad \text{in } \Omega$$

where  $u^n \approx u(n\Delta t)$  and  $u^{(0)} = 0$ . The variational formulation of the time-discretized problem writes:

$(VF)_n$ : Let  $u^n$  being known, find  $u^{n+1} \in H_0^1(\Omega)$  such that

$$a(u^{n+1}, v) = l^{(n)}(v), \quad \forall v \in H_0^1(\Omega).$$

where

$$a(u, v) = \int_{\Omega} (uv + \Delta t \nabla u \cdot \nabla v) v \, dx$$

$$l^{(n)}(v) = \int_{\Omega} (u^n + \Delta t f(t_{n+1})) v \, dx$$

This is a Poisson-like problem. The discretization with respect to space of this problem is similar to those presented in section 1.1.1, page 12.

Example file 2.15: heat.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 int main (int argc, char **argv) {
5     environment rheolef (argc, argv);
6     geo omega (argv[1]);
7     size_t n_max = (argc > 2) ? atoi(argv[2]) : 100;
8     Float delta_t = 0.5/n_max;
9     space Xh (omega, "P1");
10    Xh.block ("boundary");
11    trial u (Xh); test v (Xh);
12    form a = integrate (u*v + delta_t*dot(grad(u),grad(v)));
13    solver sa = ldlt (a.uu());
14    field uh (Xh, 0);
15    branch event ("t", "u");
16    dout << event (0, uh);
17    for (size_t n = 1; n <= n_max; n++) {
18        field rhs = uh + delta_t;
19        field lh = integrate (rhs*v);
20        uh.set_u() = sa.solve (lh.u() - a.ub()*uh.b());
21        dout << event (n*delta_t, uh);
22    }
23 }

```

### Comments

Notice the use of the `branch` class:

```
branch event ("t", "u");
```

this is a wrapper class that is used here to print the branch of solution  $(t_n, u^n)_{n \geq 0}$ , on the standard output in the ‘.branch’ file format. An instruction as:

```
dout << event (t, uh);
```

is equivalent to the formatted output

```
dout << catchmark("t") << t << endl
      << catchmark("u") << uh;
```



## How to run the program

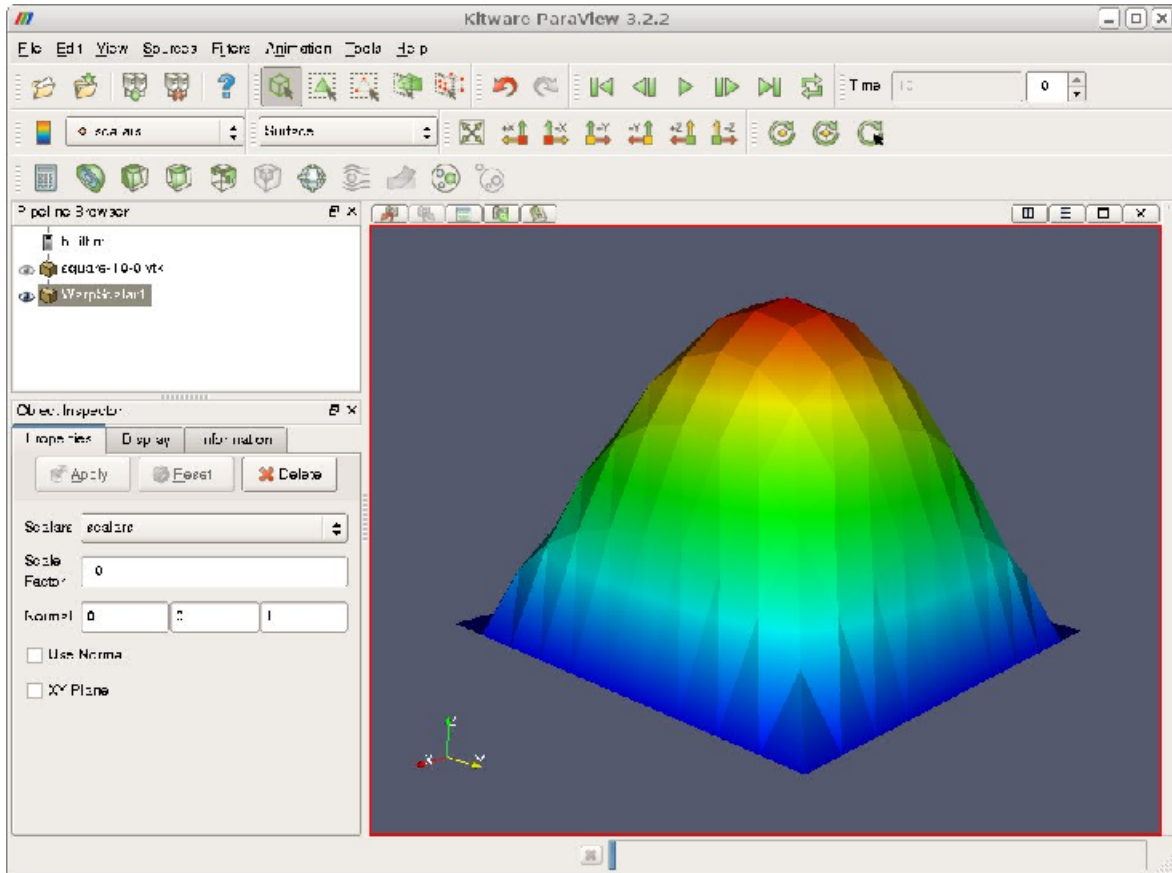


Figure 2.13: Animation of the solution of the heat problem.

We assume that the previous code is contained in the file ‘`heat.cc`’. Then, compile the program as usual (see page 14):

```
make heat
```

For a one dimensional problem, enter the commands:

```
mkgeo_grid -e 100 > line.geo
./heat line.geo > line.branch
```

The previous commands solve the problem for the corresponding mesh and write the solution in the field-family file format ‘`.branch`’. For a bidimensional one:

```
mkgeo_grid -t 10 > square.geo
./heat square.geo > square.branch
```

For a tridimensional one:

```
mkgeo_grid -T 10 > box.geo
./heat box.geo > box.branch
```

### How to run the animation

```
branch line.branch -gnuplot
```

A `gnuplot` window appears. Enter `q` to exit the window. For a bidimensional case, a more sophisticated procedure is required. Enter the following unix commands:

```
branch square.branch -paraview
paraview &
```

A window appears, that looks like a video player. Then, open the **File->open** menu and load `square-. .vtk`. The first `'.'` stands for a wildcard, i.e. the time index family. Then, press the `apply` green button and, click a first time on the video `play` button, at the top of the window. Note that the range of the color bar is not well defined: it has been set from the values at the first time step, which is zero. Thus, we have to rescale the data range for the color bar. There is a special small button **rescale to data range** over all time steps in the **properties** window: activate it and then click a second time on the video `play` button. An elevation view can be also obtained: first, click on **2D** button at the top of the **layout** window: it changes to **3D**. Then, select the **Filter->alphabetical->wrap by scalar** menu, choose **10** as **scale factor** and press the `apply` green button. If the view has been loss, click on the `reset` button on the top button bar of the `paraview` window. Then, click on the graphic window, rotate the view and finally re-play the animation

To generate an animation file, go to the **File->save animation** menu and enter as file name `square` and as file type `ogg vorbis/theora`. The animation file `square.ogv` can now be started from any video player, such as `vlc`:

```
vlc --loop square.ogv
```

For the tridimensional case, the animation feature is similar.

## 2.3.2 The convection-diffusion problem

### Formulation

Let  $T > 0$  and  $\nu > 0$ . The convection-diffusion problem writes:

(P): find  $\phi$ , defined in  $\Omega \times ]0, T[$ , such that

$$\begin{aligned} \frac{\partial \phi}{\partial t} + \mathbf{u} \cdot \nabla \phi - \nu \Delta \phi + \sigma \phi &= 0 \text{ in } \Omega \times ]0, T[ \\ \phi(0) &= \phi_0 \text{ in } \Omega \\ \phi(t) &= \phi_\Gamma(t) \text{ on } \partial\Omega \times ]0, T[ \end{aligned}$$

where  $\mathbf{u}$ ,  $\sigma \geq 0$ ,  $\phi_0$  and  $\phi_\Gamma$  being known. Notice the additional  $\mathbf{u} \cdot \nabla$  operator.

### Time approximation

This problem is approximated by the following first-order implicit Euler scheme:

$$\frac{\phi^{n+1} - \phi^n \circ X^n}{\Delta t} - \nu \Delta \phi^{n+1} + \sigma \phi^{n+1} = 0 \text{ in } \Omega$$

where  $\Delta t > 0$ ,  $\phi^n \approx \phi(n\Delta t)$  and  $\phi^{(0)} = \phi_0$ .

Let  $t_n = n\Delta t$ ,  $n \geq 0$ . The term  $X^n(x)$  is the position at  $t_n$  of the particle that is in  $x$  at  $t_{n+1}$  and is transported by  $\mathbf{u}^n$ . Thus,  $X^n(x) = X(t_n, x)$  where  $X(t, x)$  is the solution of the differential equation

$$\begin{cases} \frac{dX}{dt} = \mathbf{u}(X(t, x), t) & p.p. \ t \in ]t_n, t_{n+1}[ \\ X(t_{n+1}, x) = x. \end{cases}$$

Then  $X^n(x)$  is approximated by the first-order Euler approximation

$$X^n(x) \approx x - \Delta t \mathbf{n}^n(x).$$

This algorithm has been introduced by O. Pironneau (see e.g. [74]), and is known as the method of characteristic in the finite difference context and as the Lagrange-Galerkin in the finite element one. The efficient evaluation of  $\phi_h \circ X^n(x)$  in an unstructured mesh involves a hierarchical  $d$ -tree (quadtree, octree) data structure for the localization of the element  $K$  of the mesh that contains  $x$ . When  $d = 3$  requires also sophisticated geometric predicates to test whether  $x \in K$  without rounding errors, and avoid to conclude that no elements contains a point  $x$  close to  $\partial K$  up to rounding errors. This problems is addressed in **Rheolef** based on the **cgal** library.

The following code implements the classical rotating Gaussian hill test case (see e.g. [83]).

Example file 2.16: convect.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "rotating-hill.h"
5 int main (int argc, char **argv) {
6     environment rheolef (argc, argv);
7     geo omega (argv[1]);
8     string approx = (argc > 2) ? argv[2] : "P1";
9     Float nu      = (argc > 3) ? atof(argv[3]) : 1e-2;
10    size_t n_max  = (argc > 4) ? atoi(argv[4]) : 50;
11    size_t d = omega.dimension();
12    Float delta_t = 2*acos(-1.)/n_max;
13    space Vh (omega, approx, "vector");
14    field uh = interpolate (Vh, u(d));
15    space Xh (omega, approx);
16    Xh.block ("boundary");
17    field phi_h = interpolate (Xh, phi(d, nu, 0));
18    characteristic X (-delta_t*uh);
19    quadrature_option qopt;
20    qopt.set_family (quadrature_option::gauss_lobatto);
21    qopt.set_order (Xh.degree());
22    trial phi (Xh); test psi (Xh);
23    branch event ("t", "phi");
24    dout << catchmark("nu") << nu << endl
25         << event (0, phi_h);
26    for (size_t n = 1; n <= n_max; n++) {
27        Float t = n*delta_t;
28        Float c1 = 1 + delta_t*phi::sigma(d, nu, t);
29        Float c2 = delta_t*nu;
30        form a = integrate (c1*phi*psi + c2*dot(grad(phi), grad(psi)), qopt);
31        field lh = integrate (compose(phi_h, X)*psi, qopt);
32        solver sa (a.uu());
33        phi_h.set_u() = sa.solve (lh.u() - a.ub()*phi_h.b());
34        dout << event (t, phi_h);
35    }
36 }

```

### Comments

The characteristic variable X implements the localizer  $X^n(x)$ :

```
characteristic X (-delta_t*uh);
```

Combined with the `compose` function, it perform the composition  $\phi_h \circ X^n$ . The right-hand side is then computed by using the `integrate` function:

```
field lh = integrate (compose(phi_h, X)*psi, qopt);
```

Notice the additional `qopt` argument to the `integrate` function. By default, when this argument is omitted, a Gauss quadrature formulae is assumed, and the number of point is computed such that it integrate exactly  $2k + 1$  polynoms, where  $k$  is the degree of polynoms in  $X_h$ . The Gauss-Lobatto quadrature formulae is recommended for Lagrange-Galerkin methods. Recall that this choice of quadrature formulae guaranties unconditional stability at any polynomial order. Here, we specifies a Gauss-Lobatto quadrature formulae that should be exact for  $k$  order polynoms. The bilinear form is computed via the same quadrature formulae:

```
form a = integrate (c1*phi*psi + c2*dot(grad(phi), grad(psi)), qopt);
```

A test case is described in [75]: we take  $\Omega = ]-2, 2[^d$  and  $T = 2\pi$ . This problem provides an example for a convection-diffusion equation and a known analytical solution:

$$\phi(t, x) = \exp(-\lambda t - r(t)|x - x_0(t)|^2)$$

where  $\lambda = 4\nu/t_0 \geq 0$  with  $t_0 > 0$  and  $\nu \geq 0$ ,  $x_0(t)$  is the moving center of the hill and  $r(t) = 1/(t_0 + 4\nu t)$ . The source term is time-dependent:  $\sigma(t) = \lambda - 2d\nu r(t)$  and has been adjusted such that the right-hand side is zero. The moving center of the hill  $x_0(t)$  is associated to the velocity field  $\mathbf{u}(t, x)$  as:

$d$	$\mathbf{u}(t, x)$	$x_0(t)$
1	$1/(2\pi)$	$t/(2\pi) - 1/2$
2	$(y, -x)$	$(-\cos(t)/2, \sin(t)/2)$
3	$(y, -x, 0)$	$(-\cos(t)/2, \sin(t)/2, 0)$

Example file 2.17: `rotating-hill.h`

```

1 struct u {
2   point operator() (const point & x) const {
3     return (d == 1) ? point(u0) : point(x[1], -x[0]); }
4   u (size_t d1) : d(d1), u0 (0.5/acos(Float(-1))) {}
5   protected: size_t d; Float u0;
6 };
7 struct phi {
8   static Float sigma(size_t d, Float nu1, Float t) {
9     const Float t0 = 0.2;
10    return 4*nu1/t0 - 2*d*nu1/(t0 + 4*nu1*t); }
11   Float operator() (const point& x) const {
12     point x0t;
13     if (d == 1) { x0t = point(x0[0] + u0*t); }
14     else {
15         x0t = point( x0[0]*cos(t) + x0[1]*sin(t),
16                   -x0[0]*sin(t) + x0[1]*cos(t));
17     }
18     return exp(-4*nu*(t/t0) - dist2(x, x0t)/(t0+4*nu*t));
19   }
20   phi (size_t d1, Float nu1, Float t1) : d(d1), nu(nu1), t(t1),
21     t0(0.2), u0 (0.5/acos(Float(-1))), x0(-0.5,0) {}
22   protected: size_t d; Float nu, t, t0, u0; point x0;
23 };

```

Notice the use of a class-function `phi` for the implementation of  $\phi(t)$  as a function of  $x$ . Such programming style has been introduced in the *standard template library* [62], which is a part of the standard C++ library. By this way, for a given  $t$ ,  $\phi(t)$  can be interpolated as an usual function on a mesh.

### How to run the program

We assume that the previous code is contained in the file '`convect.cc`'. Then, compile the program as usual (see page 14):

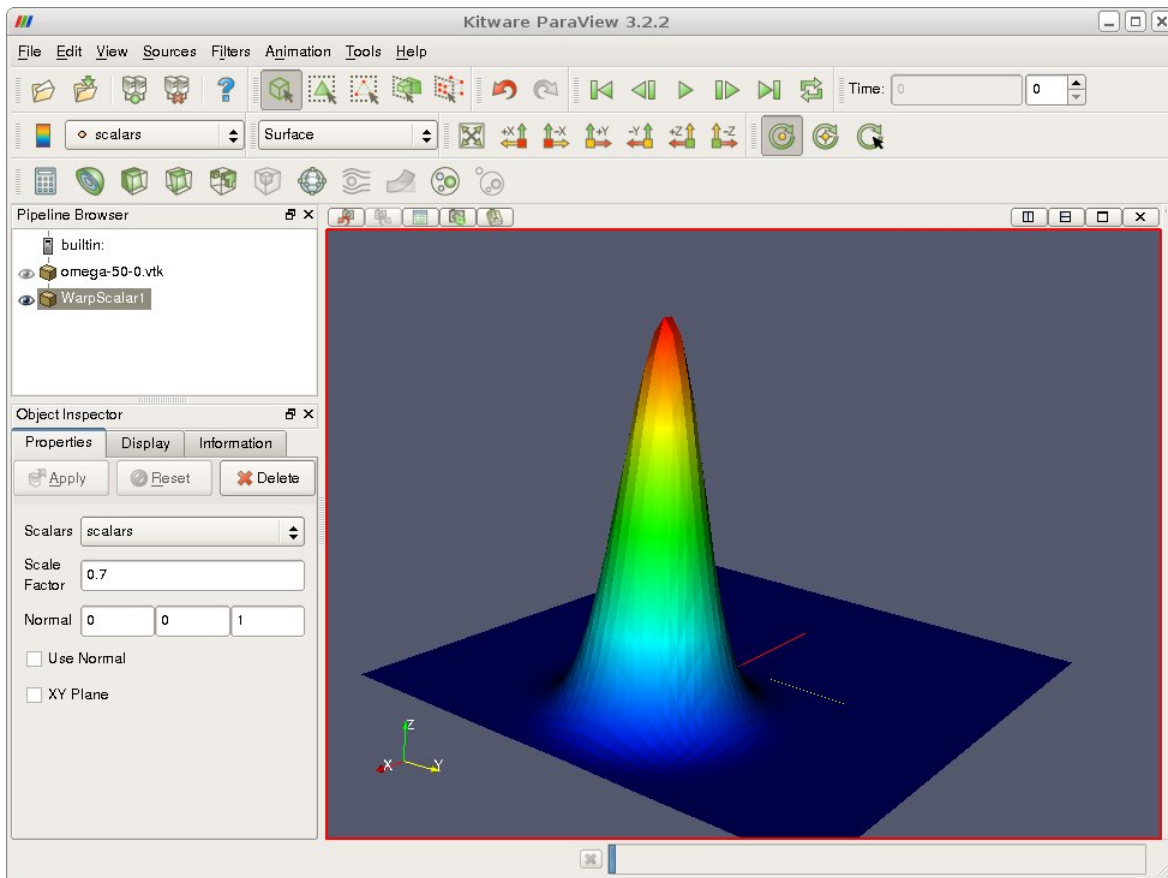


Figure 2.14: Animation of the solution of the rotating hill problem.

make convect

and enter the commands: Running the one-dimensional test case:

```
mkgeo_grid -e 500 -a -2 -b 2 > line2.geo
./convect line2.geo P1 > line2.branch
branch line2.branch -gnuplot
```

Notice the hill that moves from  $x = -1/2$  to  $x = 1/2$ . Since the exact solution is known, it is possible to analyze the error:

Example file 2.18: convect\_error.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "rotating-hill.h"
5 int main (int argc, char **argv) {
6     environment rheolef (argc,argv);
7     Float tol = (argc > 1) ? atof(argv[1]) : 1e-10;
8     Float nu;
9     din >> catchmark("nu") >> nu;
10    branch get ("t","phi");
11    branch put ("t","phi_h","pi_h_phi");
12    derr << "# t\tterror_l2\tterror_linf" << endl;
13    field phi_h;
14    Float err_l2_l2 = 0;
15    Float err_linf_linf = 0;
16    for (Float t = 0, t_prec = 0; din >> get (t, phi_h); t_prec = t) {
17        const space& Xh = phi_h.get_space();
18        size_t d = Xh.get_geo().dimension();
19        field pi_h_phi = interpolate (Xh, phi(d,nu,t));
20        trial phi (Xh); test psi (Xh);
21        form m = integrate (phi*psi);
22        field eh = phi_h - pi_h_phi;
23        Float err_l2 = sqrt(m(eh,eh));
24        Float err_linf = eh.max_abs();
25        err_l2_l2 += sqr(err_l2)*(t - t_prec);
26        err_linf_linf = max(err_linf_linf, err_linf);
27        dout << put (t, phi_h, pi_h_phi);
28        derr << t << "\t" << err_l2 << "\t" << err_linf << endl;
29    }
30    derr << "# error_l2_l2      = " << sqrt(err_l2_l2) << endl;
31    derr << "# error_linf_linf = " << err_linf_linf << endl;
32    return (err_linf_linf <= tol) ? 0 : 1;
33 }

```

The numerical error  $\phi_h - \pi_h(\phi)$  is computed as:

```

field pi_h_phi = interpolate (Xh, phi(d,nu,t));
field eh = phi_h - pi_h_phi;

```

and its  $L^2$  norm is printed on the standard error. Observe the use of the branch class as both input and output field stream.

```

make convect_error
./convect_error < line2.branch > line2-cmp.branch
branch line2-cmp.branch -gnuplot

```

The instantaneous  $L^2(\Omega)$  norm is printed at each time step and the total error in  $L^2([0, T]; L^2(\Omega))$  is finally printed at the end of the stream. A P2 approximation can be used as well:

```

./convect line2.geo P2 > line2.branch
branch line2.branch -gnuplot
./convect_error < line2.branch > line2-cmp.branch

```

On Fig. 2.15.left we observe the  $L^2(L^2)$  convergence versus  $h$  for the  $P_1$  and  $P_2$  elements when  $d = 1$ : the errors reaches a plateau that decreases versus  $\Delta t$ . On Fig. 2.15.right the  $L^\infty(L^\infty)$  norm of the error presents a similar behavior. Since the plateau are equispaced, the convergence versus  $\Delta t$  is of first order.

These computation was performed for a convection-diffusion problem with  $\nu = 10^{-2}$ . The pure transport problem ( $\nu = 0$ , without diffusion) computation is obtained by:

```

./convect line2.geo P1 0 > line2.branch
branch line2.branch -gnuplot

```

Let us turn to the two-dimensional test case:

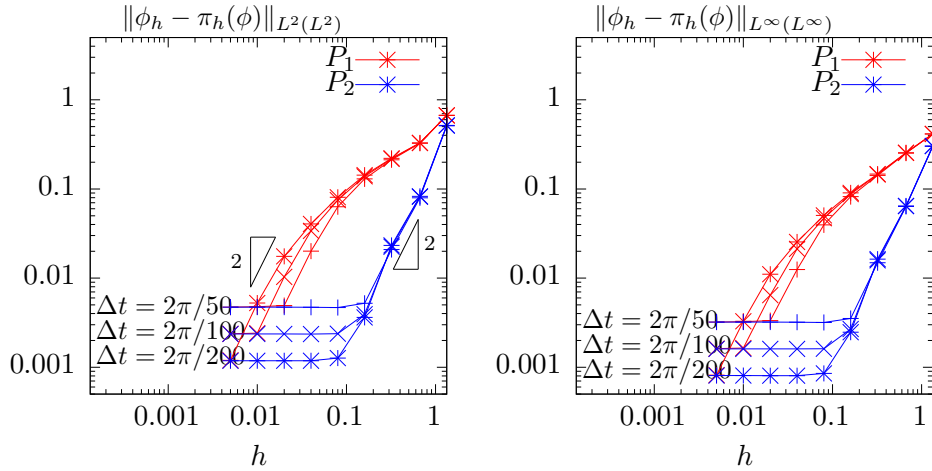


Figure 2.15: Diffusion-convection when  $d = 1$  and  $\nu = 10^{-2}$ : convergence versus  $h$  and  $\Delta t$  for  $P_1$  and  $P_2$  elements: (left) in  $L^2(L^2)$  norm; (right) in  $L^\infty(L^\infty)$  norm.

```
mkgeo_grid -t 80 -a -2 -b 2 -c -2 -d 2 > square2.geo
./convect square2.geo P1 > square2.branch
branch square2.branch -paraview
paraview &
```

The visualization and animation are similar to those of the head problem previously presented in paragraph 2.3.1. Observe the rotating hill. The result is shown on Fig. 2.14. The error analysis writes:

```
./convect_error < square2.branch > square2-cmp.branch
branch square2-cmp.branch -paraview
```

From the paraview menu, you can visualize simultaneously both the approximate solution and the Lagrange interpolate of the exact one. Finally, the three-dimensional case:

```
mkgeo_grid -T 15 -a -2 -b 2 -c -2 -d 2 -f -2 -g 2 > cube2.geo
./convect cube2.geo P1 > cube2.branch
```

The visualization is similar to the two-dimensional case.

## 2.4 The Navier-Stokes equations

### Formulation

This longer example combines most functionalities presented in the previous examples.

Let us consider the Navier-Stokes problem for the driven cavity in  $\Omega = ]0, 1[^d$ ,  $d = 2, 3$ . Let  $Re > 0$  be the Reynolds number, and  $T > 0$  a final time. The problem writes:

(NS): find  $\mathbf{u} = (u_0, \dots, u_{d-1})$  and  $p$  defined in  $\Omega \times ]0, T[$  such that:

$$\begin{aligned} \operatorname{Re} \left( \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) - \operatorname{div}(2D(\mathbf{u})) + \nabla p &= 0 \text{ in } \Omega \times ]0, T[, \\ - \operatorname{div} \mathbf{u} &= 0 \text{ in } \Omega \times ]0, T[, \\ \mathbf{u}(t=0) &= 0 \text{ in } \Omega \times \{0, T\}, \\ \mathbf{u} &= (1, 0) \text{ on } \Gamma_{\text{top}} \times ]0, T[, \\ \mathbf{u} &= 0 \text{ on } (\Gamma_{\text{left}} \cup \Gamma_{\text{right}} \cup \Gamma_{\text{bottom}}) \times ]0, T[, \\ \frac{\partial u_0}{\partial \mathbf{n}} &= \frac{\partial u_1}{\partial \mathbf{n}} = u_2 = 0 \text{ on } (\Gamma_{\text{back}} \cup \Gamma_{\text{front}}) \times ]0, T[ \text{ when } d = 3, \end{aligned}$$

where  $D(\mathbf{u}) = (\nabla \mathbf{u} + \nabla \mathbf{u}^T)/2$ . This nonlinear problem is the natural extension of the linear Stokes problem, as presented in paragraph 2.4, page 78. The boundaries are represented on Fig. 2.1, page 42.

### Time approximation

Let  $\Delta t > 0$ . Let us consider the following backward second order scheme, for all  $\phi \in C^2([0, T])$  :

$$\frac{d\phi}{dt}(t) = \frac{3\phi(t) - 4\phi(t - \Delta t) + \phi(t - 2\Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2)$$

The problem is approximated by the following second-order implicit scheme (BDF2):

$$\begin{aligned} \operatorname{Re} \frac{3\mathbf{u}^{n+1} - 4\mathbf{u}^n \circ X^n + \mathbf{u}^{n-1} \circ X^{n-1}}{2\Delta t} - \operatorname{div}(2D(\mathbf{u}^{n+1})) + \nabla p^{n+1} &= 0 \text{ in } \Omega, \\ - \operatorname{div} \mathbf{u}^{n+1} &= 0 \text{ in } \Omega, \\ \mathbf{u}^{n+1} &= (1, 0) \text{ on } \Gamma_{\text{top}}, \\ \mathbf{u}^{n+1} &= 0 \text{ on } \Gamma_{\text{left}} \cup \Gamma_{\text{right}} \cup \Gamma_{\text{bottom}}, \\ \frac{\partial u_0^{n+1}}{\partial \mathbf{n}} = \frac{\partial u_1^{n+1}}{\partial \mathbf{n}} = u_2^{n+1} &= 0 \text{ on } \Gamma_{\text{back}} \cup \Gamma_{\text{front}} \text{ when } d = 3, \end{aligned}$$

where, following [14, 36]:

$$\begin{aligned} X^n(x) &= x - \Delta t \mathbf{u}^*(x) \\ X^{n-1}(x) &= x - 2\Delta t \mathbf{u}^*(x) \\ \mathbf{u}^* &= 2\mathbf{u}^n - \mathbf{u}^{n-1} \end{aligned}$$

It is a second order extension of the method previously introduced in paragraph 2.3.2 page 73. The scheme defines a second order recurrence for the sequence  $(\mathbf{u}^n)_{n \geq -1}$ , that starts with  $\mathbf{u}^{-1} = \mathbf{u}^0 = 0$ .

### Variational formulation

The variational formulation of this problem expresses:

(NS) $_{\Delta t}$ : find  $\mathbf{u}^{n+1} \in \mathbf{V}(1)$  and  $p^{n+1} \in L_0^2(\Omega)$  such that:

$$\begin{aligned} a(\mathbf{u}^{n+1}, \mathbf{v}) + b(\mathbf{v}, p^{n+1}) &= m(\mathbf{f}^n, \mathbf{v}), \quad \forall \mathbf{v} \in \mathbf{V}(0), \\ b(\mathbf{u}^{n+1}, q) &= 0, \quad \forall q \in L_0^2(\Omega), \end{aligned}$$

where

$$\mathbf{f}^n = \frac{\operatorname{Re}}{2\Delta t} (4\mathbf{u}^n \circ X^n - \mathbf{u}^{n-1} \circ X^n)$$

and

$$a(\mathbf{u}, \mathbf{v}) = \frac{3\operatorname{Re}}{2\Delta t} \int_{\Omega} \mathbf{u} \cdot \mathbf{v} \, dx + \int_{\Omega} 2D(\mathbf{u}) : D(\mathbf{v}) \, dx$$

and  $b(\cdot, \cdot)$  and  $\mathbf{V}(\alpha)$  was already introduced in paragraph 2.1.4, page 51, while studying the Stokes problem.



### Space approximation

The Taylor-Hood [49] finite element approximation of this generalized Stokes problem was also considered in paragraph 2.1.4, page 51. We introduce a mesh  $\mathcal{T}_h$  of  $\Omega$  and the finite dimensional spaces  $\mathbf{X}_h$ ,  $\mathbf{V}_h(\alpha)$  and  $Q_h$ . The approximate problem writes:

$$(NS)_{\Delta t, h}: \text{find } \mathbf{u}_h^{n+1} \in \mathbf{V}_h(1) \text{ and } p^{n+1} \in Q_h \text{ such that:}$$

$$\begin{aligned} a(\mathbf{u}_h^{n+1}, \mathbf{v}) + b(\mathbf{v}, p^{n+1}) &= m(\mathbf{f}_h^n, \mathbf{v}), & \forall \mathbf{v} \in \mathbf{V}_h(0), \\ b(\mathbf{u}_h^{n+1}, q) &= 0, & \forall q \in Q_h. \end{aligned} \quad (2.6)$$

where

$$\mathbf{f}_h^n = \frac{Re}{2\Delta t} (4\mathbf{u}_h^n \circ X^n - \mathbf{u}_h^{n-1} \circ X^n)$$

The problem reduces to a sequence resolution of a generalized Stokes problems.

Example file 2.19: `navier_stokes_solve.icc`

```

1  using namespace std;
2  int navier_stokes_solve (
3      Float Re, Float delta_t, field l0h, field& uh, field& ph,
4      size_t& max_iter, Float& tol, odistream *p_derr=0) {
5      const space& Xh = uh.get_space();
6      const space& Qh = ph.get_space();
7      string label = "navier-stokes-" + Xh.get_geo().name();
8      quadrature_option qopt;
9      qopt.set_family(quadrature_option::gauss_lobatto);
10     qopt.set_order(Xh.degree());
11     trial u (Xh), p (Qh);
12     test v (Xh), q (Qh);
13     form mp = integrate (p*q, qopt);
14     form m = integrate (dot(u,v), qopt);
15     form a = integrate (2*ddot(D(u),D(v)) + 1.5*(Re/delta_t)*dot(u,v), qopt);
16     form b = integrate (-div(u)*q, qopt);
17     solver sa (a.uu());
18     solver_abtb stokes (a.uu(), b.uu(), mp.uu());
19     if (p_derr != 0) *p_derr << "[" << label << "]" #n |du/dt|" << endl;
20     field uh1 = uh;
21     for (size_t n = 0; true; n++) {
22         field uh2 = uh1;
23         uh1 = uh;
24         field uh_star = 2.0*uh1 - uh2;
25         characteristic X1 (-delta_t*uh_star);
26         characteristic X2 (-2.0*delta_t*uh_star);
27         field l1h = integrate (dot(compose(uh1,X1),v), qopt);
28         field l2h = integrate (dot(compose(uh2,X2),v), qopt);
29         field lh = l0h + (Re/delta_t)*(2*l1h - 0.5*l2h);
30         stokes.solve (lh.u() - a.uu()*uh.b(), -(b.uu()*uh.b()),
31                     uh.set_u(), ph.set_u());
32         field duh_dt = (3*uh - 4*uh1 + uh2)/(2*delta_t);
33         Float residual = sqrt(m(duh_dt,duh_dt));
34         if (p_derr != 0) *p_derr << "[" << label << "]" " << n << " " << residual << endl;
35         if (residual < tol) {
36             tol = residual;
37             max_iter = n;
38             return 0;
39         }
40         if (n == max_iter-1) {
41             tol = residual;
42             return 1;
43         }
44     }
45 }

```

### Comments

The `navier_stokes_solve` function is similar to the '`stokes_cavity.cc`'. It solves here a generalized Stokes problem and manages a right-hand side  $\mathbf{f}_h$ :

```

characteristic X1 ( -delta_t*uh_star);
characteristic X2 (-2.0*delta_t*uh_star);
field l1h = integrate (compose(uh1,X1)*v, qopt);
field l2h = integrate (compose(uh2,X2)*v, qopt);
field lh = l0h + (Re/delta_t)*(2*l1h - 0.5*l2h);

```

This last computation is similar to those done in the ‘convect.cc’ example. The generalized Stokes problem is solved by the solver\_abtb class. The stopping criterion is related to the stationary solution or the maximal iteration number.

Example file 2.20: navier\_stokes\_cavity.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "navier_stokes_solve.icc"
5 #include "navier_stokes_criterion.icc"
6 #include "cavity.icc"
7 int main (int argc, char**argv) {
8     environment rheolef (argc, argv);
9     if (argc < 2) {
10         cerr << "usage: " << argv[0] << " <geo> <Re> <err> <n_adapt>" << endl;
11         exit (1);
12     }
13     geo omega (argv[1]);
14     adapt_option options;
15     Float Re = (argc > 2) ? atof(argv[2]) : 100;
16     options.err = (argc > 3) ? atof(argv[3]) : 1e-2;
17     size_t n_adapt = (argc > 4) ? atoi(argv[4]) : 5;
18     Float delta_t = 0.05;
19     options.hmin = 0.004;
20     options.hmax = 0.1;
21     space Xh = cavity::velocity_space (omega, "P2");
22     space Qh (omega, "P1");
23     field uh = cavity::velocity_field (Xh, 1.0);
24     field ph (Qh, 0);
25     field fh (Xh, 0);
26     for (size_t i = 0; true; i++) {
27         size_t max_iter = 1000;
28         Float tol = 1e-5;
29         navier_stokes_solve (Re, delta_t, fh, uh, ph, max_iter, tol, &derr);
30         odistream o (omega.name(), "field");
31         o << catchmark("Re") << Re << endl
32           << catchmark("delta_t") << delta_t << endl
33           << catchmark("u") << uh
34           << catchmark("p") << ph;
35         o.close();
36         if (i >= n_adapt) break;
37         field ch = navier_stokes_criterion (Re,uh);
38         omega = adapt (ch, options);
39         o.open (omega.name(), "geo");
40         o << omega;
41         o.close();
42         Xh = cavity::velocity_space (omega, "P2");
43         Qh = space (omega, "P1");
44         uh = cavity::velocity_field (Xh, 1.0);
45         ph = field (Qh, 0);
46         fh = field (Xh, 0);
47     }
48 }

```

Example file 2.21: navier\_stokes\_criterion.icc

```

1 field navier_stokes_criterion (Float Re, const field& uh) {
2     space T0h (uh.get_geo(), "P1d");
3     return interpolate (T0h, sqrt(Re*norm2(uh) + 4*norm2(D(uh))));
4 }

```

## Comments

The code performs a computation by using adaptive mesh refinement, in order to capture recirculation zones. The `adapt_option` declaration is used by `rheolef` to send options to the mesh generator. The code reuse the file `'cavity.icc'` introduced page 52. This file contains two functions that defines boundary conditions associated to the cavity driven problem.

The `criteria` function computes the adaptive mesh refinement criteria:

$$c_h = (Re|\mathbf{u}_h|^2 + 2|D(\mathbf{u}_h)|^2)^{1/2}$$

The `criteria` function is similar to those presented in the `'embankment_adapt.cc'` example.

## How to run the program

The mesh loop adaptation is initiated from a `bamg` mesh (see also appendix A.2.1).

```
bamg -g square.bamgcad -o square.bamg
bamg2geo square.bamg square.dmn > square.geo
```

Then, compile and run the Navier-Stokes solver for the driven cavity for  $Re = 100$ :

```
make navier_stokes_cavity
time mpirun -np 8 ./navier_stokes_cavity square.geo 100
```

The program performs a computation with  $Re = 100$ . By default the time step is  $\Delta t = 0.05$  and the computation loops for five mesh adaptations. At each time step, the program prints an approximation of the time derivative, and stops when a stationary solution is reached. The `mpirun -np 8` prefix allows a parallel and distributed run while the `time` one returns the wall (real) and the cumulated CPU used by the computation. These prefixes are optionnal and you can omit the `mpirun` one if you are running with a sequential installation of **Rheolef**. Then, we visualize the `'square-005.geo'` adapted mesh and its associated solution:

```
geo square-005.geo
field square-005.field.gz -velocity
```

Notice the `-scale` option that applies a multiplicative factor to the arrow length when plotting. The representation of the stream function writes:

```
make streamf_cavity
zcat square-005.field.gz | ./streamf_cavity | field -bw -n-iso-negative 10 -
```

The programs `'streamf_cavity.cc'`, already introduced page 57, is here reused. The last options of the `field` program draws isocontours of the stream function using lines, as shown on Fig. 2.16. The zero isovalue separates the main flow from recirculations, located in corners at the bottom of the cavity.

For  $Re = 400$  and  $1000$  the computation writes:

```
./navier_stokes_cavity square.geo 400
./navier_stokes_cavity square.geo 1000
```

The visualization of the cut of the horizontal velocity along the vertical median line writes:

```
field square-005.field.gz -comp 0 -cut -normal -1 0 -origin 0.5 0 -gnuplot
field square-005.field.gz -comp 1 -cut -normal 0 1 -origin 0 0.5 -gnuplot
```

Fig. 2.18 compare the cuts with data from [40], table 1 and 2 (see also [44]). Observe that the solution is in good agreement with these previous computations.

Finally, table 2.19 compares the primary vortex position and its associated stream function value. Notice also the good agreement with previous simulations. The stream function extremal values are obtained by:

```
zcat square-005.field.gz | ./streamf_cavity | field -min -
zcat square-005.field.gz | ./streamf_cavity | field -max -
```

The maximal value has not yet been communicated to our knowledge and is provided in table 2.19 for cross validation purpose. The small program that computes the primary vortex position is showed below.

```
make vortex_position
zcat square-005.field.gz | ./streamf_cavity | ./vortex_position
```

Example file 2.22: vortex\_position.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 int main (int argc, char** argv) {
4     environment rheolef (argc, argv);
5     check_macro (communicator().size() == 1, "please, use sequentially");
6     field psi_h;
7     din >> psi_h;
8     size_t idof_min = 0;
9     Float psi_min = std::numeric_limits<Float>::max();
10    for (size_t idof = 0, ndof = psi_h.ndof(); idof < ndof; idof++) {
11        if (psi_h.dof(idof) >= psi_min) continue;
12        psi_min = psi_h.dof(idof);
13        idof_min = idof;
14    }
15    const disarray<point>& xdof = psi_h.get_space().get_xdofs();
16    point xmin = xdof [idof_min];
17    dout << "xc\t\tyc\t\tps" << std::endl
18         << xmin[0] << "\t" << xmin[1] << "\t" << psi_min << std::endl;
19 }

```

For higher Reynolds number, Shen [102] showed in 1991 that the flow converges to a stationary state for Reynolds numbers up to 10 000; for Reynolds numbers larger than a critical value  $10\,000 < Re_1 < 10\,500$  and less than another critical value  $15\,000 < Re_2 < 16\,000$ , these authors founded that the flow becomes periodic in time which indicates a Hopf bifurcation; the flow loses time periodicity for  $Re \geq Re_2$ . In 1998, Ould Salihi [67] founded a loss of stationarity between 10 000 and 20 000. In 2002, Auteri et al. [10] estimated the critical value for the apparition of the first instability to  $Re_1 \approx 8018$ . In 2005, Erturk et al. [35] computed steady driven cavity solutions up to  $Re \leq 21\,000$ . Also in 2005, this result was infirmed by [37]: these authors estimated  $Re_1$  close to 8000, in agreement with [10]. The 3D driven cavity has been investigated in [57] by the method of characteristic (see also [56] for 3D driven cavity computations). In conclusion, the exploration of the driven cavity at large Reynolds number is a fundamental challenge in computational fluid dynamics.

Note that, instead of using a time-dependent scheme, that requires many time steps, it is possible to directly compute the stationary solution of the Navier-Stokes problem, thanks to a nonlinear solver. Such approach is presented in the second volume of the Rheolef documentation [93], based on discontinuous FEM methods.

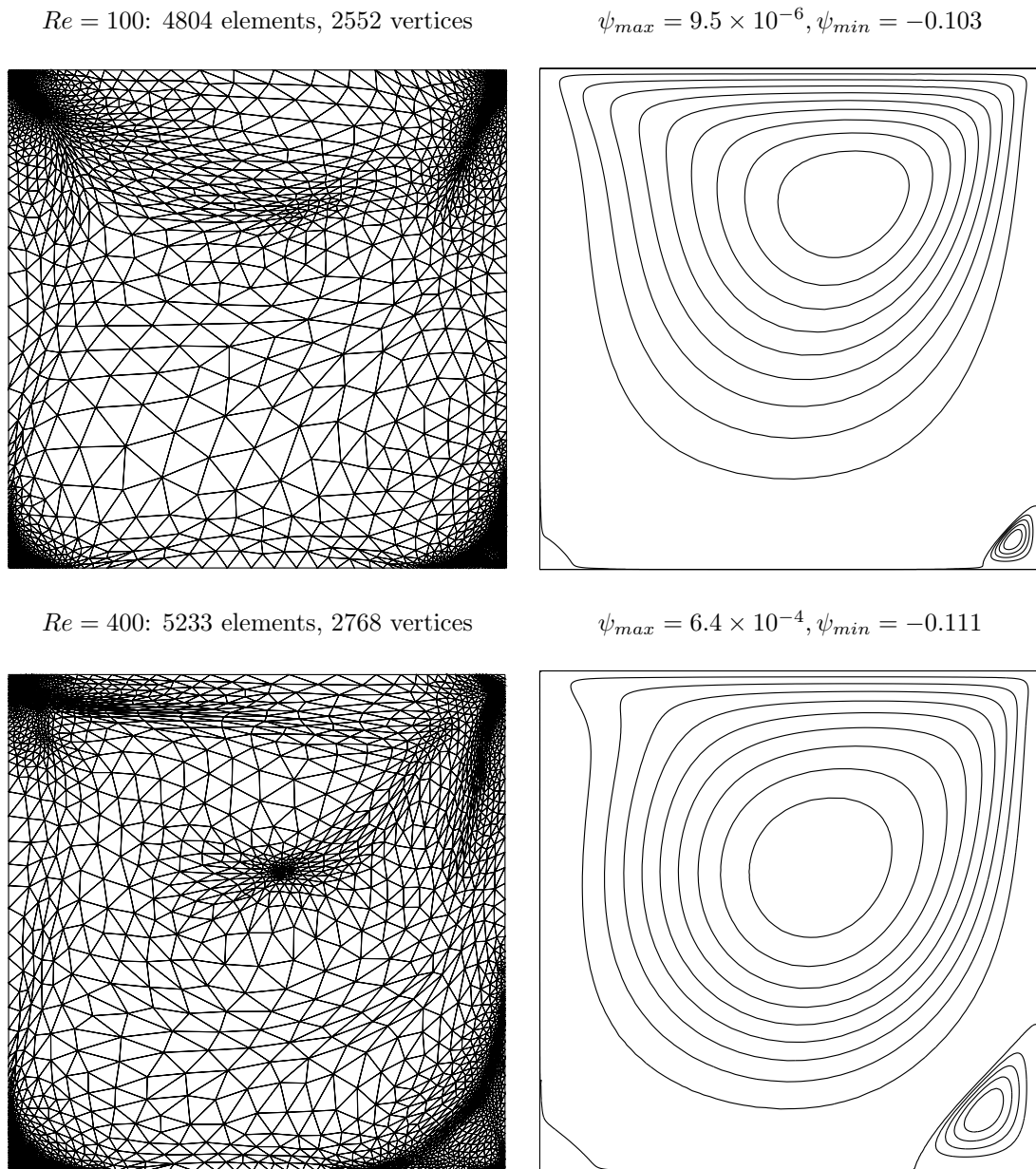


Figure 2.16: Meshes and stream functions associated to the solution of the Navier-Stokes equations for  $Re = 100$  (top) and  $Re = 400$  (bottom).

$Re = 1000$ : 5873 elements, 3106 vertices

$\psi_{max} = 1.64 \times 10^{-3}, \psi_{min} = -0.117$

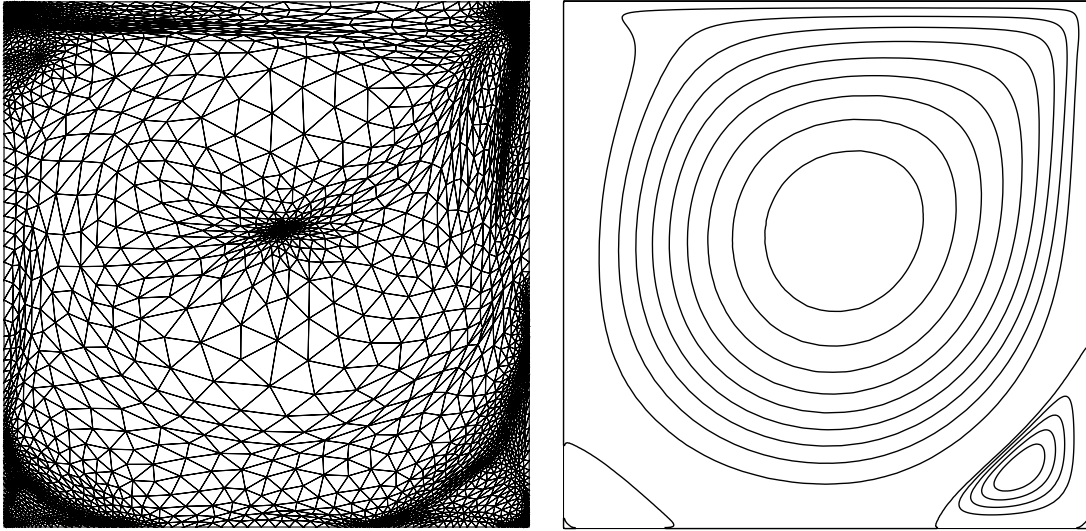


Figure 2.17: Meshes and stream functions associated to the solution of the Navier-Stokes equations for  $Re = 1000$ .

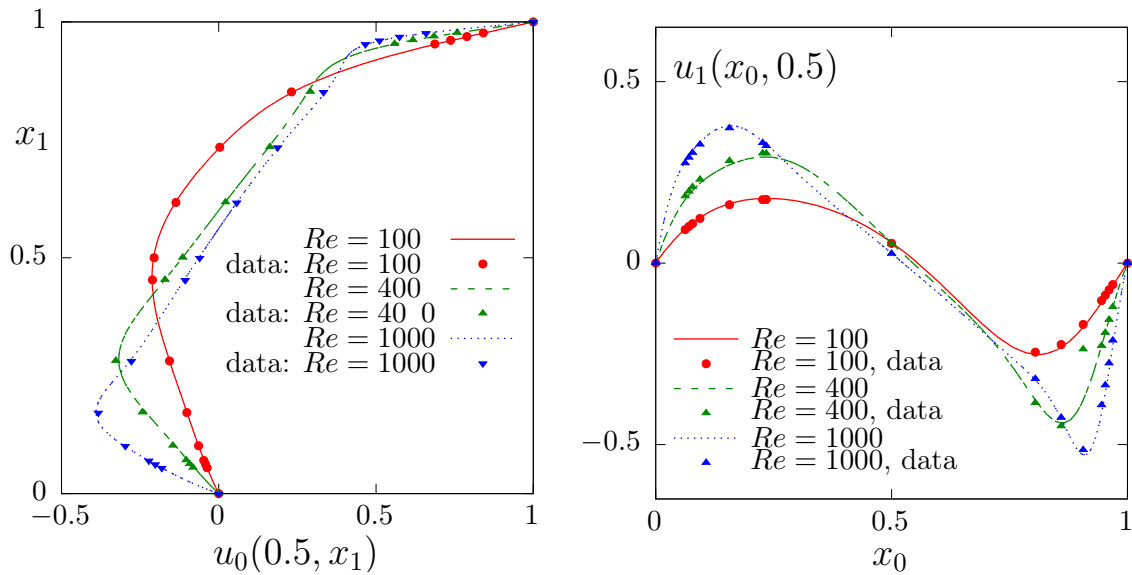


Figure 2.18: Navier-Stokes: velocity profiles along lines passing through the center of the cavity, compared with data from [40]: (a)  $u_0$  along the vertical line; (b)  $u_1$  along the horizontal line.

$Re$		$x_c$	$y_c$	$-\psi_{\min}$	$\psi_{\max}$
100	present	0.613	0.738	0.103	$9.5 \times 10^{-6}$
	Labeur and Wells [55]	0.608	0.737	0.104	-
	Donea and Huerta [33]	0.62	0.74	0.103	-
400	present	0.554	0.607	0.111	$5.6 \times 10^{-4}$
	Labeur and Wells [55]	0.557	0.611	0.115	-
	Donea and Huerta [33]	0.568	0.606	0.110	-
1000	present	0.532	0.569	0.117	$1.6 \times 10^{-3}$
	Labeur and Wells [55]	0.524	0.560	0.121	-
	Donea and Huerta [33]	0.540	0.573	0.110	-

Figure 2.19: Cavity flow: primary vortex position and stream function value.

## Chapter 3

# Advanced and highly nonlinear problems

### 3.1 Equation defined on a surface

This chapter deals with equations defined on a closed hypersurface. We present three different numerical methods: the direct resolution of the problem on an explicit surface mesh generated independently of **Rheolef**, the direct resolution on a surface mesh generated by **Rheolef** from a volume mesh, and finally a level set type method based on a volume mesh in an  $h$ -narrow band containing the surface. This last method allows to define hybrid operators between surface and volume-based finite element fields. These methods are demonstrated on two model problems and two different surfaces.

Let us consider a closed surface  $\Gamma \in \mathbb{R}^d$ ,  $d = 2$  or  $3$  and  $\Gamma$  is a connected  $C^2$  surface of dimension  $d - 1$  with  $\partial\Gamma = \emptyset$ . We first consider the following problem:

(P1) find  $u$ , defined on  $\Gamma$  such that:

$$u - \Delta_s u = f \text{ on } \Gamma \quad (3.1)$$

where  $f \in L^2(\Gamma)$ . For all function  $u$  defined on  $\Gamma$ ,  $\Delta_s$  denotes the Laplace-Beltrami operator:

$$\Delta_s u = \operatorname{div}_s(\nabla_s u)$$

where  $\nabla_s$  and  $\operatorname{div}_s$  are the tangential derivative and the surface divergence along  $\Gamma$ , defined respectively, for all scalar field  $\varphi$  and vector field  $\mathbf{v}$  by:

$$\begin{aligned} \nabla_s \varphi &= (I - \mathbf{n} \otimes \mathbf{n}) \nabla \varphi \\ \operatorname{div}_s \mathbf{v} &= (I - \mathbf{n} \otimes \mathbf{n}) : \nabla \mathbf{v} \end{aligned}$$

Here,  $\mathbf{n}$  denotes a unit normal on  $\Gamma$ .

We also consider the following variant of this problem:

(P2) find  $u$ , defined on  $\Gamma$  such that:

$$-\Delta_s u = f \text{ on } \Gamma \quad (3.2)$$

This second problem is similar to the first one: the Helmholtz operator  $I - \Delta_s$  has been replaced by the Laplace-Beltrami one  $-\Delta_s$ . In that case, the solution is defined up to a constant: if  $u$  is a solution, then  $u + c$  is also a solution for any constant  $c \in \mathbb{R}$ . Thus, we refer to (P1) as the Helmholtz-Beltrami problem and to (P2) as the Laplace-Beltrami one.



### 3.1.1 Approximation on an explicit surface mesh

#### The Helmholtz-Beltrami problem

Tanks to the surface Green formula (see appendix A.1.3), the variational formulation of problem (P1) writes:

(VF1): find  $u \in H^1(\Gamma)$  such that:

$$a(u, v) = l(v), \quad \forall v \in H^1(\Gamma)$$

where for all  $u, v \in H^1(\Gamma)$ ,

$$\begin{aligned} a(u, v) &= \int_{\Gamma} (u v + \nabla_s u \cdot \nabla_s v) \, ds \\ l(v) &= \int_{\Gamma} f v \, ds \end{aligned}$$

Let  $k \geq 1$  and consider a  $k$ -th order curved surface finite element mesh  $\Gamma_h$  of  $\Gamma$ . We define the space  $W_h$ :

$$W_h = \{v_h \in H^1(\Gamma_h); v|_S \in P_k, \forall S \in \Gamma_h\}$$

The approximate problem writes:

(VF1)<sub>h</sub>: find  $u_h \in W_h$  such that:

$$a(u_h, v_h) = l(v_h), \quad \forall v_h \in W_h$$

Example file 3.1: helmholtz\_s.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "sphere.icc"
5 int main(int argc, char**argv) {
6     environment rheolef(argc, argv);
7     geo gamma (argv[1]);
8     size_t d = gamma.dimension();
9     space Wh (gamma, argv[2]);
10    trial u (Wh); test v (Wh);
11    form a = integrate (u*v + dot(grad_s(u), grad_s(v)));
12    field lh = integrate (f(d)*v);
13    field uh (Wh);
14    solver sa (a.uu());
15    uh.set_u() = sa.solve (lh.u() - a.ub()*uh.b());
16    dout << uh;
17 }

```

#### Comments

The problem involves the Helmholtz operator and thus, the code is similar to ‘neumann-nh.cc’ presented page 28. Let us comments the only differences:

```
form a = integrate (u*v + dot(grad_s(u), grad_s(v)));
```

The form refers to the `grad_s` operator instead of the `grad` one, since only the coordinates related to the surface are involved.

```
field lh = integrate (f(d)*v);
```

The right-hand-side does not involve any boundary term, since the surface  $\Gamma$  is closed: the boundary domain  $\partial\Gamma = \emptyset$ . As test problem, the surface  $\Gamma$  is the unit circle when  $d = 2$  and the unit

sphere when  $d = 3$ . The data  $f$  has been chosen as in [29, p. 17]. This choice is convenient since the exact solution is known. Recall that the spherical coordinates  $(\rho, \theta, \phi)$  are defined from the cartesian ones  $(x_0, x_1, x_2)$  by:

$$\rho = \sqrt{x_0^2 + x_1^2 + x_2^2}, \quad \phi = \arccos(x_2/\rho), \quad \theta = \begin{cases} \arccos(x_0/\sqrt{x_0^2 + x_1^2}) & \text{when } x_1 \geq 0 \\ 2\pi - \arccos(x_0/\sqrt{x_0^2 + x_1^2}) & \text{otherwise} \end{cases}$$

Example file 3.2: sphere.icc

```

1 struct p {
2   Float operator() (const point& x) const {
3     if (d == 2) return 26*(pow(x[0],5) - 10*pow(x[0],3)*sqr(x[1])
4                   + 5*x[0]*pow(x[1],4));
5     else return 3*sqr(x[0])*x[1] - pow(x[1],3);
6   }
7   p (size_t d1) : d(d1) {}
8   protected: size_t d;
9 };
10 struct f {
11   Float operator() (const point& x) const {
12     if (d == 2) return _p(x)/pow(norm(x),5);
13     else return alpha*_p(x);
14   }
15   f (size_t d1) : d(d1), _p(d1) {
16     Float pi = acos(Float(-1));
17     alpha = -(13./8.)*sqrt(35./pi);
18   }
19   protected: size_t d; p _p; Float alpha;
20 };
21 struct u_exact {
22   Float operator() (const point& x) const {
23     if (d == 2) return _f(x)/(25+sqr(norm(x)));
24     else return sqr(norm(x))/(12+sqr(norm(x)))*_f(x);
25   }
26   u_exact (size_t d1) : d(d1), _f(d1) {}
27   protected: size_t d; f _f;
28 };
29 Float phi (const point& x) { return norm(x) - 1; }
```

### How to run the program

The program compile as usual:

```
make helmholtz_s
```

A mesh of a circle is generated by:

```
mkgeo_ball -s -e 100 > circle.geo
geo circle
```

The `mkgeo_ball` is a convenient script that generates a mesh with the `gmsh` mesh generator. Then, the problem resolution writes:

```
./helmholtz_s circle P1 > circle.field
field circle.field
field circle.field -elevation
```

The tridimensional case is similar:

```
mkgeo_ball -s -t 10 > sphere.geo
geo sphere.geo -stereo
```

```
./helmholtz_s sphere.geo P1 > sphere.field
field sphere.field
field sphere.field -stereo -gray
```

The solution is represented on Fig .3.1.left.

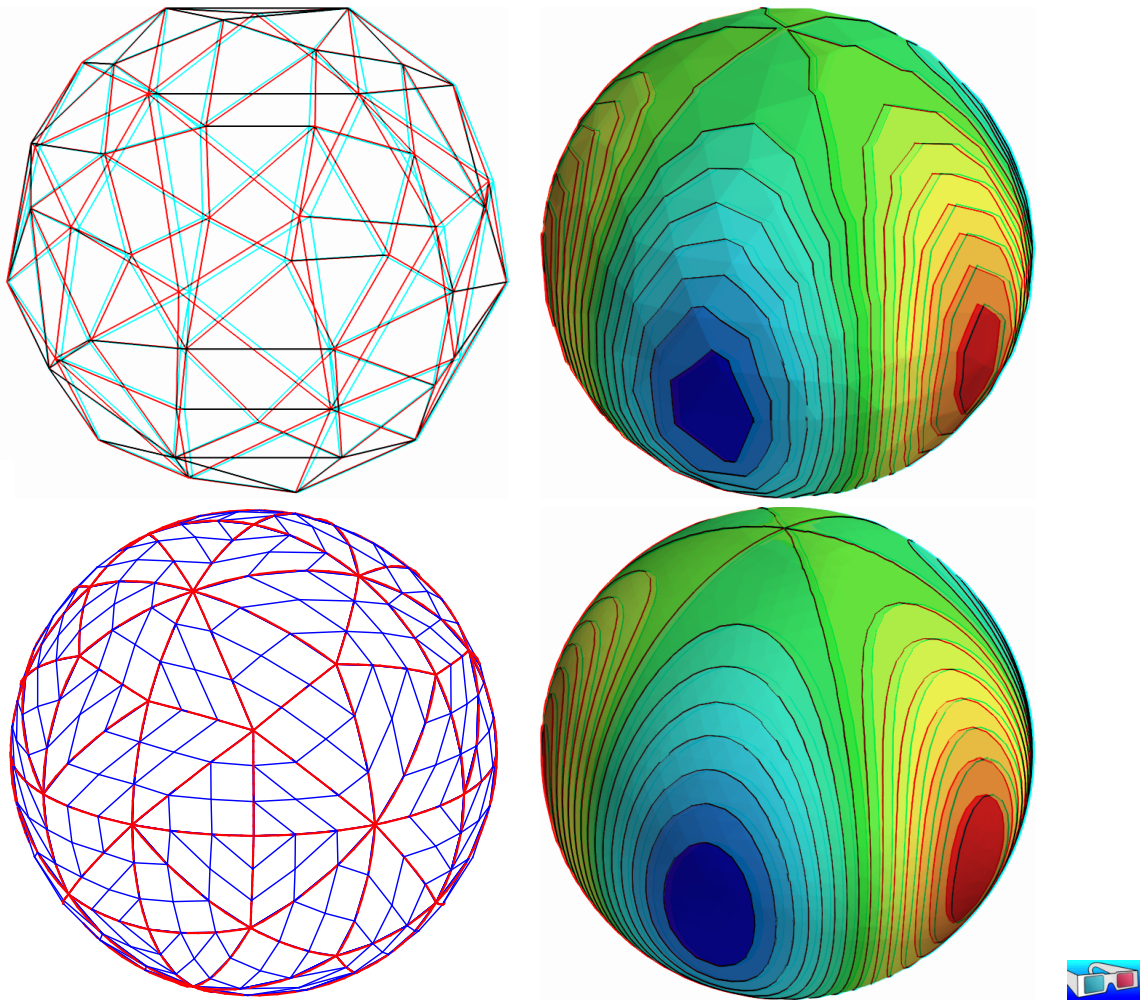


Figure 3.1: Helmholtz-Beltrami problem: high-order curved surface mesh and its corresponding isoparametric solution: (top)  $order = 1$ ; (bottom)  $order = 3$ .

Higher-order isoparametric finite elements can be considered for the curved geometry:

```
mkgeo_ball -s -e 30 -order 3 > circle-P3.geo
geo circle-P3.geo -subdivide 10
```

Observe the curved edges (see Fig .3.1). The `-subdivide` option allows a graphical representation of the curved edges by subdividing each edge in ten linear parts, since graphical softwares are not yet able to represent curved elements. The computation with the  $P_3$  isoparametric approximation writes:

```
./helmholtz_s circle-P3 P3 > circle-P3.field
field circle-P3.field -elevation
```

Notice that both the curved geometry and the finite element are second order. The tridimensional counterpart writes simply:

```

mkgeo_ball -s -t 10 -order 3 > sphere-P3.geo
geo sphere-P3.geo
./helmholtz_s sphere-P3 P3 > sphere-P3.field
field sphere-P3.field
field sphere-P3.field -stereo -gray

```

The solution is represented on Fig. 3.1).right-bottom. The graphical representation is not yet able to represent the high-order approximation: each elements is subdivided and a piecewise linear representation is used in each sub-elements.

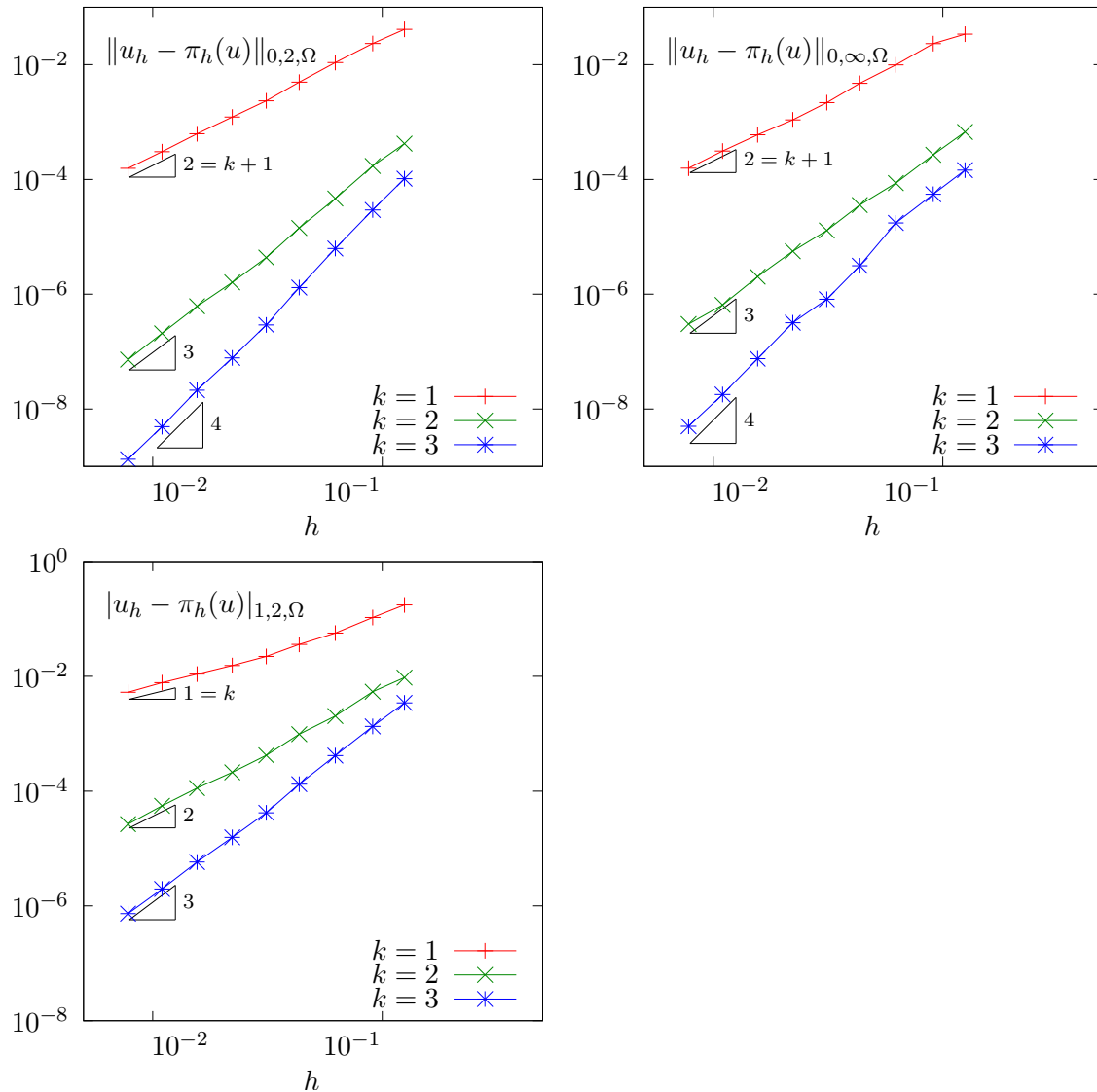


Figure 3.2: Curved non-polynomial surface: error analysis in  $L^2$ ,  $L^\infty$  and  $H^1$  norms.

Since the exact solution is known, the error can be computed: this is done by the program `helmholtz_s_error.cc`. This file is not presented here, as it is similar to some other examples, but can be found in the **Rheolef** example directory. Figure 3.2 plots the error in various norms versus element size for different isoparametric approximations.

## The Laplace-Beltrami problem

This problem has been introduced in (3.2), page 87. While the treatment of the Helmholtz-Beltrami problem was similar to the Helmholtz problem with Neumann boundary conditions, here, the treatment of the Laplace-Beltrami problem is similar to the Laplace problem with Neumann boundary conditions: see section 1.4, page 31. Notice that for both problems, the solution is defined up to a constant. Thus, the linear problem has a singular matrix. The ‘laplace\_s.cc’ code is similar to the ‘neumann-laplace.cc’ one, as presented in section 1.4. The only change lies one the definition of the right-hand side.

Example file 3.3: laplace\_s.cc

```

1  #include "rheolef.h"
2  using namespace rheolef;
3  using namespace std;
4  #include "torus.icc"
5  int main (int argc, char**argv) {
6      environment rheolef (argc, argv);
7      geo gamma (argv[1]);
8      size_t d = gamma.dimension();
9      space Wh (gamma, argv[2]);
10     trial u (Wh); test v (Wh);
11     form m = integrate (u*v);
12     form a = integrate (dot(grad_s(u), grad_s(v)));
13     field b = m*field(Wh,1);
14     field lh = integrate (f(d)*v);
15     csr<Float> A = {{ a.uu(),      b.u()},
16                   {trans(b.u()), 0  }};
17     vec<Float> B = { lh.u(),      0  };
18     solver sa (A);
19     vec<Float> U = sa.solve (B);
20     field uh(Wh);
21     uh.set_u() = U [range(0,uh.u().size())];
22     dout << uh;
23 }

```

Example file 3.4: torus.icc

```

1 static const Float R = 1;
2 static const Float r = 0.6;
3 Float phi (const point& x) {
4   return sqrt(sqrt(sqr(x[0])+sqr(x[1]))-sqr(R)) + sqrt(x[2]-sqr(r));
5 }
6 void get_torus_coordinates (const point& x,
7                             Float& rho, Float& theta, Float& phi) {
8   static const Float pi = acos(Float(-1));
9   rho = sqrt(sqr(x[2]) + sqr(sqrt(sqr(x[0]) + sqr(x[1])) - sqr(R)));
10  phi = atan2(x[1], x[0]);
11  theta = atan2(x[2], sqrt(sqr(x[0]) + sqr(x[1])) - R);
12 }
13 struct u_exact {
14   Float operator() (const point& x) const {
15     Float rho, theta, phi;
16     get_torus_coordinates (x, rho, theta, phi);
17     return sin(3*phi)*cos(3*theta+phi);
18   }
19   u_exact (size_t d=3) {}
20 };
21 struct f {
22   Float operator() (const point& x) const {
23     Float rho, theta, phi;
24     get_torus_coordinates (x, rho, theta, phi);
25     Float fx = (9*sin(3*phi)*cos(3*theta+phi))/sqr(r)
26               - (-10*sin(3*phi)*cos(3*theta+phi) - 6*cos(3*phi)*sin(3*theta+phi))
27               /sqr(R + r*cos(theta))
28               - (3*sin(theta)*sin(3*phi)*sin(3*theta+phi))
29               /(r*(R + r*cos(theta)));
30     return fx;
31   }
32   f (size_t d=3) {}
33 };

```

As test problem, the surface  $\Gamma$  is the a torus when  $d = 3$ . The data  $f$  has been chosen as in [65, p. 3355]. This choice is convenient since the exact solution is known. Let  $R$  and  $r$  denotes the large and small torus radii, respectively. The torus coordinates  $(\rho, \theta, \phi)$  are defined linked to the Cartesian ones by:

$$\begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} = R \begin{pmatrix} \cos(\phi) \\ \sin(\phi) \\ 0 \end{pmatrix} + \rho \begin{pmatrix} \cos(\phi) \cos(\theta) \\ \sin(\phi) \cos(\theta) \\ \sin(\theta) \end{pmatrix}$$

Here  $\rho$  is the distance from the point to the circle in the  $x_0x_1$  plane around 0 with radius  $R$ ,  $\theta$  is the angle from the positive  $(x_0, x_1, 0)$  to  $x_0$  and  $\phi$  is the angle from the positive  $x_0$  axis to  $(x_0, x_1, 0)$ .

### How to run the program ?

The surface mesh of the torus is generated by:

```

gmsht -2 torus.mshcad -o torus.msh
msh2geo torus.msh > torus.geo
geo torus.geo -stereo

```

The ‘torus.mshcad’ is not presented here: it can be founded in the **Rheolef** example directory. Then, the computation and visualization writes:

```

make laplace_s
./laplace_s torus.geo P1 > torus.field
field torus.field
field torus.field -stereo -gray

```

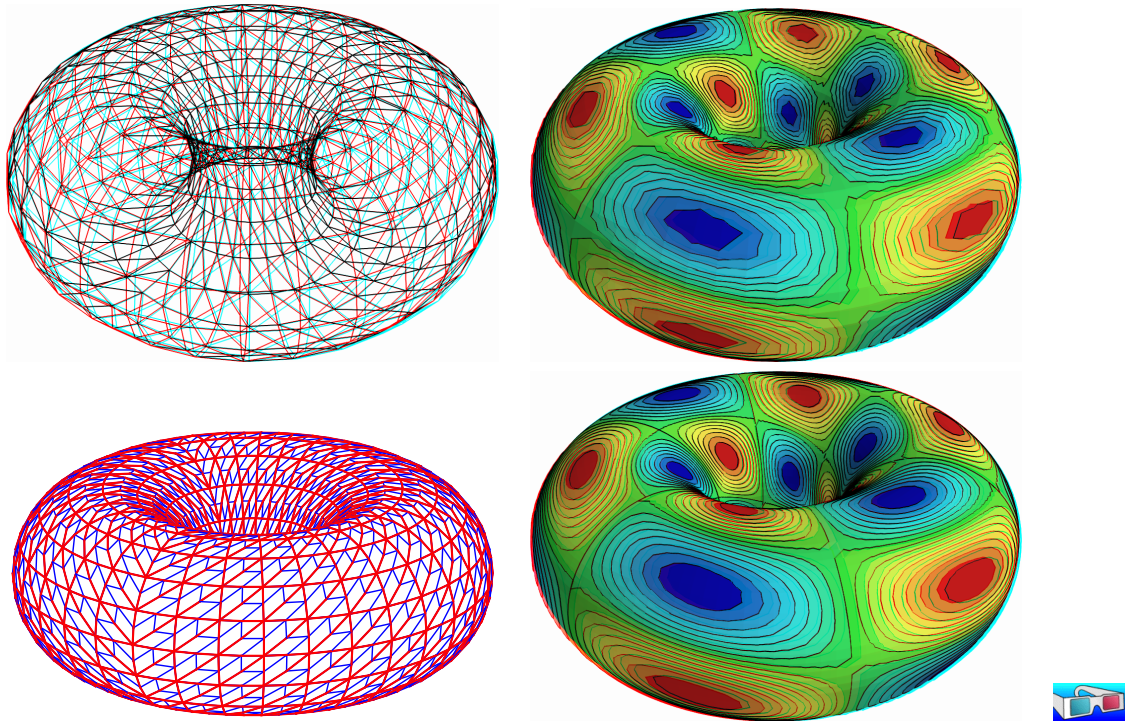


Figure 3.3: Laplace-Beltrami problem on a torus: high-order curved surface mesh and its corresponding isoparametric solution: (top)  $order = 1$ ; (bottom)  $order = 2$ .

For a higher-order approximation:

```
gmsht -2 -order 2 torus.mshcad -o torus-P2.msh
msh2geo torus-P2.msh > torus-P2.geo
geo torus-P2.geo -gnuplot
./laplace_s torus-P2.geo P2 > torus-P2.field
field torus-P2.field
```

The solution is represented on Fig. 3.3. By editing ‘[torus.mshcad](#)’ and changing the density of discretization, we can improve the approximate solution and converge to the exact solution. Due to a bug [98] in the current gmsht version 2.5.1 the convergence is not optimal  $\mathcal{O}(h^k)$  for higher values of  $k$ .

### 3.1.2 Building a surface mesh from a level set function

The previous method is limited to not-too-complex surface  $\Gamma$ , that can be described by a regular finite element surface mesh  $\Gamma_h$ . When the surface change, as in a time-dependent process, complex change of topology often occurs and the mesh  $\Gamma_h$  can degenerate or be too complex to be efficiently meshed. In that case, the surface is described implicitly as the zero isosurface, or zero *level set*, of a function:

$$\Gamma = \{x \in \Lambda; \phi(x) = 0\}$$

where  $\Lambda \subset \mathbb{R}^d$  is a bounding box of the surface  $\Gamma$ .

The following code automatically generates the mesh  $\Gamma_h$  of the surface described by the zero isosurface of a discrete  $\phi_h \in X_h$  level set function:

$$\Gamma_h = \{x \in \Lambda; \phi_h(x) = 0\}$$

where  $X_h$  is a piecewise affine functional space over a mesh  $\mathcal{T}_h$  of  $\Lambda$ :

$$X_h = \{\varphi \in L^2(\Lambda) \cap C^0(\Lambda); \varphi|_K \in P_1, \forall K \in \mathcal{T}_h\}$$

The polynomial approximation is actually limited here to first order: building higher order curved finite element surface meshes from a level set function is planned for the future versions of **Rheolef**.

Finally, a computation, as performed in the previous paragraph can be done using  $\Gamma_h$ . We also point out the limitations of this approach.

Example file 3.5: level\_set\_sphere.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "sphere.icc"
5 int main (int argc, char**argv) {
6     environment rheolef (argc,argv);
7     geo lambda (argv[1]);
8     level_set_option opts;
9     opts.split_to_triangle
10    = (argc > 2 && argv[2] == std::string("-tq")) ? false : true;
11     space Xh (lambda, "P1");
12     field phi_h = interpolate(Xh, phi);
13     geo gamma = level_set (phi_h, opts);
14     dout << gamma;
15 }
```

#### Comments

All the difficult work of building the intersection mesh  $\Gamma_h$ , defined as the zero level set of the  $\phi_h$  function, is performed by the `level_set` function:

```
geo gamma = level_set (phi_h, opts);
```

When  $d = 3$ , intersected tetrahedra leads to either triangular or quadrangular faces. By default, quadrangular faces are split into two triangles. An optional `-tq` program flag allows to conserve quadrangles in the surface mesh: it set the `split_to_triangle` optional field to false.

#### How to run the program ?

After the compilation, generates the mesh of a bounding box  $\Lambda = [-2, 2]^d$  of the surface and run the program:

```

make level_set_sphere
mkgeo_grid -t 20 -a -2 -b 2 -c -2 -d 2 > square2.geo
./level_set_sphere square2.geo > circle.geo
geo circle.geo
```



The computation of the previous paragraph can be reused:

```
./helmholtz_s circle.geo P1 | field -
```

Notice that, while the bounding box mesh was uniform, the intersected mesh could present arbitrarily small edge length (see also Fig. 3.4):

```
geo -min-element-measure circle.geo
geo -max-element-measure circle.geo
```

Let us turn to the  $d = 3$  case:

```
mkgeo_grid -T 20 -a -2 -b 2 -c -2 -d 2 -f -2 -g 2 > cube2.geo
./level_set_sphere cube2.geo | geo -upgrade - > sphere.geo
geo sphere.geo -stereo
./helmholtz_s sphere.geo P1 | field -
```

While the bounding box mesh was uniform, the triangular elements obtained by intersecting the 3D bounding box mesh with the level set function can present arbitrarily irregular sizes and shapes (see also Fig. 3.4):

```
geo -min-element-measure -max-element-measure sphere.geo
```

Nevertheless, there is a recent theoretical guaranty [66] for the finite element method to converge on these irregular families of meshes.

This approach can be extended to the Laplace-Beltrami problem on a torus:

```
sed -e 's/sphere/torus/' < level_set_sphere.cc > level_set_torus.cc
make level_set_torus
./level_set_torus cube2.geo | geo -upgrade - > torus.geo
geo torus.geo -stereo
./laplace_s torus.geo P1 | field -
```

Note that the intersected mesh is also irregular:

```
geo -min-element-measure -max-element-measure torus.geo
```

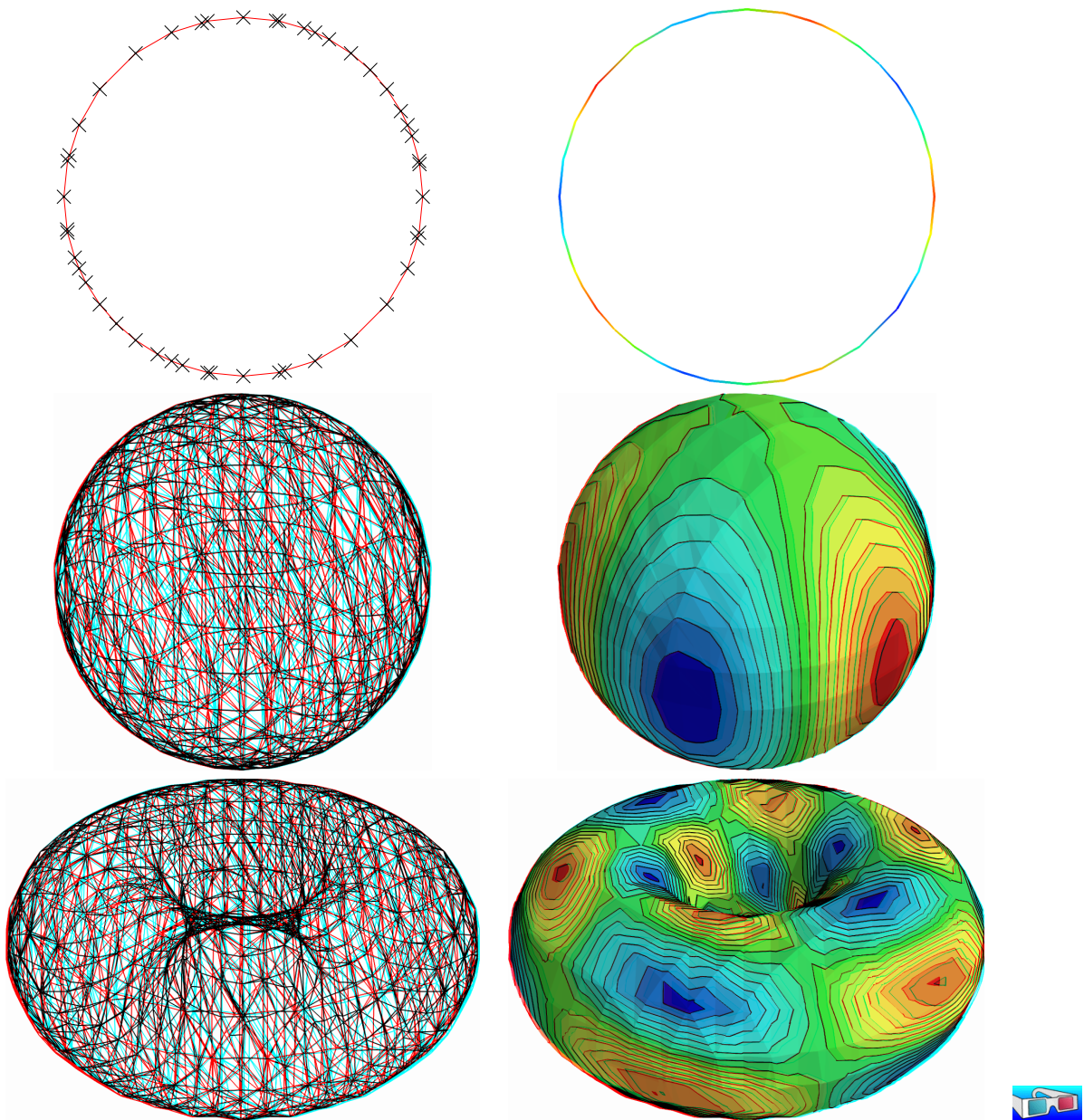


Figure 3.4: Building an explicit surface mesh from level set: (top) circle; (center) sphere; (bottom) torus.

### 3.1.3 The banded level set method

The banded level set method presents the advantages of the two previous methods without their drawback: it applies to very general geometries, as described by a level set function, and stronger convergence properties, as usual finite element methods. The previous intersection mesh can be circumvented by enlarging the surface  $\Gamma_h$  to a band  $\beta_h$  containing all the intersected elements of  $\mathcal{T}_h$  (see [2, 32, 65]):

$$\beta_h = \{K \in \mathcal{T}_h; K \cap \Gamma_h \neq \emptyset\}$$

Then, we introduce  $B_h$  the piecewise affine functional space over  $\beta_h$ :

$$B_h = \{v \in L^2(\beta_h) \cap C^0(\beta_h); v|_K \in P_1, \forall K \in \mathcal{T}_h\}$$

The problem is extended from  $\Gamma_h$  to  $\beta_h$  as:

$(VF)_h$ : find  $u_h \in B_h$  such that:

$$a(u_h, v_h) = l(v_h), \forall v_h \in B_h$$

where, for all  $u, v \in B_h$ ,

$$a(u, v) = \int_{\Gamma_h} (uv + \nabla_s u \cdot \nabla_s v) \, ds$$

$$l(v) = \int_{\Gamma_h} f v \, ds$$

for all  $u_h, v_h \in B_h$ . Notice that while  $u_h$  and  $v_h$  are defined over  $\beta_h$ , the summations in the variational formulations are restricted only to  $\Gamma_h \subset \beta_h$ .

Example file 3.6: `helmholtz_band_iterative.cc`

```

1 #include "rheolef.h"
2 using namespace std;
3 using namespace rheolef;
4 #include "sphere.icc"
5 int main (int argc, char**argv) {
6     environment rheolef(argc, argv);
7     geo lambda (argv[1]);
8     size_t d = lambda.dimension();
9     space Xh (lambda, "P1");
10    field phi_h = interpolate(Xh, phi);
11    band gamma_h (phi_h);
12    space Bh (gamma_h.band(), "P1");
13    trial u (Bh); test v (Bh);
14    form a = integrate (gamma_h, u*v + dot(grad_s(u), grad_s(v)));
15    field lh = integrate (gamma_h, f(d)*v);
16    field uh (Bh, 0);
17    solver_option sopt;
18    sopt.max_iter = 10000;
19    sopt.tol = 1e-10;
20    minres (a.uu(), uh.set_u(), lh.u(), eye(), sopt);
21    dout << catchmark("phi") << phi_h
22         << catchmark("u") << uh;
23 }

```

#### Comments

The band is build directly from the level set function as:

```
band gamma_h (phi_h);
```

The band structure is a small class that groups the surface mesh  $\Gamma_h$ , available as `gamma_h.level_set()`, and the  $\beta_h$  mesh, available as `gamma_h.band()`. It also manages some correspondance between both meshes. Then, the space of piecewise affine functions over the band is introduced:

```
space Bh (gamma_h.band(), "P1");
```

Next, two forms are computed by using the `integrate` function, with the band `gamma_h` as a domain-like argument:

```
form m = integrate (gamma_h, u*v);
form a = integrate (gamma_h, dot(grad_s(u), grad_s(v)));
```

The right-hand side also admits the `gamma_h` argument:

```
field lh = integrate (gamma_h, f(d)*v);
```

Recall that summations for both forms and right-hand side will be performed on  $\Gamma_h$ , represented by `gamma_h.level_set()`, while the approximate functional space is  $B_h$ . Due to this summation on  $\Gamma_h$  instead of  $\beta_h$ , the matrix of the system is singular [2, 64, 65] and the MINRES algorithm has been chosen to solve the linear system:

```
pminres (a.uu(), uh.set_u(), lh.u(), eye(), max_iter, tol, &derr);
```

The `eye()` argument represents here the identity preconditioner, i.e. no preconditioner at all. It has few influence of the convergence properties of the matrix and could be replaced by another simple one: the diagonal of the matrix `diag(a.uu())` without sensible gain of performance:

```
pminres (a.uu(), uh.set_u(), lh.u(), diag(a.uu()), max_iter, tol, &derr);
```

### How to run the program

The compilation and run writes:

```
make helmholtz_band_iterative
mkgeo_grid -T 20 -a -2 -b 2 -c -2 -d 2 -f -2 -g 2 > cube-20.geo
./helmholtz_band_iterative cube-20.geo > sphere-band.field
```

The run generates also two meshes (see Fig. 3.5): the intersection mesh and the band around it. The solution is here defined on this band: this extension has no interpretation in terms of the initial problem and can be restricted to the intersection mesh for visualization purpose:

```
make proj_band
./proj_band < sphere-band.field | field -
```

The ‘`proj_band.cc`’ is presented below. The run generates also the  $\Gamma_h$  mesh (see Fig. 3.5), required for the visualization. The two-dimensional case is obtained simply by replacing the 3D bounding box by a 2D one:

```
mkgeo_grid -t 20 -a -2 -b 2 -c -2 -d 2 > square-20.geo
./helmholtz_band_iterative square-20.geo > circle-band.field
./proj_band < circle-band.field | field -
./proj_band < circle-band.field | field -elevation -bw -stereo -
```

Example file 3.7: proj\_band.cc

```

1 #include "rheolef.h"
2 using namespace std;
3 using namespace rheolef;
4 int main (int argc, char**argv) {
5     environment rheolef (argc, argv);
6     field phi_h;
7     din >> catchmark("phi") >> phi_h;
8     const space& Xh = phi_h.get_space();
9     band gamma_h (phi_h);
10    space Bh (gamma_h.band(), "P1");
11    field uh(Bh);
12    din >> catchmark("u") >> uh;
13    space Wh (gamma_h.level_set(), "P1");
14    gamma_h.level_set().save();
15    dout << interpolate (Wh, uh);
16 }

```

### 3.1.4 Improving the banded level set method with a direct solver

The iterative algorithm previously used for solving the linear system is not optimal: for 3D problems on a surface, the bidimensionnal connectivity of the sparse matrix suggests that a direct sparse factorisation would be much more efficient.

Recall that  $\phi_h = 0$  on  $\Gamma_h$ . Thus, if  $u_h \in B_h$  is solution of the problem, then  $u_h + \alpha\phi_{h|\beta_h} \in B_h$  is also solution for any  $\alpha \in \mathbb{R}$ , where  $\phi_{h|\beta_h} \in B_h$  denotes the restriction of the level set function  $\phi_h \in X_h$  on the band  $\beta_h$ . Thus there is multiplicity of solutions and the matrix of the problem is singular. The direct resolution is still possible on a modified linear system with additional constraints in order to recover the unicity of the solution. We impose the constraint that the solution  $u_h$  should be othogonal to  $\phi_{h|\beta_h} \in B_h$ . In some special cases, the band is composed of several connected components (see Fig. 3.6): this appends when a vertex of the bounding box mesh belongs to  $\Gamma_h$ . In that case, the constaint could be expressed on each connected component. Fig. 3.6 shows also the case when a full side of an element is included in  $\Gamma_h$ : such an element of the band is called *isolated*.

Example file 3.8: helmholtz\_band.cc

```

1 #include "rheolef.h"
2 using namespace std;
3 using namespace rheolef;
4 #include "sphere.icc"
5 int main (int argc, char**argv) {
6     environment rheolef(argc, argv);
7     geo lambda (argv[1]);
8     size_t d = lambda.dimension();
9     space Xh (lambda, "P1");
10    field phi_h = interpolate(Xh, phi);
11    band gamma_h (phi_h);
12    field phi_h_band = phi_h [gamma_h.band()];
13    space Bh (gamma_h.band(), "P1");
14    Bh.block ("isolated");
15    Bh.unblock ("zero");
16    trial u (Bh); test v (Bh);
17    form a = integrate (gamma_h, u*v + dot(grad_s(u), grad_s(v)));
18    field lh = integrate (gamma_h, f(d)*v);
19    vector<vec<Float>> > b (gamma_h.n_connected_component());
20    vector<Float> z (gamma_h.n_connected_component(), 0);
21    for (size_t i = 0; i < b.size(); i++) {
22        const domain& cci = gamma_h.band() ["cc"+itos(i)];
23        field phi_h_cci (Bh, 0);
24        phi_h_cci [cci] = phi_h_band [cci];
25        b[i] = phi_h_cci.u();
26    }
27    csr<Float> A = { { a.uu(), trans(b)},
28                  { b, 0 } };
29    vec<Float> F = { lh.u(), z };
30    A.set_symmetry(true);
31    solver sa = ldlt(A);
32    vec<Float> U = sa.solve (F);
33    field uh(Bh,0);
34    uh.set_u() = U [range(0,uh.u().size())];
35    dout << catchmark("phi") << phi_h
36          << catchmark("u") << uh;
37 }

```

### Comments

The management of the special sides and vertices that are fully included in  $\Gamma_h$  is performed by:

```

Bh.block ("isolated");
Bh.unblock ("zero");

```

The addition of linear constraints is similar to the 'neumann-laplace.cc' code, as presented in section 1.4:

```

csr<Float> A = { { a.uu(), trans(b)},
                { b, 0 } };

```

Here `b` is a `vector<vec<Float>>`, i.e. a vector of linear constraints, one per connected component of the band  $\beta_h$ .

### How to run the program

The commands are similar to the previous iterative implementation, just replacing `helmholtz_band_iterative` by `helmholtz_band`.

This approach could be also adapted to the Laplace-Beltrami problem on the torus.

Example file 3.9: laplace\_band.cc

```

1 #include "rheolef.h"
2 using namespace std;
3 using namespace rheolef;
4 #include "torus.icc"
5 int main (int argc, char**argv) {
6     environment rheolef(argc, argv);
7     geo lambda (argv[1]);
8     size_t d = lambda.dimension();
9     space Xh (lambda, "P1");
10    field phi_h = interpolate(Xh, phi);
11    band gamma_h (phi_h);
12    field phi_h_band = phi_h [gamma_h.band()];
13    space Bh (gamma_h.band(), "P1");
14    Bh.block ("isolated");
15    Bh.unblock ("zero");
16    trial u (Bh); test v (Bh);
17    form m = integrate (gamma_h, u*v);
18    form a = integrate (gamma_h, dot(grad_s(u), grad_s(v)));
19    field lh = integrate (gamma_h, f(d)*v);
20    vector<vec<Float>> > b (gamma_h.n_connected_component());
21    vector<Float> z (gamma_h.n_connected_component(), 0);
22    for (size_t i = 0; i < b.size(); i++) {
23        const domain& cci = gamma_h.band() ["cc"+itos(i)];
24        field phi_h_cci (Bh, 0);
25        phi_h_cci [cci] = phi_h_band [cci];
26        b[i] = phi_h_cci.u();
27    }
28    field c = m*field(Bh,1);
29    csr<Float> A = { { a.uu(),          trans(b), c.u()},
30                  { b,                0,          0   },
31                  { trans(c.u()), 0,          0   } };
32    vec<Float> F = { lh.u(),          z,          0};
33    A.set_symmetry(true);
34    solver sa = ldlt(A);
35    vec<Float> U = sa.solve (F);
36    field uh(Bh,0);
37    uh.set_u() = U [range(0,uh.u().size())];
38    dout << catchmark("phi") << phi_h
39         << catchmark("u")   << uh;
40 }

```

## Comments

The code is similar to the previous one `helmholtz_band.cc`. Since the solution is defined up to a constant, an additional linear constraint has to be inserted:

$$\int_{\Gamma_h} u_h dx = 0$$

This writes:

```

field c = m*field(Bh,1);
csr<Float> A = { { a.uu(),          trans(b), c.u()},
                { b,                0,          0   },
                { trans(c.u()), 0,          0   } };

```

## How to run the program

```

make laplace_band
mkgeo_grid -T 20 -a -2 -b 2 -c -2 -d 2 -f -2 -g 2 > cube-20.geo
./laplace_band cube-20.geo > torus-band.field
./proj_band < torus-band.field | field -stereo -

```

The solution is represented on Fig. 3.5.bottom.

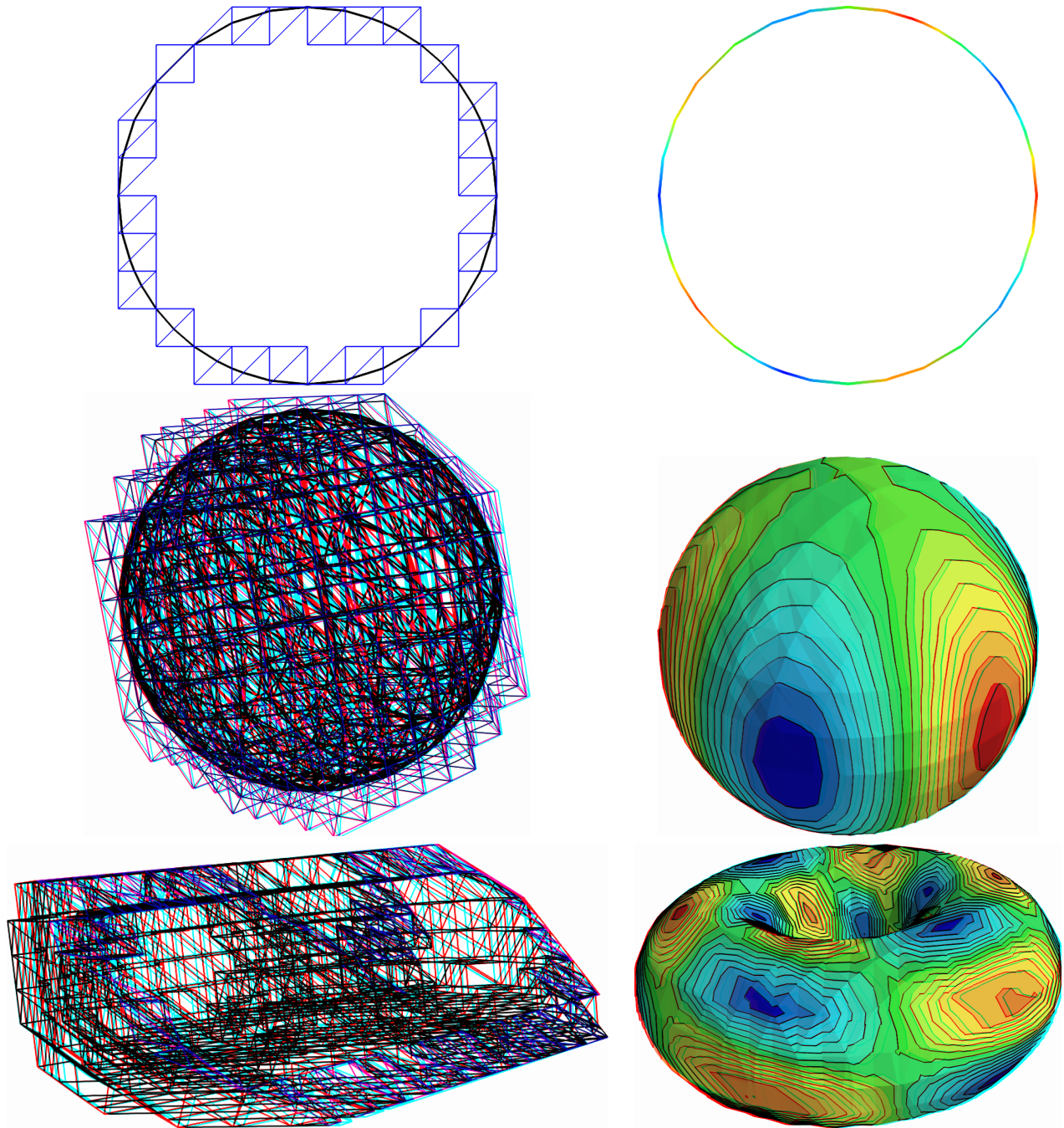


Figure 3.5: The banded level set method: (top) circle; (center) sphere; (bottom) torus.



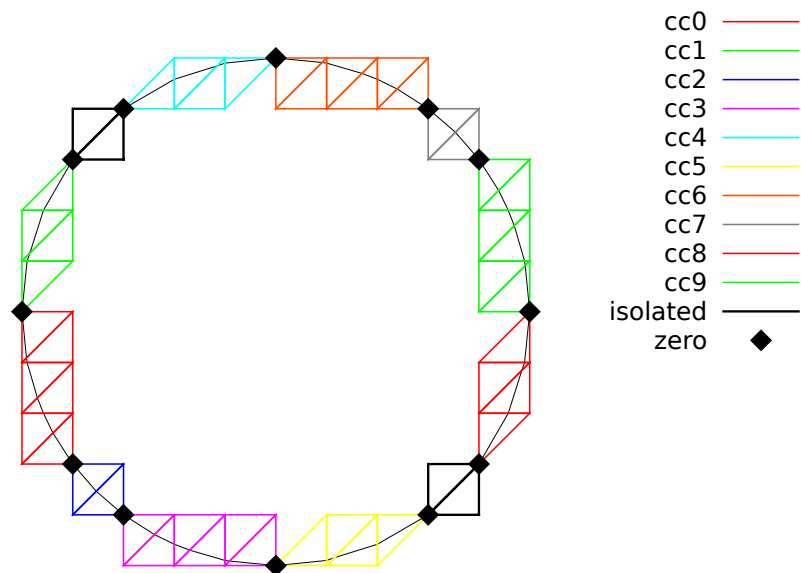


Figure 3.6: The banded level set method: the band is composed of several connected components.

## 3.2 The highly nonlinear $p$ -laplacian problem

### 3.2.1 Problem statement

Let us consider the classical  $p$ -Laplacian problem with homogeneous Dirichlet boundary conditions in a domain bounded  $\Omega \subset \mathbb{R}^d$ ,  $d = 1, 2, 3$ :

(P): find  $u$ , defined in  $\Omega$  such that:

$$\begin{aligned} -\operatorname{div}(\eta(|\nabla u|^2) \nabla u) &= f \text{ in } \Omega \\ u &= 0 \text{ on } \partial\Omega \end{aligned}$$

where  $\eta : z \in \mathbb{R}^+ \mapsto z^{\frac{p-2}{2}} \in \mathbb{R}^+$ . Several variants of the  $\eta$  can be considered: see [94] for practical and useful examples: this problem represents a pipe flow of a non-Newtonian power-law fluid. Here  $p \in ]1, +\infty[$  and  $f$  are known. For the computational examples, we choose  $f = 1$ . When  $p = 2$ , this problem reduces to the linear Poisson problem with homogeneous Dirichlet boundary conditions. Otherwise, for any  $p > 1$ , the nonlinear problem is equivalent to the following minimization problem:

(MP): find  $u \in W_0^{1,p}(\Omega)$  such that:

$$u = \arg \min_{v \in W_0^{1,p}(\Omega)} \frac{1}{2} \int_{\Omega} H(|\nabla v|^2) \, dx - \int_{\Omega} f v \, dx,$$

where  $H$  denotes the primitive of  $\eta$ :

$$H(z) = \int_0^z \eta(z) \, dz = \frac{2z^p}{p}$$

Here  $W_0^{1,p}(\Omega)$  denotes the usual Sobolev spaces of functions in  $W^{1,p}(\Omega)$ . We also assume that  $f \in W^{-1,p}(\Omega)$ , where  $W_0^{-1,p}(\Omega)$  denotes the dual space of  $W_0^{1,p}(\Omega)$  that vanishes on the boundary [16, p. 118]. The variational formulation of this problem expresses:

(VF): find  $u \in W_0^{1,p}(\Omega)$  such that:

$$a(u; u, v) = l(v), \quad \forall v \in W_0^{1,p}(\Omega)$$

where  $a(\cdot, \cdot)$  and  $l(\cdot)$  are defined for any  $u_0, u, v \in W^{1,p}(\Omega)$  by

$$a(u_0; u, v) = \int_{\Omega} \eta(|\nabla u_0|^2) \nabla u \cdot \nabla v \, dx, \quad \forall u, v \in W_0^{1,p}(\Omega) \quad (3.3)$$

$$l(v) = \int_{\Omega} f v \, dx, \quad \forall u, v \in L^2(\Omega) \quad (3.4)$$

The quantity  $a(u; u, u)^{1/p} = \|\nabla u\|_{0,p,\Omega}$  induces a norm in  $W_0^{1,p}$ , equivalent to the standard norm. The form  $a(\cdot; \cdot, \cdot)$  is bilinear with respect to the two last variable and is related to the *energy* form.

### 3.2.2 The fixed-point algorithm

#### Principle of the algorithm

This nonlinear problem is then reduced to a sequence of linear subproblems by using the fixed-point algorithm. The sequence  $(u^{(n)})_{n \geq 0}$  is defined by recurrence as:

- $n = 0$ : let  $u^{(0)} \in W_0^{1,p}(\Omega)$  be known.
- $n \geq 0$ : suppose that  $u^{(n)} \in W_0^{1,p}(\Omega)$  is known and find  $u^* \in W_0^{1,p}(\Omega)$  such that:

$$a(u^{(n)}; u^*, v) = l(v), \quad \forall v \in W_0^{1,p}(\Omega)$$

and then set

$$u^{(n+1)} = \omega u^* + (1 - \omega) * u^{(n)}$$

Here  $\omega > 0$  is the relaxation parameter: when  $\omega = 1$  we obtain the usual un-relaxed fixed point algorithm. For stiff nonlinear problems, we will consider the under-relaxed case  $0 < \omega < 1$ . Let  $u^{(n+1)} = G(u^{(n)})$  denotes the operator that solve the previous linear subproblem for a given  $u^{(n)}$ . Since the solution  $u$  satisfies  $u = G(u)$ , it is a fixed-point of  $G$ .

Let us introduce a mesh  $\mathcal{T}_h$  of  $\Omega$  and the finite dimensional space  $X_h$  of continuous piecewise polynomial functions and  $V_h$ , the subspace of  $X_h$  containing elements that vanishes on the boundary of  $\Omega$ :

$$\begin{aligned} X_h &= \{v_h \in C_0^0(\bar{\Omega}); v_{h/K} \in P_k, \forall K \in \mathcal{T}_h\} \\ V_h &= \{v_h \in X_h; v_h = 0 \text{ on } \partial\Omega\} \end{aligned}$$

where  $k = 1$  or  $2$ . The approximate problem expresses: suppose that  $u_h^{(n)} \in V_h$  is known and find  $u_h^* \in V_h$  such that:

$$a(u_h^{(n)}; u_h^*, v_h) = l(v_h), \forall v_h \in V_h$$

By developing  $u_h^*$  on a basis of  $V_h$ , this problem reduces to a linear system.

Example file 3.10: p\_laplacian\_fixed\_point.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "eta.icc"
5 #include "dirichlet.icc"
6 int main(int argc, char**argv) {
7     environment rheolef (argc,argv);
8     geo omega (argv[1]);
9     Float eps = std::numeric_limits<Float>::epsilon();
10    string approx = (argc > 2) ? argv[2] : "P1";
11    Float p = (argc > 3) ? atof(argv[3]) : 1.5;
12    Float w = (argc > 4) ? (is_float(argv[4]) ? atof(argv[4]) :2/p) :1;
13    Float tol = (argc > 5) ? atof(argv[5]) : 1e5*eps;
14    size_t max_it = (argc > 6) ? atoi(argv[6]) : 500;
15    derr << "# P-Laplacian problem by fixed-point:" << endl
16          << "# geo = " << omega.name() << endl
17          << "# approx = " << approx << endl
18          << "# p = " << p << endl
19          << "# w = " << w << endl
20          << "# tol = " << tol << endl;
21    space Xh (omega, approx);
22    Xh.block ("boundary");
23    trial u (Xh); test v (Xh);
24    form m = integrate (u*v);
25    solver sm (m.uu());
26    quadrature_option qopt;
27    qopt.set_family (quadrature_option::gauss);
28    qopt.set_order (2*Xh.degree()-1);
29    field uh (Xh);
30    uh ["boundary"] = 0;
31    field lh = integrate (v);
32    dirichlet (lh, uh);
33    derr << "# n r v" << endl;
34    Float r = 1, r0 = 1;
35    size_t n = 0;
36    do {
37        form a = integrate(compose(eta(p), norm2(grad(uh)))*dot(grad(u), grad(v)),
38                          qopt);
39        field mrh = a*uh - lh;
40        field rh (Xh, 0);
41        rh.set_u() = sm.solve (mrh.u());
42        r = rh.max_abs();
43        if (n == 0) { r0 = r; }
44        Float v = (n == 0) ? 0 : log10(r0/r)/n;
45        derr << n << " " << r << " " << v << endl;
46        if (r <= tol || n++ >= max_it) break;
47        solver sa (a.uu());
48        vec<Float> u_star = sa.solve (lh.u() - a.ub()*uh.b());
49        uh.set_u() = w*u_star + (1-w)*uh.u();
50    } while (true);
51    dout << catchmark("p") << p << endl
52          << catchmark("u") << uh;
53    return (r <= tol) ? 0 : 1;
54 }

```

### Comments

The implementation with **Rheolef** involves a weighted forms: the tensor-valued weight  $\eta \left( \left| \nabla u_h^{(n)} \right|^2 \right)$  is inserted in the variational expression passed to the `integrate` function. The construction of the weighted form `a(.,.,.)` writes:

```

form a = integrate(compose(eta(p), norm2(grad(uh)))*dot(grad(u), grad(v)),
                  qopt);

```

Remarks the usage of the `compose`, `norm2` and `grad` library functions. The weight  $\eta \left( \left| \nabla u_h^{(n)} \right|^2 \right)$  is represented by the `compose(eta(p), norm2(grad(uh)))` sub-expression. This weight is evaluated on the fly at the quadrature nodes during the assembly process implemented by the `integrate` function. Also, notice the distinction between  $u_h$ , that represents the value of the solution at step  $n$ , and the trial  $u$  and test  $v$  functions, that represents any elements of the function space  $X_h$ . These functions appear in the `dot(grad(u), grad(v))` sub-expression. As the integrals involved by this weighted form cannot be computed exactly for a general  $\eta$  function, a quadrature formula is used:

$$\int_K f(x) dx = \sum_{q=0}^{n_K-1} f(x_{K,q}) \omega_{K,q} + \mathcal{O}(h^{k'+1})$$

where  $(x_{K,q}, \omega_{K,q})_{0 \leq q < n_K}$  are the quadrature nodes and weights on  $K$  and  $k'$  is the order of the quadrature: when  $f$  is a polynomial of degree less than  $k'$ , the integral is exact. The bilinear form  $a(.,.)$  introduced in (3.3) is then re-defined for all  $u_0, u, v \in X_h$  by:

$$a(u_0; u, v) = \sum_{K \in \mathcal{T}_h} \sum_{q=0}^{n_K-1} \eta \left( \left| \nabla u_0(x_{K,q}) \right|^2 \right) \nabla u(x_{K,q}) \cdot \nabla v(x_{K,q}) \omega_{K,q} \quad (3.5)$$

We choose the Gauss quadrature formula and the order  $k'$  is chosen as  $k' = 2k - 1$ : the number  $n_K$  of nodes and weights in  $K$  is adjusted correspondingly. This choice writes:

```
quadrature_option qopt;
qopt.set_family (quadrature_option::gauss);
qopt.set_order (2*Xh.degree()-1);
```

while the `qopt` variable is send as an optional argument to the weighted form  $a(.,.)$  declaration. Remark that the integral would be exact for a constant weight. For a general weight, this choice also guarantee that the approximate solution  $u_h$  converges optimally with mesh refinements to the exact solution  $u$  (see [77, p. 129]). Notice also that the Gauss quadrature formula is convenient here, as quadrature nodes are internal to the elements: evaluation of  $\eta$  does not occurs at the domain boundaries, where the weight function could be singular when  $p < 2$  and where the gradient vanishes, e.g. at corners.

Example file 3.11: `eta.icc`

```
1 struct eta {
2   Float operator() (const Float& z) const {
3     check_macro(z != 0 || p > 2, "eta: division by zero (HINT: check mesh)");
4     return pow(z, (p-2)/2);
5   }
6   Float derivative (const Float& z) const {
7     check_macro(z != 0 || p > 4, "eta': division by zero (HINT: check mesh)");
8     return 0.5*(p-2)*pow(z, (p-4)/2);
9   }
10  eta (const Float& q) : p(q) {}
11  Float p;
12 };
```

The  $\eta$  function is implemented separately, in file named `eta.icc` in order to easily change its definition. The `derivative` member function is not yet used here: it is implemented for a forthcoming application (the Newton method). Notice the guards that check for division by zero and send a message related to the mesh: this will be commentated in the next paragraph. Finally, the fixed-point algorithm is initiated with  $u^{(0)}$  as the solution of the linear problem associated to  $p = 2$ , i.e. the standard Poisson problem with Dirichlet boundary conditions.

Example file 3.12: dirichlet.icc

```

1 void dirichlet (const field& lh, field& uh) {
2   const space& Xh = lh.get_space();
3   trial u (Xh); test v (Xh);
4   form a = integrate (dot(grad(u),grad(v)));
5   solver sa (a.uu());
6   uh.set_u() = sa.solve (lh.u() - a.ub()*uh.b());
7 }

```

### Running the program

Compile the program, as usual:

```
make p_laplacian_fixed_point
```

and enter the commands:

```
mkgeo_ugrid -t 50 > square.geo
geo square.geo
```

The triangular mesh has a boundary domain named boundary.

```
./p_laplacian_fixed_point square.geo P1 1.5 > square.field
```

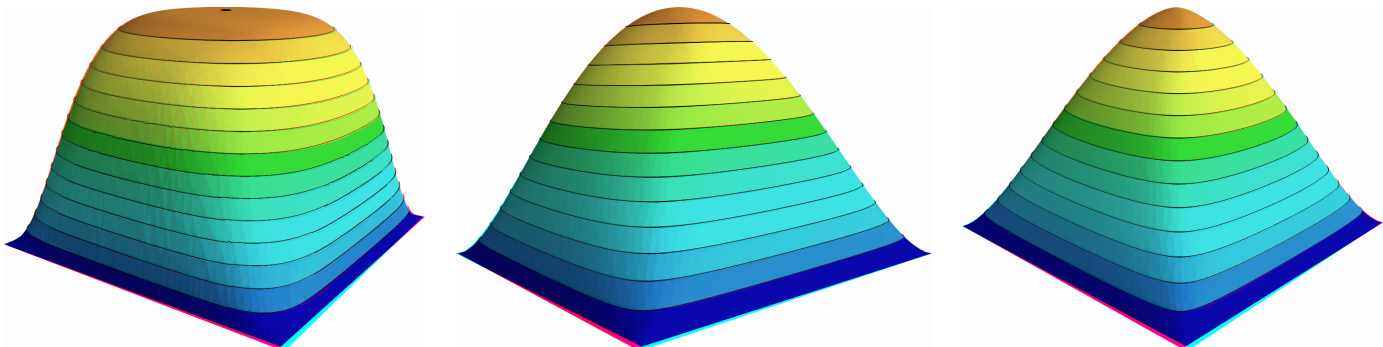


Figure 3.7: The  $p$ -Laplacian for  $d = 2$ : elevation view for  $p = 1.25$  (left),  $p = 2$  (center) and  $p = 2.5$  (right).

Run the field visualization:

```
field square.field -elevation -stereo
field square.field -cut -origin 0.5 0.5 -normal 1 1 -gnuplot
```

The first command shows an elevation view of the solution (see 3.7) while the second one shows a cut along the first bisector  $x_0 = x_1$ . Observe that the solution becomes flat at the center when  $p$  decreases. The  $p = 2$  case, corresponding to the linear case, is showed for the purpose of comparison.

There is a technical issue concerning the mesh: the computation could failed on some mesh that presents at least one triangle with two edges on the boundary:

```
mkgeo_grid -t 50 > square-bedge.geo
geo square-bedge.geo
./p_laplacian_fixed_point square-bedge.geo P1 1.5 > square-bedge.field
```

The computation stops and claims a division by zero: the three nodes of such a triangle, the three nodes are on the boundary, where  $u_h = 0$  is prescribed: thus  $\nabla u_h = 0$  uniformly inside this element. Notice that this failure occurs only for linear approximations: the computation works well on such meshes for  $P_k$  approximations with  $k \geq 2$ . While the `mkgeo_grid` generates uniform meshes that have such triangles, the `mkgeo_ugrid` calls the `gmsh` generator that automatically splits the triangles with two boundary edges. When using `bang`, you should consider the `-splitpbedge` option.

### Convergence properties of the fixed-point algorithm

The fixed-point algorithm prints also  $r_n$ , the norm of the residual term, at each iteration  $n$ , and the convergence rate  $v_n = \log_{10}(r_n/r_0)/n$ . The residual term of the non-linear variational formulation is defined by:

$$r_h^{(n)} \in V_h \quad \text{and} \quad m(r_h^{(n)}, v_h) = a(u_h^{(n)}; u_h^{(n)}, v_h) - l(v_h), \quad \forall v_h \in V_h$$

where  $m(\cdot, \cdot)$  denotes the  $L^2$  scalar product. Clearly,  $u_h^{(n)}$  is a solution if and only if  $r_h^{(n)} = 0$ .

For clarity, let us drop temporarily the  $n$  index of the current iteration. The field  $r_h \in V_h$  can be extended as a field  $r_h \in X_h$  with vanishing components on the boundary. The previous relation writes, after expansion of the bilinear forms and fields on the unknown and blocked parts (see page 14 for the notations):

$$\begin{aligned} \text{m.uu*rh.u} &= \text{a.uu*uh.u} + \text{a.ub*ub.b} - \text{lh.u} \\ \text{rh.b} &= 0 \end{aligned}$$

This relation expresses that the residual term  $r_h$  is obtained by solving a linear system involving the mass matrix.

It remains to choose a good norm for estimating this residual term. For the corresponding continuous formulation, we have:

$$r = -\text{div}(\eta(|\nabla u|^2) \nabla u) - f \in W^{-1,p}(\Omega)$$

Thus, for the continuous formulation, the residual term may be measured with the  $W^{-1,p}(\Omega)$  norm. It is defined, for all  $\varphi \in W^{-1,p}(\Omega)$ , by duality:

$$\|\varphi\|_{-1,p,\Omega} = \sup_{\substack{\varphi \in W_0^{-1,p}(\Omega) \\ v \neq 0}} \frac{\langle \varphi, v \rangle}{\|v\|_{1,p,\Omega}} = \sup_{\substack{v \in W_0^{1,p}(\Omega) \\ \|v\|_{1,p,\Omega}=1}} \langle \varphi, v \rangle$$

where  $\langle \cdot, \cdot \rangle$  denotes the duality bracketed between  $W_0^{-1,p}(\Omega)$  and  $W^{-1,p}(\Omega)$ .

By analogy, let us introduce the discrete  $W^{-1,p}(\Omega)$  norm, denoted as  $\|\cdot\|_{-1,h}$ , defined by duality for all  $\varphi_h \in V_h$  by:

$$\|\varphi_h\|_{-1,h} = \sup_{\substack{v_h \in V_h \\ \|v_h\|_{1,p,\Omega}=1}} \langle \varphi_h, v_h \rangle$$

The dual of space of the finite element space  $V_h$  is identified to  $V_h$  and the duality bracketed is the Euclidian scalar product of  $\mathbb{R}^{\dim(V_h)}$ . Then,  $\|\varphi_h\|_{-1,h}$  is the largest absolute value of components of  $\varphi_h$  considered as a vector of  $\mathbb{R}^{\dim(V_h)}$ . With the notations of the **Rheolef** library, it simply writes:

$$\text{Float r} = \text{rh.u}().\text{max\_abs}()$$

Fig 3.8.top-left shows that the residual term decreases exponentially versus  $n$ , since the slope of the plot in semi-log scale tends to be strait. Moreover, observe that the slope is independent of

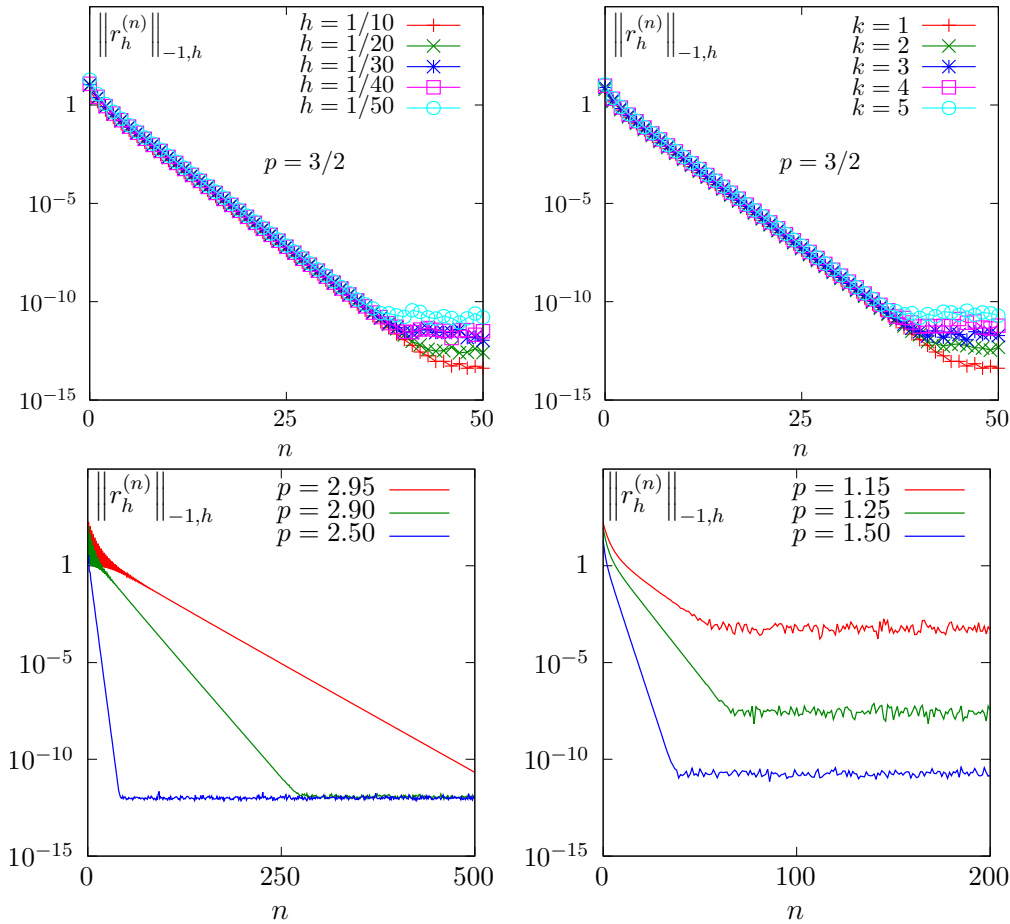


Figure 3.8: The fixed-point algorithm on the  $p$ -Laplacian for  $d = 2$ : when  $p = 3/2$ , independence of the convergence properties of the residue (top-left) with mesh refinement; (top-right) with polynomial order  $P_k$ ; when  $h = 1/50$  and  $k = 1$ , convergence (bottom-left) for  $p > 2$  and (bottom-right) for  $p < 2$ .

the mesh size  $h$ . Also, by virtue of the previous careful definition of the residual term and its corresponding norm, all the slopes falls into a master curve.

These invariance properties applies also to the polynomial approximation  $P_k$  : Fig 3.8.top-right shows that all the curves tends to collapse when  $k$  increases. Thus, the convergence properties of the algorithm are now investigated on a fixed mesh  $h = 1/50$  and for a fixed polynomial approximation  $k = 1$ .

Fig 3.8.bottom-left and 3.8.bottom-right show the convergence versus the power-law index  $p$ : observe that the convergence becomes easier when  $p$  approaches  $p = 2$ , where the problem is linear. In that case, the convergence occurs in one iteration. Nevertheless, it appears two limitations. From one hand, when  $p \rightarrow 3$  the convergence starts to slow down and  $p \geq 3$  cannot be solved by this algorithm (it will be solved later in this chapter). From other hand, when  $p \rightarrow 1$ , the convergence slows down too and numerical rounding effects limits the convergence: the machine precision cannot be reached. Let us introduce the convergence rate  $v_n = \log_{10}(r_n/r_0)/n$  it tends to a constant, denoted as  $\bar{v}$  and:  $r_n \approx r_0 \times 10^{-\bar{v}n}$ . Observe on Fig 3.9.left that  $\bar{v}$  tends to  $+\infty$  when  $p = 2$ , since the system becomes linear and the algorithm converge in one iteration. Observe also that  $\bar{v}$  tends to zero for  $p = 1$  and  $p = 3$  since the algorithm diverges. Fig 3.9.right shows the same plot in semi-log scale and shows that  $\bar{v}$  behaves as:  $\bar{v} \approx -\log_{10} |p - 2|$ . This study shows



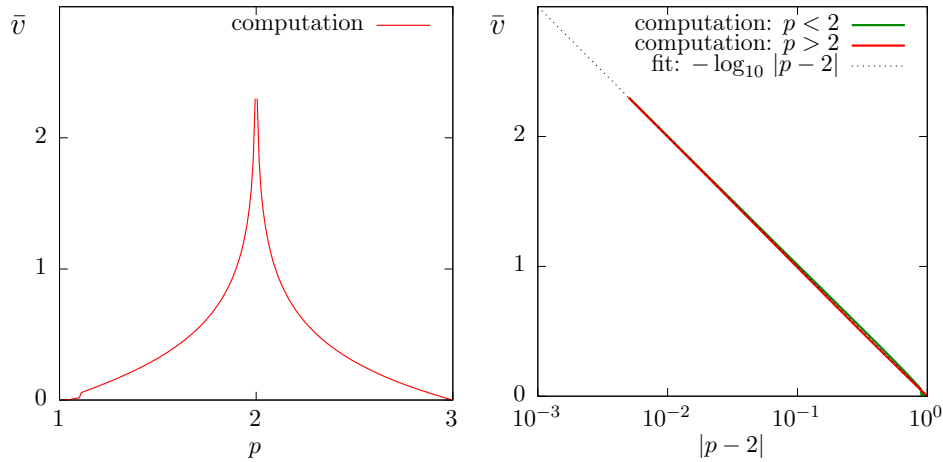


Figure 3.9: The fixed-point algorithm on the  $p$ -Laplacian for  $d = 2$ : (left) convergence rate versus  $p$ ; (right) convergence rate versus  $p$  in semi-log scale.

that the residual term of the fixed point algorithm behaves as:

$$r_n \approx r_0 |p - 2|^n$$

### Improvement by relaxation

The relaxation parameter can improve the fixed-point algorithm: for instance, for  $p = 3$  and  $\omega = 0.5$  we get a convergent sequence:

```
./p_laplacian_fixed_point square.geo P1 3 0.5 > square.field
```

Observe on Fig. 3.10 the effect on the relaxation parameter  $\omega$  upon the convergence rate  $\bar{v}$ : for  $p < 2$  it can improve it and for  $p > 2$ , it can converge when  $p > 3$ . For each  $p$ , there is clearly an optimal relaxation parameter, denoted by  $\omega_{\text{opt}}$ . A simple fit shows that (see Fig. 3.10.bottom-left):

$$\omega_{\text{opt}} = 2/p$$

Let us denote  $\bar{v}_{\text{opt}}$  the corresponding rate of convergence. Fig. 3.10.top-right shows that the convergence is dramatically improved when  $p > 2$  while the gain is less pronounced when  $p < 2$ . Conveniently replacing the extra parameter  $\omega$  on the command line by  $-$  leads to compute automatically  $\omega = \omega_{\text{opt}}$ : the fixed-point algorithm is always convergent with an optimal convergent rate, e.g.:

```
./p_laplacian_fixed_point square.geo P1 4.0 - > square.field
```

There is no way to improve more the fixed point algorithm: the next paragraph shows a different algorithm that dramatically accelerates the computation of the solution.

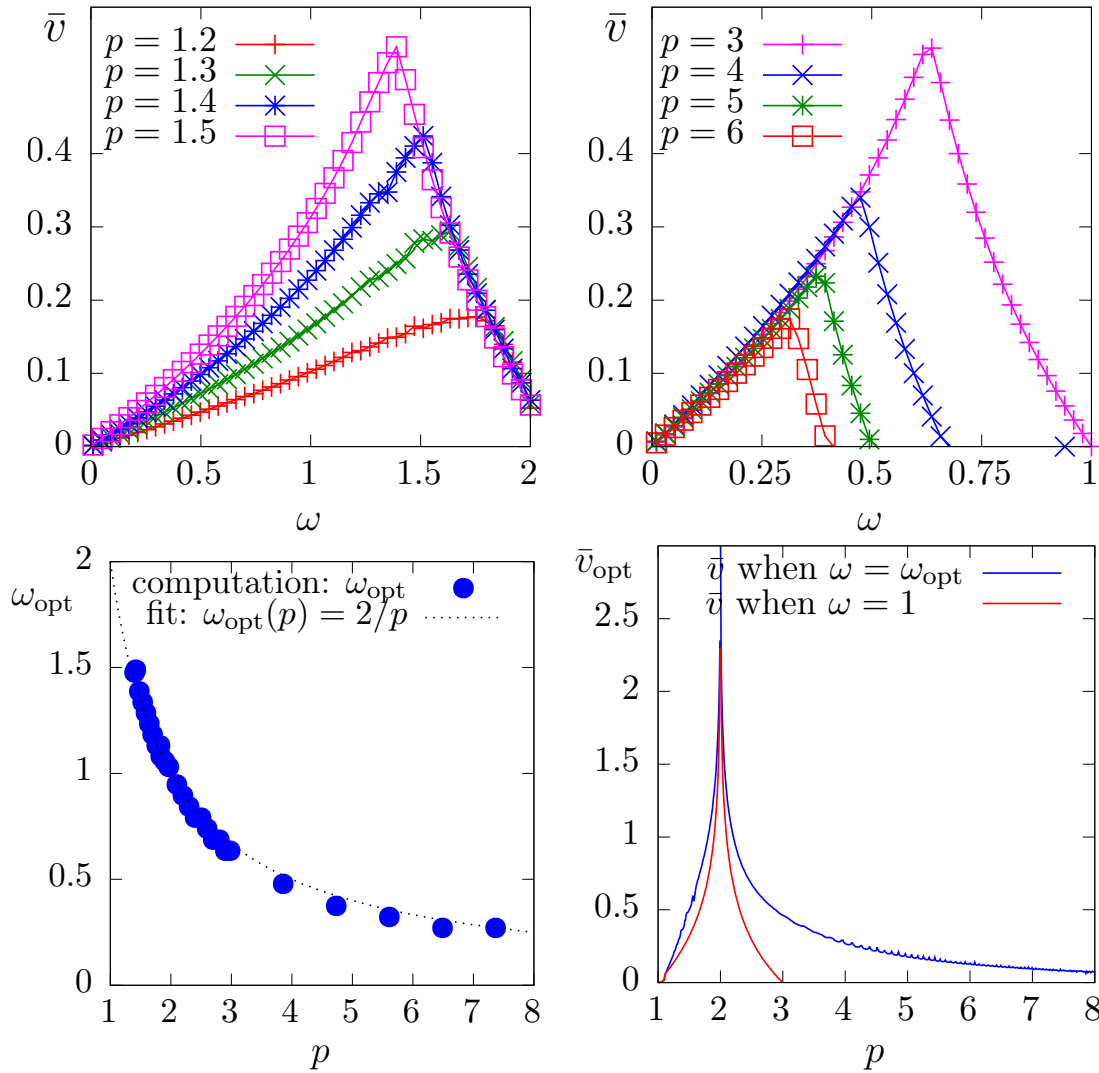


Figure 3.10: The fixed-point algorithm on the  $p$ -Laplacian for  $d = 2$ : effect of the relaxation parameter  $\omega$  (top-left) when  $p < 2$ ; (top-right) when  $p > 2$ ; (bottom-left) optimal  $\omega_{\text{opt}}$ ; (bottom-right) optimal  $\bar{v}_{\text{opt}}$ .

### 3.2.3 The Newton algorithm

#### Principle of the algorithm

An alternative to the fixed-point algorithm is to solve the nonlinear problem  $(P)$  by using the Newton algorithm. Let us consider the following operator:

$$\begin{aligned} F &: W_0^{1,p}(\Omega) \longrightarrow W^{-1,p}(\Omega) \\ u &\longmapsto F(u) = -\operatorname{div}(\eta(|\nabla u|^2) \nabla u) - f \end{aligned}$$

The  $F$  operator computes simply the residual term and the problem expresses now as: find  $u \in W_0^{1,p}(\Omega)$  such that  $F(u) = 0$ .

The Newton algorithm reduces the nonlinear problem into a sequence of linear subproblems: the sequence  $(u^{(n)})_{n \geq 0}$  is classically defined by recurrence as:

- $n = 0$ : let  $u^{(0)} \in W_0^{1,p}(\Omega)$  be known.
- $n \geq 0$ : suppose that  $u^{(n)}$  is known, find  $\delta u^{(n)}$ , defined in  $\Omega$ , such that:

$$F'(u^{(n)}) \delta u^{(n)} = -F(u^{(n)})$$

and then compute explicitly:

$$u^{(n+1)} := u^{(n)} + \delta u^{(n)}$$

The notation  $F'(u)$  stands for the Fréchet derivative of  $F$ , as an operator from  $W^{-1,p}(\Omega)$  into  $W_0^{1,p}(\Omega)$ . For any  $r \in W^{-1,p}(\Omega)$ , the linear tangent problem writes:  
find  $\delta u \in W_0^{1,p}(\Omega)$  such that:

$$F'(u) \delta u = -r$$

After the computation of the Fréchet derivative, we obtain the strong form of this problem:  
( $LT$ ): find  $\delta u$ , defined in  $\Omega$ , such that

$$\begin{aligned} -\operatorname{div}(\eta(|\nabla u|^2) \nabla(\delta u) + 2\eta'(|\nabla u|^2) \{\nabla u, \nabla(\delta u)\} \nabla u) &= -r \quad \text{in } \Omega \\ \delta u &= 0 \quad \text{on } \partial\Omega \end{aligned}$$

where

$$\eta'(z) = \frac{1}{2}(p-2)z^{\frac{p-4}{2}}, \quad \forall z > 0$$

This is a Poisson-like problem with homogeneous Dirichlet boundary conditions and a non-constant tensorial coefficient. The variational form of the linear tangent problem writes:  
( $VLT$ ): find  $\delta u \in W_0^{1,p}(\Omega)$  such that

$$a_1(u; \delta u, \delta v) = l_1(v), \quad \forall \delta v \in W_0^{1,p}(\Omega)$$

where the  $a_1(\cdot; \cdot, \cdot)$  is defined for any  $u, \delta u, \delta v \in W_0^{1,p}(\Omega)$  by:

$$\begin{aligned} a_1(u; \delta u, \delta v) &= \int_{\Omega} (\eta(|\nabla u|^2) \nabla(\delta u) \cdot \nabla(\delta v) + 2\eta'(|\nabla u|^2) \{\nabla u, \nabla(\delta u)\} \{\nabla u, \nabla(\delta v)\}) \, dx \\ l_1(v) &= - \int_{\Omega} r v \, dx \end{aligned}$$

For any  $\xi \in \mathbb{R}^d$  let us denote by  $\nu(\xi)$  the following  $d \times d$  matrix:

$$\nu(\xi) = \eta(|\xi|^2) I + 2\eta'(|\xi|^2) \xi \otimes \xi$$

where  $I$  stands for the  $d$ -order identity matrix. Then the  $a_1$  expresses in a more compact form:

$$a_1(u; \delta u, \delta v) = \int_{\Omega} (\nu(\nabla u) \nabla(\delta u)) \cdot \nabla(\delta v) dx$$

Clearly  $a_1$  is linear and symmetric with respect to the two last variables.

Example file 3.13: p\_laplacian\_newton.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "p_laplacian.h"
5 int main(int argc, char**argv) {
6     environment rheolef (argc, argv);
7     geo omega (argv[1]);
8     Float eps = std::numeric_limits<Float>::epsilon();
9     string approx = (argc > 2) ? argv[2] : "P1";
10    Float p = (argc > 3) ? atof(argv[3]) : 1.5;
11    Float tol = (argc > 4) ? atof(argv[4]) : 1e5*eps;
12    size_t max_iter = (argc > 5) ? atoi(argv[5]) : 500;
13    derr << "# P-Laplacian problem by Newton:" << endl
14         << "# geo = " << omega.name() << endl
15         << "# approx = " << approx << endl
16         << "# p = " << p << endl
17         << "# tol = " << tol << endl
18         << "# max_iter = " << max_iter << endl;
19    p_laplacian F (p, omega, approx);
20    field uh = F.initial ();
21    int status = newton (F, uh, tol, max_iter, &derr);
22    dout << setprecision(numeric_limits<Float>::digits10)
23         << catchmark("p") << p << endl
24         << catchmark("u") << uh;
25    return status;
26 }

```

Example file 3.14: p\_laplacian.h

```

1 class p_laplacian {
2 public:
3     typedef field value_type;
4     typedef Float float_type;
5     p_laplacian (Float p, const geo& omega, string approx);
6     field initial() const;
7     field residue (const field& uh) const;
8     void update_derivative (const field& uh) const;
9     field derivative_solve (const field& mrh) const;
10    field derivative_trans_mult (const field& mrh) const;
11    Float space_norm (const field& uh) const;
12    Float dual_space_norm (const field& mrh) const;
13    Float p;
14    space Xh;
15    field lh;
16    form m;
17    solver sm;
18    quadrature_option qopt;
19    mutable form a1;
20    mutable solver sa1;
21 };
22 #include "p_laplacian1.icc"
23 #include "p_laplacian2.icc"

```

## Comments

The Newton algorithm is implemented in a generic way, for any  $F$  function, by the `newton` function from the `Rheolef` libraries. The reference manual for the `newton` generic function is available online:

`man newton`

The function  $F$  and its derivative  $F'$  are provided by a template class argument. Here, the `p_laplacian` class describes our  $F$  function, i.e. our problem to solve: its interface is defined in the file `'p_laplacian.h'` and its implementation in `'p_laplacian1.icc'` and `'p_laplacian2.icc'`. The introduction of the class `p_laplacian` will allow an easy exploration of some variants of the Newton algorithm for this problem, as we will see in the next section.

Example file 3.15: `p_laplacian1.icc`

```

1 #include "eta.icc"
2 #include "nu.icc"
3 #include "dirichlet.icc"
4 p_laplacian::p_laplacian (Float p1, const geo& omega, string approx)
5 : p(p1), Xh(), lh(), m(), sm(), qopt(), a1(), sa1() {
6   Xh = space (omega, approx);
7   Xh.block ("boundary");
8   qopt.set_family(quadrature_option::gauss);
9   qopt.set_order(2*Xh.degree()-1);
10  trial u (Xh); test v (Xh);
11  lh = integrate (v);
12  m = integrate (u*v);
13  sm = solver (m.uu());
14 }
15 field p_laplacian::initial() const {
16   field uh (Xh, 0);
17   dirichlet (lh, uh);
18   return uh;
19 }
20 field p_laplacian::residue (const field& uh) const {
21   trial u (Xh); test v (Xh);
22   form a = integrate (compose(eta(p), norm2(grad(uh)))*dot(grad(u), grad(v)),
23                       qopt);
24   field mrh = a*uh - lh;
25   mrh.set_b() = 0;
26   return mrh;
27 }
28 void p_laplacian::update_derivative (const field& uh) const {
29   size_t d = Xh.get_geo().dimension();
30   trial u (Xh); test v (Xh);
31   a1 = integrate (dot(compose(nu<eta>(eta(p), d), grad(uh))*grad(u), grad(v)),
32                  qopt);
33   sa1 = ldlt (a1.uu());
34 }
35 field p_laplacian::derivative_solve (const field& rh) const {
36   field delta_uh (Xh, 0);
37   delta_uh.set_u() = sa1.solve(rh.u());
38   return delta_uh;
39 }

```

The residual term  $F(u_h)$  is computed by the member function `residual` while the resolution of  $F'(u_h)\delta u_h = Mr_h$  is performed by the function `derivative_solve`. The derivative  $F'(u_h)$  is computed separately by the function `update_derivative`:

```

a1 = integrate (dot (compose (nu<eta>(eta(p), d), grad(uh))*grad(u), grad(v)),
                qopt);

```

Notice that the  $a_1(u; \cdot, \cdot)$  bilinear form is a tensorial weighted form, where  $\nu = \nu(\nabla u)$  is the weight tensor. The tensorial weight  $\nu$  is inserted as  $(\nu \nabla u) \cdot \nabla v$  in the variational expression for the `integrate` function. As the tensor  $\nu$  is symmetric, the bilinear form  $a_1(\cdot, \cdot)$  is also symmetric. As the weight is non-polynomial for general  $\eta$  function and a quadrature formula is used:

$$a_1(u_0; u, v) = \sum_{K \in \mathcal{T}_h} \sum_{q=0}^{n_K-1} (\nu(\nabla u_0(x_{K,q})) \nabla u(x_{K,q}) \cdot \nabla v(x_{K,q})) \omega_{K,q} \quad (3.6)$$

By using exactly the same quadrature for computing both  $a_1(\cdot, \cdot)$  and  $a(\cdot, \cdot)$  in (3.6), then we have that  $F'$  is always the derivative of  $F$  at the discrete level: while, in general, the derivation and the discretization of problems does not commute, it is the case when using the same quadrature

formulae on both problems. This is an important aspect of the Newton method at discrete level, for conservating the optimal convergence rate of the residual terms versus  $n$ .

The linear system involving the derivative  $F'(u_h)$  is solved by the `p_laplacian` member function `derivative_solve`. Finally, applying the generic Newton method requires a stopping criteria on the residual term: this is the aim of the member function `dual_space_norm`. The three last member functions are not used by the Newton algorithm, but by its extension, the damped Newton method, that will be presented later.

Example file 3.16: `p_laplacian2.icc`

```

1 field p_laplacian::derivative_trans_mult (const field& mrh) const {
2   field rh (Xh, 0);
3   rh.set_u() = sm.solve(mrh.u());
4   field mgh = a1*rh;
5   mgh.set_b() = 0;
6   return mgh;
7 }
8 Float p_laplacian::space_norm (const field& uh) const {
9   return sqrt (m(uh,uh));
10 }
11 Float p_laplacian::dual_space_norm (const field& mrh) const {
12   field rh (Xh, 0);
13   rh.set_u() = sm.solve (mrh.u());
14   return sqrt (dual(mrh, rh));
15 }

```

The  $\nu$  function is implemented for a generic  $\eta$  function, as a class-function that accept as template argument another class-function.

Example file 3.17: `nu.icc`

```

1 template<class Function>
2 struct nu {
3   tensor operator() (const point& grad_u) const {
4     Float x2 = norm2 (grad_u);
5     Float a = f(x2);
6     Float b = 2*f.derivative(x2);
7     tensor value;
8     for (size_t i = 0; i < d; i++) {
9       value(i,i) = a + b*grad_u[i]*grad_u[i];
10      for (size_t j = 0; j < i; j++)
11        value(j,i) = value(i,j) = b*grad_u[i]*grad_u[j];
12    }
13    return value;
14  }
15  nu (const Function& f1, size_t d1) : f(f1), d(d1) {}
16  Function f;
17  size_t d;
18 };

```

### Running the program

Enter:

```

make p_laplacian_newton
mkgeo_ugrid -t 50 > square.geo
./p_laplacian_newton square.geo P1 3 > square.field
field square.field -elevation -stereo

```

The program prints at each iteration  $n$ , the residual term  $r_n$  in discrete  $L^2(\Omega)$  norm. Convergence occurs in less than ten iterations: it dramatically improves the previous algorithm (see Fig. 3.11). Observe that the slope is no more constant in semi-log scale: the convergence rate accelerates and the slope tends to be vertical, the so-called super-linear convergence. This is the major advantage of the Newton method. Figs. 3.12.top-left and. 3.12.top-bottom shows that the algorithm converge

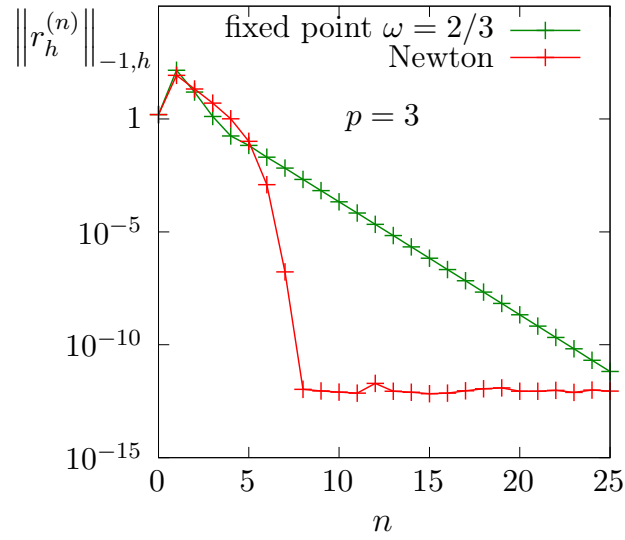


Figure 3.11: The Newton algorithm on the  $p$ -laplacian for  $d = 2$ : comparison with the fixed-point algorithm.

when  $p \geq 3$  and that the convergence properties are independent of the mesh size  $h$  and the polynomial order  $k$ . There are still two limitations of the method. From one hand, the Newton algorithm is no more independent of  $h$  and  $k$  when  $p \leq 3/2$  and tends to diverge in that case when  $h$  tends to zero (see Fig. 3.12.bottom-left). From other hand, when  $p$  becomes large (see Fig. 3.12.bottom-right), an overshoot in the convergence tends to increase and destroys the convergence, due to rounding problems. In order to circumvent these limitations, another strategy is considered in the next section: the damped Newton algorithm.

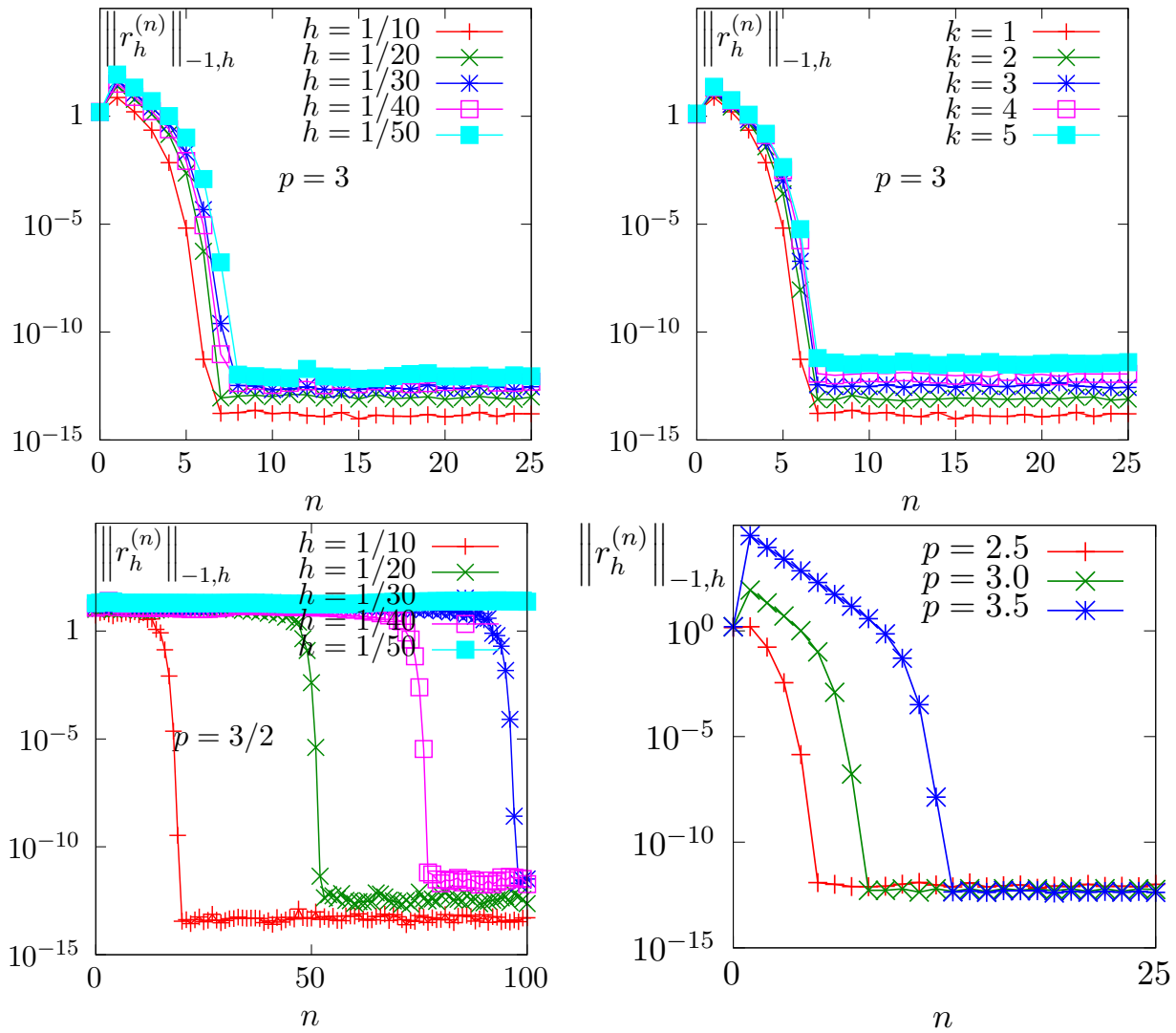


Figure 3.12: The Newton algorithm on the  $p$ -laplacian for  $d = 2$ : (top-left) comparison with the fixed-point algorithm; when  $p = 3$ , independence of the convergence properties of the residue (top-left) with mesh refinement; (top-right) with polynomial order  $P_k$ ; (bottom-left) mesh-dependence convergence when  $p < 2$ ; (bottom-right) overshoot when  $p > 2$ .



### 3.2.4 The damped Newton algorithm

#### Principe of the algorithm

The Newton algorithm diverges when the initial  $u^{(0)}$  is too far from a solution, e.g. when  $p$  is not at the vicinity of 2. Our aim is to modify the Newton algorithm and to obtain a *globally convergent algorithm*, i.e to converge to a solution for any initial  $u^{(0)}$ . By this way, the algorithm should converge for any value of  $p \in ]1, +\infty[$ . The basic idea is to decrease the step length while maintaining the direction of the original Newton algorithm:

$$u^{(n+1)} := u^{(n)} + \lambda_n \delta u^{(n)}$$

where  $\lambda^{(n)} \in ]0, 1]$  and  $\delta u^{(n)}$  is the direction from the Newton algorithm, given by:

$$F' \left( u^{(n)} \right) \delta u^{(n)} = -F \left( u^{(n)} \right)$$

Let  $V$  a Banach space and let  $T : V \rightarrow \mathbb{R}$  defined for any  $v \in V$  by:

$$T(v) = \frac{1}{2} \|C^{-1}F(v)\|_V^2,$$

where  $C$  is some non-singular operator, easy to invert, used as a non-linear preconditioner. The simplest case, without preconditioner, is  $C = I$ . The  $T$  function furnishes a measure of the residual term in  $L^2$  norm. The convergence is global when for any initial  $u^{(0)}$ , we have for any  $n \geq 0$ :

$$T \left( u^{(n+1)} \right) \leq T \left( u^{(n)} \right) + \alpha \left\langle T' \left( u^{(n)} \right), u^{(n+1)} - u^{(n)} \right\rangle_{V',V} \quad (3.7)$$

where  $\langle \cdot, \cdot \rangle_{V',V}$  is the duality product between  $V$  and its dual  $V'$ , and  $\alpha \in ]0, 1[$  is a small parameter. Notice that

$$T'(u) = \{C^{-1}F'(u)\}^* C^{-1}F(u)$$

where the superscript  $*$  denotes the adjoint operator, i.e. the transpose matrix the in finite dimensional case. In practice we consider  $\alpha = 10^{-4}$  and we also use a minimal step length  $\lambda_{\min} = 1/10$  in order to avoid too small steps. Let us consider a fixed step  $n \geq 0$ : for convenience the  $n$  superscript is dropped in  $u^{(n)}$  and  $\delta u^{(n)}$ . Let  $g : \mathbb{R} \rightarrow \mathbb{R}$  defined for any  $\lambda \in \mathbb{R}$  by:

$$g(\lambda) = T(u + \lambda \delta u)$$

Then :

$$\begin{aligned} g'(\lambda) &= \langle T'(u + \lambda \delta u), \delta u \rangle_{V',V} \\ &= \langle C^{-1}F(u + \lambda \delta u), F'(u + \lambda \delta u)C^{-1}\delta u \rangle_{V',V} \end{aligned}$$

where the superscript  $*$  denotes the adjoint operator, i.e. the transpose matrix the in finite dimensional case. The practical algorithm for obtaining  $\lambda$  was introduced first in [51] and is also presented in [76, p. 385]. The step length  $\lambda$  that satisfy (3.7) is computed by using a finite sequence  $\lambda_k, k = 0, 1 \dots$  with a second order recurrence:

- $k = 0$  : initialization  $\lambda_0 = 1$ . If (3.7) is satisfied with  $u + \lambda_0 d$  then let  $\lambda := \lambda_0$  and the sequence stop here.
- $k = 1$  : first order recursion. The quantities  $g(0) = f(u)$  et  $g'(0) = \langle f'(u), d \rangle$  are already computed at initialization. Also, we already have computed  $g(1) = f(u + d)$  when verifying whether (3.7) was satisfied. Thus, we consider the following approximation of  $g(\lambda)$  by a second order polynomial:

$$\tilde{g}_1(\lambda) = \{g(1) - g(0) - g'(0)\}\lambda^2 + g'(0)\lambda + g(0)$$

After a short computation, we find that the minimum of this polynomial is:

$$\tilde{\lambda}_1 = \frac{-g'(0)}{2\{g(1) - g(0) - g'(0)\}}$$

Since the initialization at  $k = 0$  does not satisfy (3.7), it is possible to show that, when  $\alpha$  is small enough, we have  $\tilde{\lambda}_1 \leq 1/2$  and  $\tilde{\lambda}_1 \approx 1/2$ . Let  $\lambda_1 := \max(\lambda_{\min}, \tilde{\lambda}_1)$ . If (3.7) is satisfied with  $u + \lambda_1 d$  then let  $\lambda := \lambda_1$  and the sequence stop here.

- $k \geq 2$  : second order recurrence. The quantities  $g(0) = f(u)$  et  $g'(0) = f'(u)$ ,  $d$  are available, together with  $\lambda_{k-1}$ ,  $g(\lambda_{k-1})$ ,  $\lambda_{k-2}$  and  $g(\lambda_{k-2})$ . Then,  $g(\lambda)$  is approximated by the following third order polynomial:

$$\tilde{g}_k(\lambda) = a\lambda^3 + b\lambda^2 + g'(0)\lambda + g(0)$$

where  $a$  et  $b$  are expressed by:

$$\begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{\lambda_{k-1} - \lambda_{k-2}} \begin{pmatrix} \frac{1}{\lambda_{k-1}^2} & -\frac{1}{\lambda_{k-2}^2} \\ -\frac{\lambda_{k-2}}{\lambda_{k-1}^2} & \frac{\lambda_{k-1}}{\lambda_{k-2}^2} \end{pmatrix} \begin{pmatrix} g(\lambda_{k-1}) - g'(0)\lambda_{k-1} - g(0) \\ g(\lambda_{k-2}) - g'(0)\lambda_{k-2} - g(0) \end{pmatrix}$$

The minimum of  $\tilde{g}_k(\lambda)$  is

$$\tilde{\lambda}_k = \frac{-b + \sqrt{b^2 - 3ag'(0)}}{3a}$$

Let  $\lambda_k = \min(1/2 \lambda_k, \max(\tilde{\lambda}_k/10, \tilde{\lambda}_{k+1}))$  in order for  $\lambda_k$  to be at the same order of magnitude as  $\lambda_{k-1}$ . If (3.7) is satisfied with  $u + \lambda_k d$  then let  $\lambda := \lambda_k$  and the sequence stop here.

The sequence  $(\lambda_k)_{k \geq 0}$  is strictly decreasing: when the stopping criteria is not satisfied until  $\lambda_k$  reaches the machine precision  $\varepsilon_{\text{mach}}$  then the algorithm stops with an error.

Example file 3.18: p\_laplacian\_damped\_newton.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "p_laplacian.h"
5 int main(int argc, char**argv) {
6     environment rheolef (argc,argv);
7     geo omega (argv[1]);
8     Float eps = numeric_limits<Float>::epsilon();
9     string approx = (argc > 2) ? argv[2] : "P1";
10    Float p = (argc > 3) ? atof(argv[3]) : 1.5;
11    Float tol = (argc > 4) ? atof(argv[4]) : eps;
12    size_t max_iter = (argc > 5) ? atoi(argv[5]) : 500;
13    derr << "# P-Laplacian problem by damped Newton:" << endl
14         << "# geo = " << omega.name() << endl
15         << "# approx = " << approx << endl
16         << "# p = " << p << endl;
17    p_laplacian F (p, omega, approx);
18    field uh = F.initial ();
19    int status = damped_newton (F, uh, tol, max_iter, &derr);
20    dout << catchmark("p") << p << endl
21         << catchmark("u") << uh;
22    return status;
23 }

```

### Comments

The `damped_newton` function implements the damped Newton algorithm for a generic  $T(u)$  function, i.e. a generic nonlinear preconditioner. This algorithm uses a backtrack strategy implemented

in the file ‘`newton-backtrack.h`’ of the **Rheolef** library. The simplest choice of the identity preconditioner  $C = I$  i.e.  $T(u) = \|F(u)\|_{V'}^2/2$  is showed in file `damped-newton.h`. The gradient at  $\lambda = 0$  is

$$T'(u) = F'(u)^* F(u)$$

and the slope at  $\lambda = 0$  is:

$$\begin{aligned} g'(0) &= \langle T'(u), \delta u \rangle_{V',V} \\ &= \langle F(u), F'(u)\delta u \rangle_{V',V'} \\ &= -\|F(u)\|_{V'}^2 \end{aligned}$$

The ‘`p_laplacian_damped_newton.cc`’ is the application program to the  $p$ -Laplacian problem together with the  $\|\cdot\|_{L^2(\Omega)}$  discrete norm for the function  $T$ .

### Running the program

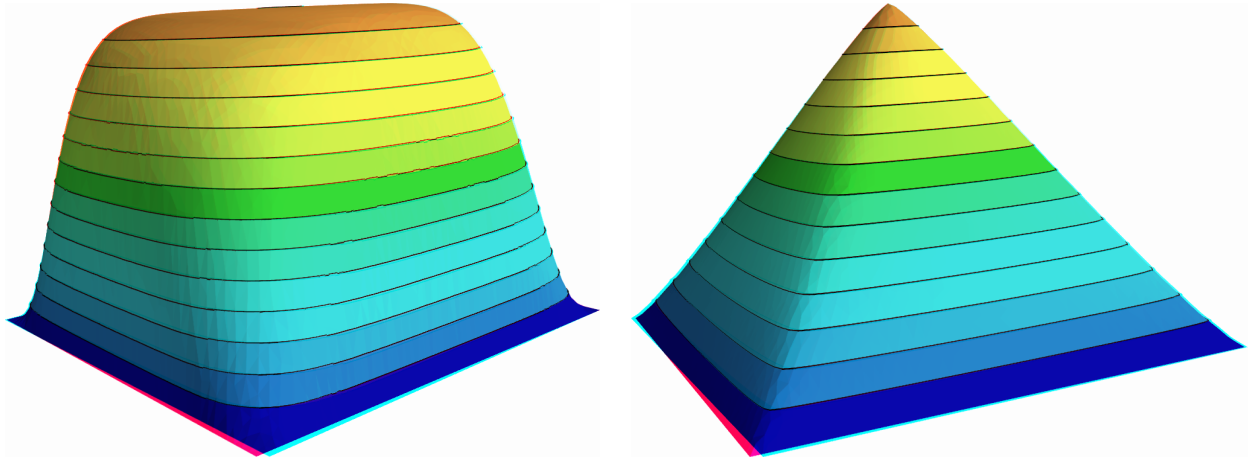


Figure 3.13: The  $p$ -Laplacian for  $d = 2$ : elevation view for  $p = 1.15$  (left) and  $p = 7$  (right).

As usual, enter:

```
make p_laplacian_damped_newton
mkgeo_ugrid -t 50 > square.geo
./p_laplacian_damped_newton square.geo P1 1.15 | field -stereo -elevation -
./p_laplacian_damped_newton square.geo P1 7 | field -stereo -elevation -
```

See Fig. 3.13 for the elevation view of the solution. The algorithm is now quite robust: the convergence occurs for quite large range of  $p > 1$  values and extends the range previously presented on Fig. 3.7. The only limitation is now due to machine roundoff on some architectures.

Figs. 3.14.top shows that the convergence properties seems to slightly depend on the mesh refinement. Nevertheless, there are quite good and support both mesh refinement and high order polynomial degree. When  $p$  is far from  $p = 2$ , i.e. either close to one or large, Figs. 3.14.bottom shows that the convergence becomes slower and that the first linear regime, corresponding to the line search, becomes longer. This first regime finishes by a brutal super-linear regime, where the residual terms fall in few iterations to the machine precision.

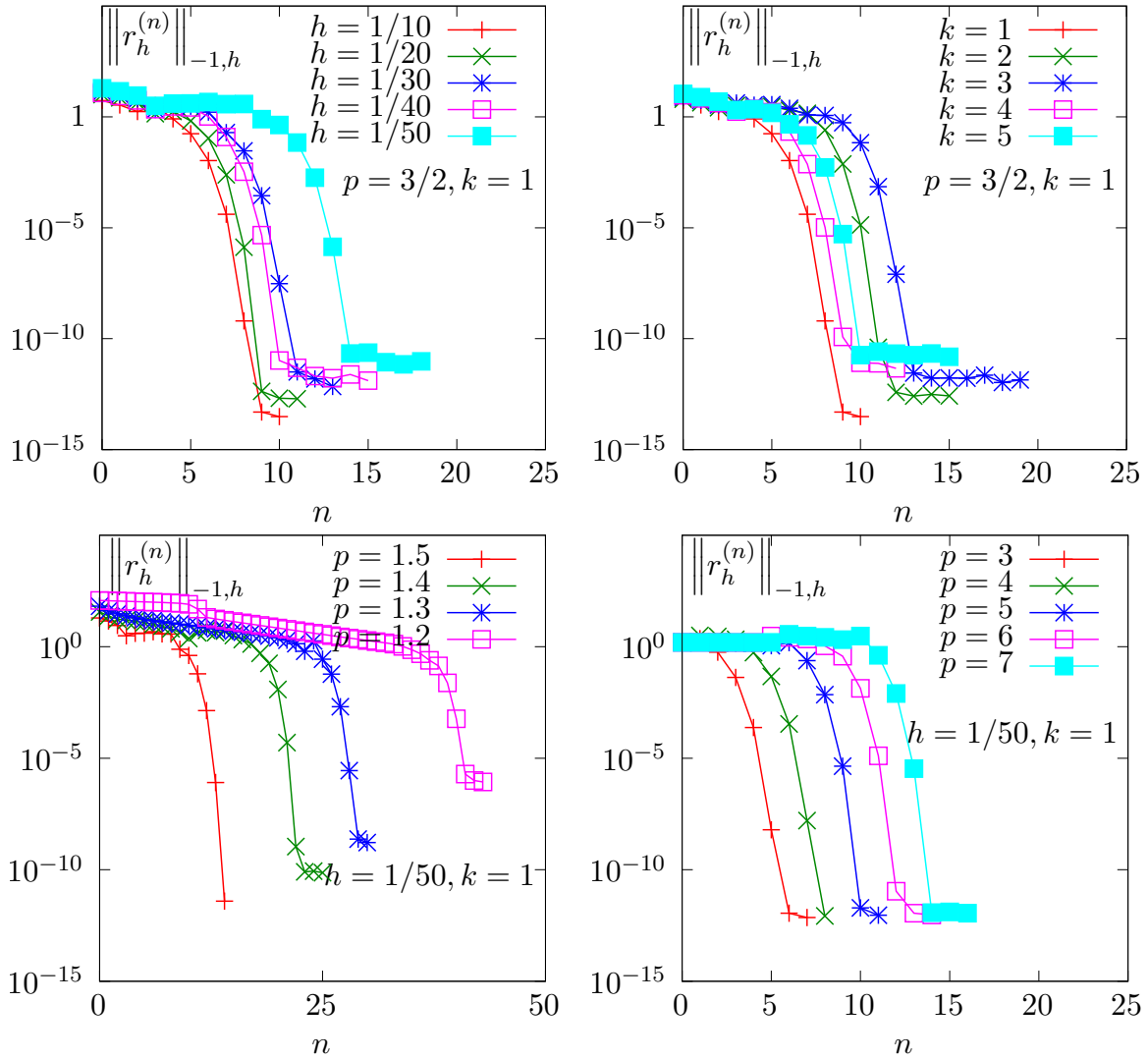
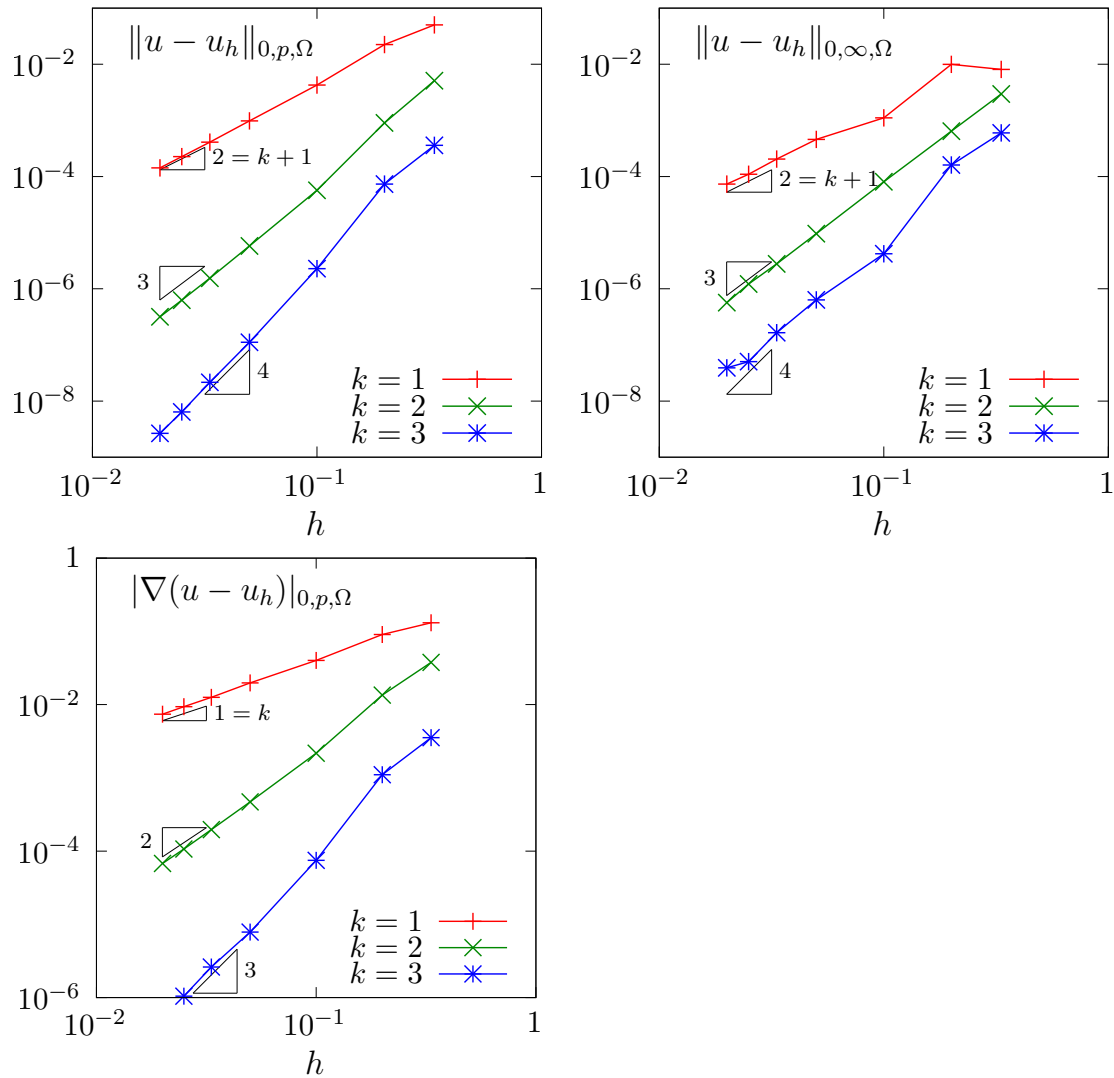


Figure 3.14: The damped Newton algorithm on the  $p$ -Laplacian for  $d = 2$ : when  $p = 1.5$  and  $h = 1/50$ , convergence properties of the residue (top-left) with mesh refinement; (top-right) with polynomial order  $P_k$ ; (bottom-left) convergence when  $p < 2$ ; (bottom-right) when  $p > 2$ .

### 3.2.5 Error analysis

While there is no simple explicit expression for the exact solution in the square  $\Omega = ]0, 1[^2$ , there is one when considering  $\Omega$  as the unit circle:

$$u(x) = \frac{(p-1) 2^{-\frac{1}{p-1}}}{p} \left( 1 - (x_0^2 + x_1^2)^{\frac{p}{p-1}} \right)$$

Figure 3.15: The  $p$ -Laplacian for  $d = 2$ : error analysis.

Example file 3.19: p\_laplacian\_circle.icc

```

1 struct u_exact {
2   Float operator() (const point& x) const {
3     return (1 - pow(norm2(x), p/(2*p-2)))/(p/(p-1))*pow(2.,1/(p-1)));
4   }
5   u_exact (Float q) : p(q) {}
6   protected: Float p;
7 };
8 struct grad_u {
9   point operator() (const point& x) const {
10    return - (pow(norm2(x), p/(2*p-2) - 1)/pow(2.,1/(p-1)))*x;
11  }
12  grad_u (Float q) : p(q) {}
13  protected: Float p;
14 };

```

Example file 3.20: p\_laplacian\_error.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "p_laplacian_circle.icc"
5 int main(int argc, char**argv) {
6     environment rheolef (argc,argv);
7     Float tol = (argc > 1) ? atof(argv[1]) : 1e-15;
8     Float p;
9     field uh;
10    din >> catchmark("p") >> p
11         >> catchmark("u") >> uh;
12    const geo& omega = uh.get_geo();
13    const space& Xh = uh.get_space();
14    field pi_h_u = interpolate (Xh, u_exact(p));
15    field eh = pi_h_u - uh;
16    quadrature_option qopt;
17    qopt.set_family(quadrature_option::gauss);
18    qopt.set_order(2*Xh.degree());
19    Float err_lp = pow(integrate (omega,
20        pow(fabs(uh - u_exact(p)), p), qopt), 1./p);
21    Float err_wlp = pow(integrate (omega,
22        pow(norm(grad(uh) - grad_u(p)), p), qopt), 1./p);
23    Float err_linf = eh.max_abs();
24    dout << "err_linf = " << err_linf << endl
25         << "err_lp = " << err_lp << endl
26         << "err_wlp = " << err_wlp << endl;
27    return (err_linf < tol) ? 0 : 1;
28 }

```

Notice, in the file ‘p\_laplacian\_error.cc’, the usage of the `integrate` function, together with a quadrature formula specification, for computing the errors in  $L^p$  norm and  $W^{1,p}$  semi-norm. Notice also the flexibility of expressions, mixing together fields as `uh` and functors, as `u_exact`. The whole expression is evaluated by the `integrate` function at quadrature points inside each element of the mesh.

By this way, the error analysis investigation becomes easy:

```

make p_laplacian_error
mkgeo_ball -t 10 -order 2 > circle-10-P2.geo
./p_laplacian_damped_newton circle-10-P2.geo P2 1.5 | ./p_laplacian_error

```

We can vary both the mesh size and the polynomial order and the error plots are showed on Fig. 3.15 for both the  $L^2$ ,  $L^\infty$  norms and the  $W^{1,p}$  semi-norm. Observe the optimal error behavior: the slopes in the log-log scale are the same as those obtained by a direct Lagrange interpolation of the exact solution.

### 3.3 [New] Continuation and bifurcation methods

This chapter is an introduction to continuation and bifurcation methods with Rheolef. We consider a model nonlinear problem that depends upon a parameter. This problem is inspired from application to combustion. Solutions exists for a limited range of this parameter and there is a limit point: beyond this limit point there is no more solution. Our first aim is to compute the branch of solutions until this limit point, thanks to the continuation algorithm. Moreover, the limit point is a turning point for the branch of solutions: there exists a second branch of solutions, continuing the first one. Our second aim is to compute the second branch of solutions after this limit point with the Keller continuation algorithm. For simplicity in this presentation, the discretization is in one dimension: the extension to high order space dimension is immediate.

### 3.3.1 Problem statement and the Newton method

Let us consider the following model problem (see [71, p. 59] or [28, p. 2]), defined for all  $\lambda \in \mathbb{R}$ :

(P): find  $u$ , defined in  $\Omega$  such that

$$\begin{aligned} -\Delta u + \lambda \exp(u) &= 0 \text{ in } \Omega \\ u &= 0 \text{ on } \partial\Omega \end{aligned}$$

In order to apply a Newton method to the whole problem, let us introduce:

$$F(\lambda, u) = -\Delta u + \lambda \exp(u)$$

Then, the Gâteaux derivative at any  $(\lambda, u) \in \mathbb{R} \times H_0^1(\Omega)$  is given by:

$$\frac{\partial F}{\partial u}(\lambda, u).(v) = -\Delta v + \lambda \exp(u)v, \forall v \in H_0^1(\Omega)$$

Example file 3.21: combustion.h

```

1 struct combustion {
2     typedef Float float_type;
3     typedef field value_type;
4     combustion(const geo& omega=geo(), string approx="");
5     void reset(const geo& omega, string approx);
6     field initial (std::string restart="");
7     idiststream& get (idiststream& is, field& uh);
8     odiststream& put (odiststream& os, const field& uh) const;
9     string parameter_name() const { return "lambda"; }
10    float_type parameter() const { return lambda; }
11    void set_parameter(float_type lambda1) { lambda = lambda1; }
12    bool stop (const field& xh) const { return xh.max_abs() > 10; }
13    field residue (const field& uh) const;
14    csr<float_type> derivative (const field& uh) const;
15    field derivative_versus_parameter (const field& uh) const;
16    solver::determinant_type update_derivative (const field& uh) const;
17    field derivative_solve (const field& mrh) const;
18    field derivative_trans_mult (const field& mrh) const;
19    field massify (const field& uh) const { return m*uh; }
20    field unmassify (const field& uh) const;
21    float_type space_dot (const field& xh, const field& yh) const;
22    float_type dual_space_dot (const field& mrh, const field& msh) const;
23 protected:
24    float_type lambda;
25    space Xh;
26    form m;
27    solver sm;
28    mutable csr<float_type> a1;
29    mutable solver sa1;
30    mutable branch event;
31 };
32 #include "combustion1.icc"
33 #include "combustion2.icc"

```

Example file 3.22: combustion1.icc

```
1 combustion::combustion (const geo& omega, string approx)
2 : lambda(0), Xh(), m(), sm(), a1(), sa1(), event("lambda","u") {
3   if (approx != "") reset (omega, approx);
4 }
5 void combustion::reset (const geo& omega, string approx) {
6   Xh = space (omega, approx);
7   Xh.block ("boundary");
8   m = form (Xh, Xh, "mass");
9   sm = solver (m.uu());
10 }
11 field combustion::initial (std::string restart) {
12   if (restart == "") return field (Xh, 0);
13   idiststream in (restart);
14   field xh0;
15   get (in, xh0);
16   derr << "# restart from lambda=" << lambda << endl;
17   return xh0;
18 }
19 odiststream& combustion::put (odiststream& os, const field& uh) const {
20   return os << event(lambda,uh);
21 }
22 idiststream& combustion::get (idiststream& is, field& uh) {
23   is >> event(lambda,uh);
24   if (!is) return is;
25   if (Xh.stamp() == "") reset (uh.get_geo(), uh.get_approx());
26   if (uh.b().dis_size() == 0) {
27     // re-allocate the field with right blocked/unblocked sizes
28     field tmp = field(Xh, 0);
29     std::copy (uh.begin_dof(), uh.end_dof(), tmp.begin_dof());
30     uh = tmp;
31   }
32   return is;
33 }
```



Example file 3.23: combustion2.icc

```

1 field combustion::residue (const field& uh) const {
2   test v(Xh);
3   field mrh = integrate(dot(grad(uh),grad(v)) - lambda*exp(uh)*v);
4   mrh.set_b() = 0;
5   return mrh;
6 }
7 csr<Float> combustion::derivative (const field& uh) const {
8   trial du(Xh); test v(Xh);
9   form df_du = integrate(dot(grad(du),grad(v)) - lambda*exp(uh)*du*v);
10  return df_du.uu();
11 }
12 solver::determinant_type
13 combustion::update_derivative (const field& uh) const {
14   a1 = derivative (uh);
15   solver_option sopt;
16   sopt.compute_determinant = true;
17   sal = ldlt (a1, sopt);
18   return sal.det();
19 }
20 field combustion::derivative_versus_parameter (const field& uh) const {
21   test v(Xh);
22   return - integrate(exp(uh)*v);
23 }
24 field combustion::derivative_solve (const field& rh) const {
25   field delta_uh (Xh,0);
26   delta_uh.set_u() = sal.solve(rh.u());
27   return delta_uh;
28 }
29 field combustion::derivative_trans_mult (const field& mrh) const {
30   field rh = unmassify(mrh);
31   field mgh(Xh,0);
32   mgh.set_u() = a1*rh.u();
33   return mgh;
34 }
35 field combustion::unmassify (const field& mrh) const {
36   field rh (Xh, 0);
37   rh.set_u() = sm.solve(mrh.u());
38   return rh;
39 }
40 Float combustion::space_dot (const field& xh, const field& yh) const {
41   return m(xh,yh); }
42 Float combustion::dual_space_dot (const field& mrh, const field& msh) const {
43   return dual(unmassify(mrh), msh); }

```

Example file 3.24: combustion\_newton.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "combustion.h"
5 int main(int argc, char**argv) {
6   environment rheolef (argc,argv);
7   geo omega (argv[1]);
8   Float eps = numeric_limits<Float>::epsilon();
9   string approx = (argc > 2) ? argv[2] : "Pi";
10  Float lambda = (argc > 3) ? atof(argv[3]) : 0.1;
11  Float tol = (argc > 4) ? atof(argv[4]) : eps;
12  size_t max_iter = (argc > 5) ? atoi(argv[5]) : 100;
13  combustion F (omega, approx);
14  F.set_parameter (lambda);
15  field uh = F.initial ();
16  Float residue = tol;
17  size_t n_iter = max_iter;
18  damped_newton (F, uh, residue, n_iter, &derr);
19  F.put (dout, uh);
20  return (residue <= sqrt(tol)) ? 0 : 1;
21 }

```

Let us choose  $\alpha = 1/2$  and  $\lambda = 8(\alpha/\cosh(\alpha))^2 \approx 1.57289546593186$ . Compilation and run are:

```
make combustion_newton
mkgeo_grid -e 10 > line-10.geo
./combustion_newton line-10 P1 1.57289546593186 > line-10.field
field line-10.field
```

### 3.3.2 Error analysis and multiplicity of solutions

In one dimension, when  $\Omega = ]0, 1[$ , the problem can be solved explicitly [71, p. 59]:

- when  $\lambda \leq 0$  the solution is parametrized by  $\alpha \in ]0, \pi/2[$  and:

$$\lambda = -\frac{8\alpha^2}{\cos^2(\alpha)}$$

$$u(x) = 2 \log \left( \frac{\cos(\alpha)}{\cos(\alpha(1-2x))} \right), \quad x \in ]0, 1[$$

- when  $0 \leq \lambda \leq \lambda_c$  there is two solutions. The smallest one is parametrized by  $\alpha \in ]0, \alpha_c]$  and: and the largest by  $\alpha \in ]\alpha_c, +\infty[$  with:

$$\lambda = \frac{8\alpha^2}{\cosh^2(\alpha)}$$

$$u(x) = 2 \log \left( \frac{\cosh(\alpha)}{\cosh(\alpha(1-2x))} \right), \quad x \in ]0, 1[$$

- when  $\lambda > \lambda_c$  there is no more solution.

The critical parameter value  $\lambda_c = 8\alpha_c^2/\cosh^2(\alpha_c)$  where  $\alpha_c$  is the unique positive solution to  $\tanh(\alpha) = 1/\alpha$ . The following code compute  $\alpha_c$  and  $\lambda_c$  by using a Newton method.

Example file 3.25: lambda\_c.icc

```
1 struct alpha_c_fun {
2     typedef Float value_type;
3     typedef Float float_type;
4     alpha_c_fun() : _f1(0) {}
5     Float residue (const Float& a) const { return tanh(a) - 1/a; }
6     void update_derivative (const Float& a) const {
7         _f1 = 1/sqr(cosh(a)) + 1/sqr(a); }
8     Float derivative_solve (const Float& r) const { return r/_f1; }
9     Float dual_space_norm (const Float& r) const { return abs(r); }
10    mutable Float _f1;
11 };
12 Float alpha_c() {
13     Float tol = numeric_limits<Float>::epsilon();
14     size_t max_iter = 100;
15     alpha_c_fun f;
16     Float ac = 1;
17     newton (alpha_c_fun(), ac, tol, max_iter);
18     return ac;
19 }
20 Float lambda_c() {
21     Float ac = alpha_c();
22     return 8*sqr(ac/cosh(ac));
23 }
```

Example file 3.26: lambda\_c.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "lambda_c.icc"
5 int main(int argc, char**argv) {
6     environment rheolef (argc,argv);
7     dout << setprecision(numeric_limits<Float>::digits10)
8         << "alpha_c = " << alpha_c() << endl
9         << "lambda_c = " << lambda_c() << endl;
10 }

```

Compilation and run write:

```

make lambda_c
./lambda_c

```

and then  $\alpha_c \approx 1.19967864025773$  and  $\lambda_c \approx 3.51383071912516$ . The exact solution and its gradient at the limit point are computed by the following functions:

Example file 3.27: combustion\_exact.icc

```

1 #include "lambda2alpha.icc"
2 struct u_exact {
3     Float operator() (const point& x) const {
4         return 2*log(cosh(a)/cosh(a*(1-2*x[0]))); }
5     u_exact (Float lambda, bool is_upper)
6         : a(lambda2alpha(lambda,is_upper)) {}
7     u_exact (Float a1) : a(a1) {}
8     Float a;
9 };
10 struct grad_u {
11     point operator() (const point& x) const {
12         return point(4*a*tanh(a*(1-2*x[0]))); }
13     grad_u (Float lambda, bool is_upper)
14         : a(lambda2alpha(lambda,is_upper)) {}
15     grad_u (Float a1) : a(a1) {}
16     Float a;
17 };

```

The lambda2alpha function converts  $\lambda$  into  $\alpha$ . When  $0 < \lambda < \lambda_c$ , there is two solutions to the equation  $8(\alpha/\cosh(\alpha))^2 = \lambda$  and thus we specify with the boolean `is_upper` which one is expected. Then  $\alpha$  is computed by a dichotomy algorithm.

Example file 3.28: lambda2alpha.icc

```

1 #include "lambda_c.icc"
2 Float lambda2alpha (Float lambda, bool up = false) {
3     static const Float ac = alpha_c();
4     Float tol = 1e2*numeric_limits<Float>::epsilon();
5     size_t max_iter = 1000;
6     Float a_min = up ? ac : 0;
7     Float a_max = up ? 100 : ac;
8     for (size_t k = 0; abs(a_max - a_min) > tol; ++k) {
9         Float a1 = (a_max + a_min)/2;
10        Float lambda1 = 8*sqrt(a1/cosh(a1));
11        if ((up && lambda > lambda1) || (!up && lambda < lambda1))
12            { a_max = a1; }
13        else { a_min = a1; }
14        check_macro (k < max_iter, "lambda2alpha: max_iter=" << k
15                    << " reached and err=" << a_max - a_min);
16    }
17    return(a_max + a_min)/2;
18 };

```

Finally, the errors in various norms are available:

Example file 3.29: combustion\_error.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "combustion_exact.icc"
5 int main(int argc, char**argv) {
6     environment rheolef (argc,argv);
7     bool is_upper = (argc > 1) && (argv[1][0] == '1');
8     bool is_crit = (argc > 1) && (argv[1][0] == 'c');
9     Float tol = (argc > 2) ? atof(argv[2]) : 1e-15;
10    Float lambda_h;
11    field uh;
12    din >> catchmark("lambda") >> lambda_h
13         >> catchmark("u") >> uh;
14    Float lambda = (is_crit ? lambda_c() : lambda_h);
15    const geo& omega = uh.get_geo();
16    const space& Xh = uh.get_space();
17    field pi_h_u = interpolate (Xh, u_exact(lambda, is_upper));
18    field eh = pi_h_u - uh;
19    quadrature_option qopt;
20    qopt.set_family(quadrature_option::gauss);
21    qopt.set_order(2*Xh.degree()+1);
22    Float err_l2
23    = sqrt(integrate(omega, norm2(uh - u_exact(lambda, is_upper)), qopt));
24    Float err_h1
25    = sqrt(integrate(omega, norm2(grad(uh)-grad_u(lambda, is_upper)), qopt));
26    Float err_linf = eh.max_abs();
27    dout << "err_linf = " << err_linf << endl
28         << "err_l2 = " << err_l2 << endl
29         << "err_h1 = " << err_h1 << endl;
30    return (err_h1 < tol) ? 0 : 1;
31 }

```

The computation of the error writes:

```

make combustion_error
./combustion_error < line-10.field

```

The solution is represented on Fig. 3.16. Then we consider the vicinity of  $\lambda_c \approx 3.51383071912516$ .

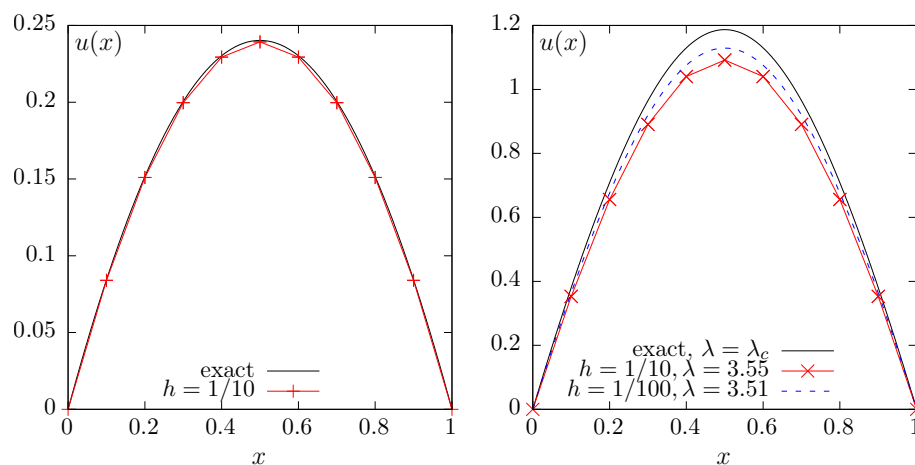


Figure 3.16: Combustion problem: (left)  $\alpha = 1/2$ . (right) near  $\alpha_c$ .

```

./combustion_newton line-10 P1 3.55 > line-10.field
field line-10.field
mkgeo_grid -e 100 > line-100.geo
./combustion_newton line-100 P1 3.51 > line-100.field

```

The Newton method fails when the parameter  $\lambda$  is greater to some critical value  $\lambda_{c,h}$  that depends upon the mesh size. Moreover, while the approximate solution is close to the exact one for moderate  $\lambda = 1.57289546593186$ , the convergence seems to be slower at the vicinity of  $\lambda_c$ .

In the next section, we compute accurately  $\lambda_{c,h}$  for each mesh and observe the convergence of  $\lambda_{c,h}$  to  $\lambda_c$  and the convergence of the associated approximate solution to the exact one.

### 3.3.3 The Euler-Newton continuation algorithm

The Euler-Newton continuation algorithm writes [71, p. 176]:

**algorithm 1** (*continuation*)

- $n = 0$ : Let  $(\lambda_0, u_0)$  be given. Compute

$$\dot{u}_0 = - \left( \frac{\partial F}{\partial u}(\lambda_0, u_0) \right)^{-1} \frac{\partial F}{\partial \lambda}(\lambda_0, u_0)$$

- $n \geq 0$ : Let  $(\lambda_n, u_n)$  and  $\dot{u}_n$  being known.

- 1) First choose a step  $\Delta\lambda_n$  and set  $\lambda_{n+1} = \lambda_n + \Delta\lambda_n$ .
- 2) Then, perform a prediction by computing

$$w_0 = u_n - \Delta\lambda_n \left( \frac{\partial F}{\partial u}(\lambda_n, u_n) \right)^{-1} \frac{\partial F}{\partial \lambda}(\lambda_n, u_n)$$

- 3) Then, perform a correction step: for all  $k \geq 0$ , with  $w_k$  being known, compute

$$w_{k+1} = w_k - \left( \frac{\partial F}{\partial u}(\lambda_{n+1}, w_k) \right)^{-1} F(\lambda_{n+1}, w_k)$$

At convergence of the correction loop, set  $u_{n+1} = w_\infty$ .

- 4) Finally, compute

$$\dot{u}_{n+1} = - \left( \frac{\partial F}{\partial u}(\lambda_{n+1}, u_{n+1}) \right)^{-1} \frac{\partial F}{\partial \lambda}(\lambda_{n+1}, u_{n+1})$$

The step  $\Delta\lambda_n$  can be chosen from a guest  $\Delta\lambda_* = \Delta\lambda_{n-1}$  by adjusting the contraction ratio  $\kappa(\Delta\lambda_*)$  of the Newton method. Computing the two first iterates  $w_{0,*}$  and  $w_{1,*}$  with the guest step  $\Delta\lambda_*$  and  $\lambda_* = \lambda_n + \Delta\lambda_*$  we have:

$$\kappa(\Delta\lambda_*) = \frac{\left\| \left( \frac{\partial F}{\partial u}(\lambda_*, w_{1,*}) \right)^{-1} F(\lambda_*, w_{1,*}) \right\|}{\left\| \left( \frac{\partial F}{\partial u}(\lambda_*, w_{0,*}) \right)^{-1} F(\lambda_*, w_{0,*}) \right\|}$$

As the Newton method is expected to converge quadratically for small enough step, we get a practical expression for  $\Delta\lambda_n$  [71, p. 185]:

$$\frac{\kappa_0}{\Delta\lambda_n} \approx \frac{\kappa(\Delta\lambda_*)}{\Delta\lambda_*^2}$$

where  $\kappa_0 \in ]0, 1[$  is the chosen reference for the contraction ratio, for instance  $\kappa_0 = 1/2$ .

Example file 3.30: combustion\_continuation.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "combustion.h"
5 int main(int argc, char**argv) {
6     environment rheolef (argc, argv);
7     cin >> noverbose;
8     geo omega (argv[1]);
9     string approx = (argc > 2) ? argv[2] : "P1";
10    Float eps = numeric_limits<Float>::epsilon();
11    continuation_option opts;
12    opts.ini_delta_parameter = 0.1;
13    opts.max_delta_parameter = 1;
14    opts.min_delta_parameter = 1e-7;
15    opts.tol = eps;
16    derr << setprecision(numeric_limits<Float>::digits10)
17         << "# continuation in lambda:" << endl
18         << "# geo          = " << omega.name() << endl
19         << "# approx        = " << approx << endl
20         << "# dlambdai      = " << opts.ini_delta_parameter << endl
21         << "# dlambdamin    = " << opts.min_delta_parameter << endl
22         << "# dlambdamax    = " << opts.max_delta_parameter << endl
23         << "# tol          = " << opts.tol << endl;
24    combustion F (omega, approx);
25    field uh = F.initial();
26    F.put (dout, uh);
27    continuation (F, uh, &dout, &derr, opts);
28 }

```

Then, the program is compiled and run as:

```

make combustion_continuation
mkgeo_grid -e 10 > line-10.geo
./combustion_continuation line-10 > line-10.branch
branch line-10.branch -toc

```

The last command lists all the computations performed by the continuation algorithm. The last

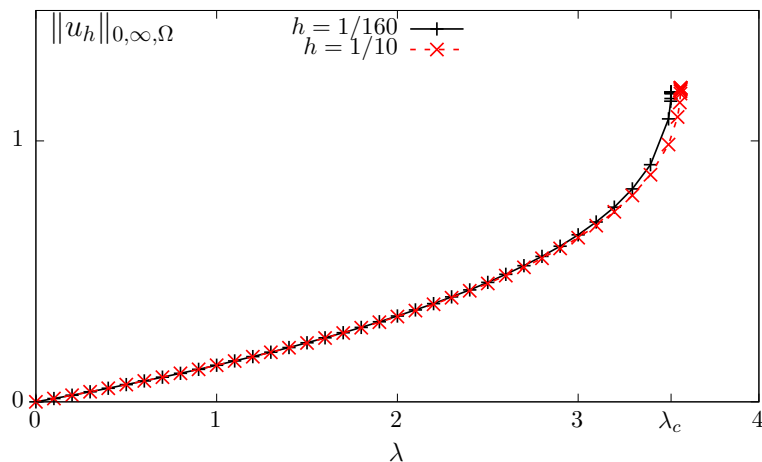


Figure 3.17: Combustion problem:  $\|u_h\|_{0,\infty,\Omega}$  vs  $\lambda$  when  $h = 1/10$ ,  $1/20$  and  $1/160$ .

recorded computation is associated to the *limit point* denoted and denoted as  $\lambda_{c,h}$ , says at index 24: Let us visualize the solution  $u_h$  at the limit point and compute its maximum  $\|u_h\|_{0,\infty,\Omega} = u_h(1/2)$ :

```
branch line-10.branch -extract 24 | field -
branch line-10.branch -extract 24 | field -max -
```

Fig. 3.17 plots  $\|u_h\|_{0,\infty,\Omega}$  versus  $\lambda$  for various meshes.

Fig. 3.18 plots its convergence to  $\lambda_c$  and also the convergence of the corresponding approximate solution  $u_h$  to the exact one  $u$ , associated to  $\lambda_c$ . Observe that  $|\lambda_{c,h} - \lambda_c|$  converges to zero

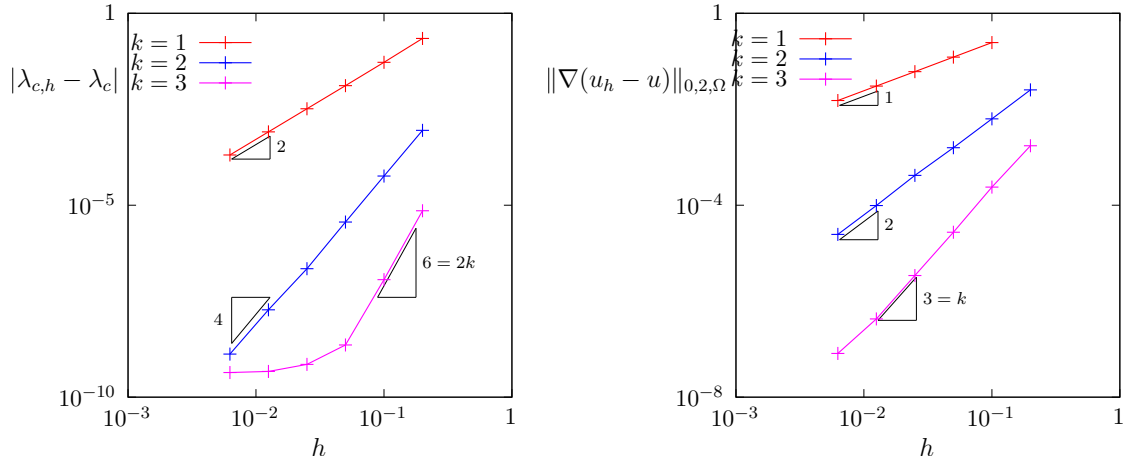


Figure 3.18: Combustion problem: (left) convergence of  $|\lambda_{c,h} - \lambda_c|$  vs  $h$  ; (right) convergence of  $|u_h - u|$  vs  $h$  at the limit point.

as  $\mathcal{O}(h^{2k})$  while  $|u_h - u| = \mathcal{O}(h^k)$ . When  $k = 3$ , the convergence of  $\lambda_{c,h}$  slows down around  $10^{-9}$ : this is due to the stopping criterion of the Newton method that detects automatically the propagation of rounding effects in the resolution of linear systems. Observe on Fig. 3.19 the plot

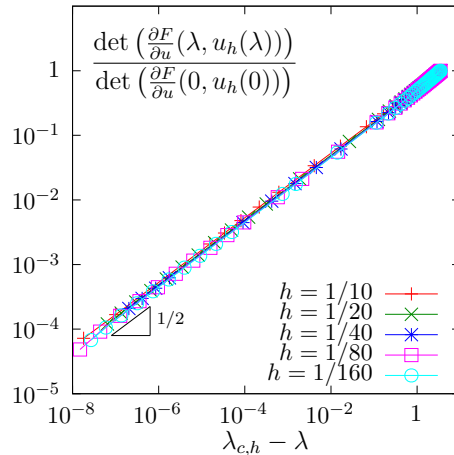


Figure 3.19: Combustion problem:  $\det \left( \frac{\partial F}{\partial u}(\lambda, u_h) \right)$  versus  $\lambda_{c,h} - \lambda$  for various  $h$  and  $k = 1$ .

of the determinant of the Jacobian  $\frac{\partial F}{\partial u}(\lambda, u_h(\lambda))$  versus  $\lambda_{c,h} - \lambda$ : it tends to zero at the vicinity of  $\lambda = \lambda_c$  and thus, the Newton method reaches increasing difficulties to solve the problem. Notice that the determinant has been normalized by its value when  $\lambda = 0$ , i.e.  $\det \left( \frac{\partial F}{\partial u}(0, u_h(0)) \right)$ : all the curves tend to superpose on a unique master curve and the asymptotic behavior is independent

of  $h$ . More precisely, Fig. 3.19 suggests a 1/2 slope in logarithmic scale and thus

$$\frac{\det\left(\frac{\partial F}{\partial u}(\lambda, u_h(\lambda))\right)}{\det\left(\frac{\partial F}{\partial u}(0, u_h(0))\right)} \approx C(\lambda_{c,h} - \lambda)^{1/2}$$

where  $C \approx 0.52$  when  $k = 1$ . This behavior explains that the Newton method is unable to reach the limit point  $\lambda_c$ .

### 3.3.4 Beyond the limit point : the Keller algorithm

Notice that the continuation method stops to the limit point (see Fig. 3.17) while the the branch continues: the limit point is a turning point for the branch of solutions. Especially, for each  $\lambda \in ]0, \lambda_c[$  there are two solutions and only one has been computed. Keller proposed a method to follow the branch beyond a turning point and this method is presented here. The main idea is to parametrize the branch  $(\lambda, u(\lambda))$  by a curvilinear abscissa  $s$  as  $(\lambda(s), u(s))$ . In order to have the count of unknown and equations we add a normalization equation:

$$\begin{aligned} \mathcal{N}(s, \lambda(s), u(s)) &= 0 \\ F(\lambda(s), u(s)) &= 0 \end{aligned}$$

where  $\mathcal{N}$  is a given normalization function. For the normalization function, Keller proposed to choose, when  $(s, \lambda, u)$  is at the vicinity of  $(s_n, \lambda(s_n), u(s_n))$  the one following orthogonal norms:

- The **orthogonal** norm:

$$\begin{aligned} \mathcal{N}_n(s, \chi=(\lambda, u)) &= (\chi'(s_n), \chi - \chi(s_n)) - (s - s_n) \\ &= \lambda'(s_n)(\lambda - \lambda(s_n)) + (u'(s_n), u - u(s_n))_V - (s - s_n) \end{aligned} \quad (3.8a)$$

- The **spherical** norm:

$$\begin{aligned} \tilde{\mathcal{N}}_n(s, \chi=(\lambda, u)) &= \|\chi - \chi_n\|^2 - |s - s_n|^2 \\ &= |\lambda - \lambda_n|^2 + \|u - u_n\|_V^2 - |s - s_n|^2 \end{aligned} \quad (3.8b)$$

The orthogonal norm induces a pseudo curvilinear arc-length  $s$ , measured on the tangent at  $s = s_n$ . The spherical norm measures is simply a distance between  $(s, \chi)$  and  $(s_n, \chi_n)$  (see also [71, pp. 179-180] and the corresponding Fig. 5.2). We add the subscript  $n$  to  $\mathcal{N}$  in order to emphasize that  $\mathcal{N}$  depends upon both  $(\lambda(s_n), u(s_n))$  and  $(\lambda'(s_n), u'(s_n))$ . For any  $s \in \mathbb{R}$  and  $\chi = (\lambda, u) \in \mathbb{R} \times V$  we introduce:

$$\mathcal{F}_n(s, \chi) = \begin{pmatrix} \mathcal{N}_n(s, \chi) \\ F(\chi) \end{pmatrix} \quad \text{and} \quad \tilde{\mathcal{F}}_n(s, \chi) = \begin{pmatrix} \tilde{\mathcal{N}}_n(s, \chi) \\ F(\chi) \end{pmatrix}$$

Then, the Keller problem with the orthogonal norm reduces to find, for any  $s \in \mathbb{R}$ ,  $\chi(s) \in \mathbb{R} \times V$  such that

$$\mathcal{F}_n(s, \chi(s)) = 0 \quad (3.9a)$$

Conversely, the Keller problem with the spherical norm reduces to find, for any  $s \in \mathbb{R}$ ,  $\chi(s) \in \mathbb{R} \times V$  such that

$$\tilde{\mathcal{F}}_n(s, \chi(s)) = 0 \quad (3.9b)$$

Both problems falls into the framework of the previous paragraph, when  $F$  is replaced by either  $\mathcal{F}_n$  or  $\tilde{\mathcal{F}}_n$ . Then, for any  $s$  and  $\chi = (\lambda, u)$ , the partial derivatives are:

$$\begin{aligned} \frac{\partial \mathcal{F}_n}{\partial s}(s, \chi) &= \begin{pmatrix} -1 \\ 0 \end{pmatrix} \\ \frac{\partial \mathcal{F}_n}{\partial \chi}(s, \chi) &= \begin{pmatrix} \frac{\partial \mathcal{N}_n}{\partial \chi}(s, \chi) \\ F'(s, \chi) \end{pmatrix} = \begin{pmatrix} \lambda'(s_n) & u'(s_n) \\ \frac{\partial F}{\partial \lambda}(\lambda, u) & \frac{\partial F}{\partial u}(\lambda, u) \end{pmatrix} \end{aligned}$$



and

$$\begin{aligned}\frac{\partial \tilde{\mathcal{F}}_n}{\partial s}(s, \chi) &= \begin{pmatrix} -2(s - s_n) \\ 0 \end{pmatrix} \\ \frac{\partial \tilde{\mathcal{F}}_n}{\partial \chi}(s, \chi) &= \begin{pmatrix} 2(\chi - \chi_n)^T \\ F'(\chi) \end{pmatrix} = \begin{pmatrix} 2(\lambda - \lambda_n) & 2(u - u_n)^T \\ \frac{\partial F}{\partial \lambda}(\lambda, u) & \frac{\partial F}{\partial u}(\lambda, u) \end{pmatrix}\end{aligned}$$

Let us focus on the orthogonal norm case, as the spherical one is similar. The continuation algorithm of the previous paragraph is able to follow the branch of solution beyond the limit point and explore the second part of the branch. Let us compute  $\lambda'(s_n)$  and  $u'(s_n)$ . By differentiating (3.9) with respect to  $s$ , we get:

$$\frac{\partial \mathcal{F}_n}{\partial s}(s, \chi(s)) + \frac{\partial \mathcal{F}_n}{\partial \chi}(s, \chi(s)) \cdot (\chi'(s)) = 0$$

that writes equivalently

$$\begin{aligned}\frac{\partial \mathcal{N}_n}{\partial s}(s, \chi(s)) + \frac{\partial \mathcal{N}_n}{\partial \chi}(s, \chi(s)) \cdot (\chi'(s)) &= 0 \\ F'(s, \chi(s)) \cdot (\chi'(s)) &= 0\end{aligned}$$

Using the expression 3.8 for  $\mathcal{N}_n$  we obtain:

$$-1 + \lambda'(s_n)\lambda'(s) + (u'(s_n), u'(s)) = 0 \quad (3.10)$$

$$\frac{\partial F}{\partial \lambda}(\lambda(s), u(s))\lambda'(s) + \frac{\partial F}{\partial u}(\lambda(s), u(s)) \cdot (u'(s)) = 0 \quad (3.11)$$

Here  $(\cdot, \cdot)$  denotes the scalar product of the  $V$  space for  $u$ . Let us choose  $s = s_n$ , for any  $n \geq 0$ : we obtain

$$\begin{aligned}|\lambda'(s_n)|^2 + \|u'(s_n)\|^2 &= 1 \\ \frac{\partial F}{\partial \lambda}(\lambda_n, u_n)\lambda'(s_n) + \frac{\partial F}{\partial u}(\lambda_n, u_n) \cdot (u'(s_n)) &= 0\end{aligned}$$

where we use the notations  $\lambda_n = \lambda(s_n)$  and  $u_n = u(s_n)$ , and where  $\|\cdot\|$  denotes the norm of the  $V$  space. Thus

$$\begin{aligned}\chi'(s_n) &= \begin{pmatrix} \lambda'(s_n) \\ u'(s_n) \end{pmatrix} \\ &= \frac{1}{\left(1 + \left\| \left( \frac{\partial F}{\partial u}(\lambda_n, u_n) \right)^{-1} \frac{\partial F}{\partial \lambda}(\lambda_n, u_n) \right\|^2\right)^{1/2}} \begin{pmatrix} 1 \\ - \left( \frac{\partial F}{\partial u}(\lambda_n, u_n) \right)^{-1} \frac{\partial F}{\partial \lambda}(\lambda_n, u_n) \end{pmatrix}\end{aligned}$$

The previous relation requires  $\frac{\partial F}{\partial u}(\lambda_n, u_n)$  to be nonsingular, e.g. the computation is not possible at a singular point  $(\lambda_n, u_n)$ .

For a singular point, suppose that  $n \geq 1$  and that both  $(\lambda_n, u_n)$ ,  $(\lambda_{n-1}, u_{n-1})$  and  $(\dot{\lambda}_{n-1}, \dot{u}_{n-1})$  are known. By differentiating (3.9) at step  $n - 1$  we get the equivalent of (3.10)-(3.11) at step  $n - 1$  that is then evaluated for  $s = s_n$ . We get:

$$\begin{aligned}\lambda'(s_{n-1})\lambda'(s_n) + (u'(s_{n-1}), u'(s_n)) &= 1 \\ \frac{\partial F}{\partial \lambda}(\lambda_n, u_n)\lambda'(s_n) + \frac{\partial F}{\partial u}(\lambda_n, u_n) \cdot (u'(s_n)) &= 0\end{aligned}$$

that writes equivalently

$$\begin{pmatrix} \lambda'(s_{n-1}) & u'(s_{n-1}) \\ \frac{\partial F}{\partial \lambda}(\lambda_n, u_n) & \frac{\partial F}{\partial u}(\lambda_n, u_n) \end{pmatrix} \begin{pmatrix} \lambda'(s_n) \\ u'(s_n) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (3.12)$$

The matrix involved in the left hand side is exactly the Jacobian of  $\mathcal{F}_{n-1}$  evaluated at point  $\chi_n = (\lambda_n, u_n)$ . This Jacobian is expected to be nonsingular at a simple limit point.

Thus, at the first step, we suppose that the initial point  $(\lambda_0, u_0)$  is non-singular and we compute  $(\dot{\lambda}_0, \dot{u}_0)$ . Then, at the beginning of the  $n$ -th step,  $n \geq 1$ , of the Keller continuation algorithm, we suppose that both  $(\lambda_{n-1}, u_{n-1})$  and  $(\lambda_{n-1}, \dot{u}_{n-1})$  are known. We consider the problem  $\mathcal{F}_{n-1}(s, \chi(s)) = 0$ . Here,  $\mathcal{F}_{n-1}(s, \chi)$  is completely defined at the vicinity of  $(s_{n-1}, \lambda_{n-1}, u_{n-1})$ . The step control procedure furnishes as usual a parameter step  $\Delta s_{n-1}$  and we set  $s_n = s_{n-1} + \Delta s_{n-1}$ . The Newton method is performed and we get  $(\lambda_n, u_n)$ . Finally, we compute  $\dot{\lambda}_n$  and  $\dot{u}_n$  from (3.12).

Recall that the function  $\mathcal{F}_{n-1}$  depends upon  $n$  and should be *refreshed* at the beginning of each iteration by using the values  $(\dot{\lambda}_{n-1}, \dot{u}_{n-1})$ .

The Keller continuation algorithm writes:

**algorithm 2** (*Keller continuation*)

- $n = 0$ : Let  $(s_0, \chi_0 = (\lambda_0, u_0))$  be given. The recurrence requires also  $\dot{\chi}_0$  and its orientation  $\varepsilon_0 \in \{-1, +1\}$ : they could either be given or computed. When computing  $(\dot{\chi}_0, \varepsilon_0)$ , the present algorithm supposes that  $(\lambda_0, u_0)$  is a regular point: on a singular point, e.g. a bifurcation one, there are several possible directions, and one should be chosen. Then, choose  $\varepsilon = \pm 1$  and compute  $\dot{\chi}_0$  in three steps:

$$\begin{aligned} \frac{du}{d\lambda}(\lambda_0) &= - \left( \frac{\partial F}{\partial u}(\lambda_0, u_0) \right)^{-1} \frac{\partial F}{\partial \lambda}(\lambda_0, u_0) \\ c &= \left( 1 + \left\| \frac{du}{d\lambda}(\lambda_0) \right\| \right)^{-1/2} \\ \dot{\chi}_0 &\stackrel{def}{=} (\dot{\lambda}_0, \dot{u}_0)^T = c \left( 1, \frac{du}{d\lambda}(\lambda_0) \right)^T \end{aligned}$$

- $n \geq 0$ : Let  $(s_n, \chi_n = (\lambda_n, u_n))$ ,  $\dot{\chi}_n = (\dot{\lambda}_n, \dot{u}_n)$  and  $\varepsilon_n$  being known.
  - 1) First choose a step  $\Delta s_n$  and set  $s_{n+1} = s_n + \Delta s_n$ , as in the classical continuation algorithm
  - 2) Then, perform a prediction, as usual:

$$y_0 = \chi_n + \varepsilon_n \Delta s_n \dot{\chi}_n$$

- 3) Also as usual, do a correction loop: for all  $k \geq 0$ ,  $y_k$  being known, compute

$$y_{k+1} = y_k - \left( \frac{\partial \mathcal{F}_n}{\partial \chi}(s_{n+1}, y_k) \right)^{-1} \mathcal{F}_n(s_{n+1}, y_k)$$

At convergence of the correction loop, set  $\chi_{n+1} = y_\infty$ . This Newton correction loop can be replaced by damped Newton one.

- 4) **Check**: if  $n \geq 1$ , compute the following angle cosinus:

$$\begin{aligned} c_1 &= (\chi_{n+1} - \chi_n, \chi_n - \chi_{n-1}) \\ &= (\lambda_{n+1} - \lambda_n)(\lambda_n - \lambda_{n-1}) + (u_{n+1} - u_n, u_n - u_{n-1})_V \\ c_2 &= (\dot{\chi}_n, \chi_{n+1} - \chi_n) \\ &= (\dot{\lambda}_n, \lambda_{n+1} - \lambda_n) + (\dot{u}_n, u_{n+1} - u_n)_V \end{aligned}$$

When either  $c_1 \leq 0$  or  $c_2 \leq 0$ , then decreases  $\Delta s_n$  and go back at step 1. Otherwise the computation of  $\chi_{n+1}$  is validated.

- 5) Finally, compute  $\dot{\chi}_{n+1} = (\dot{\lambda}_{n+1}, \dot{u}_{n+1})$  as:

$$\dot{\chi}_{n+1} = - \left( \frac{\partial \mathcal{F}_n}{\partial \chi}(s_{n+1}, \chi_{n+1}) \right)^{-1} \frac{\partial \mathcal{F}_n}{\partial s}(s_{n+1}, \chi_{n+1})$$

If  $\varepsilon_n (\dot{\chi}_{n+1}, \dot{\chi}_n) \geq 0$  then set  $\varepsilon_{n+1} = \varepsilon_n$  else  $\varepsilon_{n+1} = -\varepsilon_n$ .

The Keller algorithm with the spherical norm is simply obtained by replacing  $\mathcal{F}_n$  by  $\tilde{\mathcal{F}}_n$ . Both algorithm variants still require to save  $\dot{\chi}_n$  and  $\varepsilon_n$  at each step for restarting nicely with the same sequence of computation. The only drawback is that, at a restart of the algorithm, we skip the first **Check** step, and its possible to go back at the first iterate if  $\Delta s_0$  is too large. A possible remedy is, when retarting, to furnish two previous iterates, nammely  $\chi_{-1}$  and  $\chi_0$ , together with  $\dot{\chi}_0$ :  $\chi_{-1}$  is used only for the **Check** of a possible change of direction.

In [50], the author suggests that the Keller algorithm with spherical norm is more robust that its variant with orthogonal one. In [50], the author repport that

*[...] the spherical constraint  $N_2$  is much more efficient than the orthogonal constraint  $N_1$ , as  $N_2$  required only one step of length 0.01 to exceed the target value, while  $N_1$  required 25, with 8 additional convergence failures.*

*[...] spherical constraint was seen to be more efficient than an orthogonal constraint in a region of high curvature of the solution manifold, while both constraints performed similarly in regions of low or moderate curvature.*

Example file 3.31: combustion\_keller.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "combustion.h"
5 int main(int argc, char**argv) {
6     environment rheolef (argc, argv);
7     cin >> noverbose;
8     geo omega (argv[1]);
9     string approx = (argc > 2) ? argv[2] : "P1";
10    string metric = (argc > 3) ? argv[3] : "orthogonal";
11    Float eps = numeric_limits<Float>::epsilon();
12    continuation_option opts;
13    opts.ini_delta_parameter = 0.1;
14    opts.max_delta_parameter = 0.5;
15    opts.min_delta_parameter = 1e-10;
16    opts.tol = eps;
17    derr << setprecision(numeric_limits<Float>::digits10)
18    << "# continuation in s:" << endl
19    << "# geo      = " << omega.name() << endl
20    << "# approx   = " << approx << endl
21    << "# metric    = " << metric << endl
22    << "# ds_init   = " << opts.ini_delta_parameter << endl
23    << "# ds_min    = " << opts.min_delta_parameter << endl
24    << "# ds_max    = " << opts.max_delta_parameter << endl
25    << "# tol      = " << opts.tol << endl;
26    dout << catchmark("metric") << metric << endl;
27    keller<combustion> F (combustion(omega,approx), metric);
28    keller<combustion>::value_type xh = F.initial();
29    F.put (dout, xh);
30    continuation (F, xh, &dout, &derr, opts);
31 }

```

```

make combustion_keller
mkgeo_grid -e 20 > line-20.geo

```

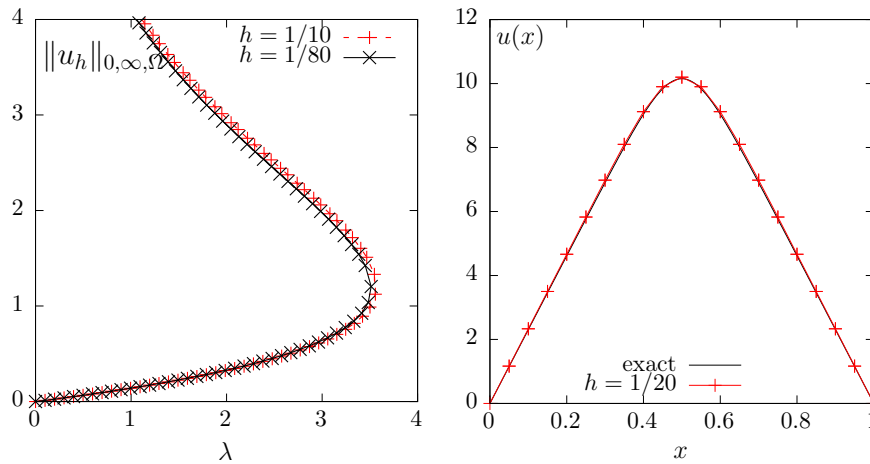


Figure 3.20: Combustion problem: (left)  $\|u_h\|_{0,\infty,\Omega}$  vs  $\lambda$  when  $h = 1/10$ ,  $1/20$  and  $1/160$ . The full branch of solutions when  $\lambda \geq 0$ , with the limit point and the upper part of the branch. (right) solution of the upper branch when  $\lambda = 10^{-2}$ .

```
./combustion_keller line-20 > line-20.branch
make combustion_keller_post
./combustion_keller_post < line-20.branch
branch line-20.branch -toc
make combustion_error
branch line-20.branch -extract 41 | ./combustion_error 1
```

Please, replace 41 by the last computed index on the branch. Observe on Fig. 3.20 the full branch of solutions when  $\lambda \geq 0$ , with the limit point and the upper part of the branch. Compare it with Fig. 3.17, where the continuation algorithm was limited to the lower part part of the branch.

Example file 3.32: combustion\_keller\_post.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "combustion.h"
5 int main(int argc, char**argv) {
6     environment rheolef (argc,argv);
7     string metric;
8     din >> catchmark("metric") >> metric;
9     keller<combustion> F (combustion(), metric);
10    keller<combustion>::value_type xh, dot_xh;
11    dout << noverbose
12         << setprecision(numeric_limits<Float>::digits10)
13         << "# metric " << metric << endl
14         << "# s lambda umax det(mantissa,base,exp) |u| |grad(u)| |residue|"
15         << endl;
16    for (size_t n = 0; F.get(din,xh); ++n) {
17        solver::determinant_type det;
18        if (n > 0 || metric != "spherical") det = F.update_derivative(xh);
19        const space& Xh = xh.second.get_space();
20        trial u (Xh); test v (Xh);
21        form a = integrate(dot(grad(u),grad(v))),
22              m = integrate(u*v);
23        const combustion& F0 = F.get_problem();
24        field mrh = F0.residue(xh.second);
25        dout << F.parameter() << " " << xh.first
26             << " " << xh.second.max_abs()
27             << " " << det.mantissa
28             << " " << det.base
29             << " " << det.exponent
30             << " " << sqrt(m(xh.second,xh.second))
31             << " " << sqrt(a(xh.second,xh.second))
32             << " " << sqrt(F0.dual_space_dot (mrh,mrh))
33             << endl;
34        dot_xh = F.direction (xh);
35        F.refresh (F.parameter(), xh, dot_xh);
36    }
37 }

```

# Chapter 4

## Discontinuous Galerkin methods

### 4.1 Scalar first-order problems

#### 4.1.1 The transport equation

The aim of this chapter is to introduce to discontinuous Galerkin methods within the **Rheolef** environment. For some recent presentations of discontinuous Galerkin methods, see [31] for theoretical aspects and [47] for algorithmic and implementation. The discontinuous Galerkin methods are in active development in **Rheolef**, and new features will appear soon.

The steady scalar transport problem writes:

(P): find  $\phi$ , defined in  $\Omega$ , such that

$$\begin{aligned} \mathbf{u} \cdot \nabla \phi + \sigma \phi &= f \text{ in } \Omega \\ \phi &= \phi_\Gamma \text{ on } \partial\Omega_- \end{aligned}$$

where  $\mathbf{u}$ ,  $\sigma > 0$ ,  $f$  and  $\phi_\Gamma$  being known. Notice that this is the steady version of the unsteady diffusion-convection problem previously introduced in section 2.3.2, page 73 and when the diffusion coefficient  $\nu$  vanishes. Here, the  $\partial\Omega_-$  notation is the *upstream* boundary part, defined by

$$\partial\Omega_- = \{x \in \partial\Omega; \mathbf{u}(x) \cdot \mathbf{n}(x) < 0\}$$

Let us suppose that  $\mathbf{u} \in W^{1,\infty}(\Omega)^d$  and introduce the space:

$$X = \{\varphi \in L^2(\Omega); (\mathbf{u} \cdot \nabla)\varphi \in L^2(\Omega)^d\}$$

and, for all  $\phi, \varphi \in X$

$$\begin{aligned} a(\phi, \varphi) &= \int_{\Omega} (\mathbf{u} \cdot \nabla \phi \varphi + \sigma \phi \varphi) \, dx + \int_{\partial\Omega} \max(0, -\mathbf{u} \cdot \mathbf{n}) \phi \varphi \, ds \\ l(\varphi) &= \int_{\Omega} f \varphi \, dx + \int_{\partial\Omega} \max(0, -\mathbf{u} \cdot \mathbf{n}) \phi_\Gamma \varphi \, ds \end{aligned}$$

Then, the variational formulation writes:

(FV): find  $\phi \in X$  such that

$$a(\phi, \varphi) = l(\varphi), \quad \forall \varphi \in X$$

Notice that the term  $\max(0, -\mathbf{u} \cdot \mathbf{n}) = (|\mathbf{u} \cdot \mathbf{n}| - \mathbf{u} \cdot \mathbf{n})/2$  is positive and vanishes everywhere except on  $\partial\Omega_-$ . Thus, the boundary condition  $\phi = \phi_\Gamma$  is weakly imposed on  $\partial\Omega_-$  via the integrals on the boundary. The *discontinuous* finite element space is defined by:

$$X_h = \{\varphi_h \in L^2(\Omega); \varphi_h|_K \in P_k, \quad \forall K \in \mathcal{T}_h\}$$

where  $k \geq 0$  is the polynomial degree. Notice that  $X_h \not\subset X$  and that the  $\nabla\phi_h$  term has no more sense for discontinuous functions  $\phi_h \in X_h$ . Following [31, p. 14], we introduce the *broken gradient*  $\nabla_h$  as a convenient notation:

$$(\nabla_h \phi_h)|_K = \nabla(\phi_h|_K), \quad \forall K \in \mathcal{T}_h$$

Thus

$$\int_{\Omega} \mathbf{u} \cdot \nabla_h \phi_h \varphi_h \, dx = \sum_{K \in \mathcal{T}_h} \int_K \mathbf{u} \cdot \nabla \phi_h \varphi_h \, dx, \quad \forall \phi_h, \varphi_h \in X_h$$

This leads to a discrete version  $a_h$  of the bilinear form  $a$ , defined for all  $\phi_h, \varphi_h \in X_h$  by (see e.g. [31, p. 57], eqn. (2.34)):

$$\begin{aligned} a_h(\phi_h, \varphi_h) &= \int_{\Omega} (\mathbf{u} \cdot \nabla_h \phi_h \varphi_h + \sigma \phi_h \varphi_h) \, dx + \int_{\partial\Omega} \max(0, -\mathbf{u} \cdot \mathbf{n}) \phi_h \varphi_h \, ds \\ &\quad + \sum_{S \in \mathcal{S}_h^{(i)}} \int_S \left( -\mathbf{u} \cdot \mathbf{n} \llbracket \phi_h \rrbracket \{\!\! \{ \varphi_h \}\!\!\} + \frac{\alpha}{2} |\mathbf{u} \cdot \mathbf{n}| \llbracket \phi_h \rrbracket \llbracket \varphi_h \rrbracket \right) \, ds \end{aligned}$$

The last term involves a sum over  $\mathcal{S}_h^{(i)}$ , the set of *internal sides* of the mesh  $\mathcal{T}_h$ . Each internal side

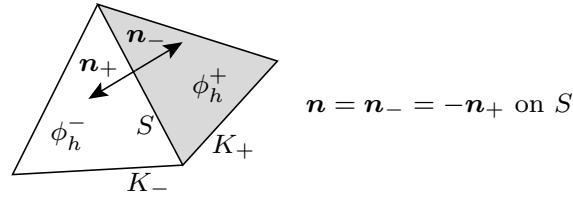


Figure 4.1: Discontinuous Galerkin method: an internal side, its two neighbor elements and their opposite normals.

$S \in \mathcal{S}_h^{(i)}$  has two possible orientations: one is chosen definitively. In practice, this orientation is defined in the ‘.geo’ file containing the mesh, where all sides are listed, together with their orientation. Let  $\mathbf{n}$  the normal to the oriented side  $S$ : as  $S$  is an internal side, there exists two elements  $K_-$  and  $K_+$  such that  $S = \partial K_- \cap \partial K_+$  and  $\mathbf{n}$  is the outward unit normal of  $K_-$  on  $\partial K_- \cap S$  and the inward unit normal of  $K_+$  on  $\partial K_+ \cap S$ , as shown on Fig. 4.1. For all  $\phi_h \in X_h$ , recall that  $\phi_h$  is in general discontinuous across the internal side  $S$ . We define on  $S$  the *inner value*  $\phi_h^- = \phi_h|_{K_-}$  of  $\phi_h$  as the restriction  $\phi_h|_{K_-}$  of  $\phi_h$  in  $K_-$  along  $\partial K_- \cap S$ . Conversely, we define the *outer value*  $\phi_h^+ = \phi_h|_{K_+}$ . We also denote on  $S$  the *jump*  $\llbracket \phi_h \rrbracket = \phi_h^- - \phi_h^+$  and the *average*  $\{\!\! \{ \phi_h \}\!\!\} = (\phi_h^- + \phi_h^+)/2$ . The last term in the definition of  $a_h$  is ponderated by a coefficient  $\alpha \geq 0$ . Choosing  $\alpha = 0$  correspond to the so-called *centered flux* approximation, while  $\alpha = 1$  is the *upwinding flux* approximation. The case  $\alpha = 1$  and  $k = 0$  (piecewise constant approximation) leads to the popular upwinding finite volume scheme. Finally, the discrete variational formulation writes:

(FV) $_h$ : find  $\phi_h \in X_h$  such that

$$a_h(\phi_h, \varphi_h) = l(\varphi_h), \quad \forall \varphi_h \in X_h$$

The following code implement this problem in the **Rheolef** environment.

Example file 4.1: transport\_dg.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 int main(int argc, char**argv) {
5     environment rheolef (argc, argv);
6     geo omega (argv[1]);
7     space Xh (omega, argv[2]);
8     Float alpha = (argc > 3) ? atof(argv[3]) : 1;
9     Float sigma = (argc > 4) ? atof(argv[4]) : 3;
10    point u (1,0,0);
11    trial phi (Xh); test psi (Xh);
12    form ah = integrate (dot(u,grad_h(phi))*psi + sigma*phi*psi)
13              + integrate ("boundary", max(0, -dot(u,normal()))*phi*psi)
14              + integrate ("internal_sides",
15                          - dot(u,normal())*jump(phi)*average(psi)
16                          + 0.5*alpha*abs(dot(u,normal()))*jump(phi)*jump(psi));
17    field lh = integrate ("boundary", max(0, -dot(u,normal()))*psi);
18    solver sah (ah.uu());
19    field phi_h(Xh);
20    phi_h.set_u() = sah.solve(lh.u());
21    dout << catchmark("sigma") << sigma << endl
22          << catchmark("phi") << phi_h;
23 }

```

### Comments

The data are  $\phi_\gamma = 1$  and  $\mathbf{u} = (1, 0, 0)$ , and then the exact solution is known:  $\phi(x) = \exp(-\sigma x_0)$ . The numerical tests are running with  $\sigma = 3$  by default. The one-dimensional case writes:

```

make transport_dg
mkgeo_grid -e 10 > line.geo
./transport_dg line P0 | field -
./transport_dg line P1d | field -
./transport_dg line P2d | field -

```

Observe the jumps across elements: these jumps decrease with mesh refinement or when polynomial approximation increases. The two-dimensional case writes:

```

mkgeo_grid -t 10 > square.geo
./transport_dg square P0 | field -elevation -
./transport_dg square P1d | field -elevation -
./transport_dg square P2d | field -elevation -

```

The elevation view shows details on inter-element jumps. Finally, the three-dimensional case writes:

```

mkgeo_grid -T 5 > cube.geo
./transport_dg cube P0 | field -
./transport_dg cube P1d | field -
./transport_dg cube P2d | field -

```

Fig. 4.2 plots the solution when  $d = 1$  and  $k = 0$ : observe that the boundary condition  $\phi = 1$  at  $x_0 = 0$  is only weakly satisfied. It means that the approximation  $\phi_h(0)$  tends to 1 when  $h$  tends to zero. Fig. 4.3 plots the error  $\phi - \phi_h$  in  $L^2$  and  $L^\infty$  norms: these errors behave as  $\mathcal{O}(h^{k+1})$  for all  $k \geq 0$ , which is optimal. A theoretical  $\mathcal{O}(h^{k+1/2})$  error bound was shown in [52]. The present numerical results confirm that these theoretical error bounds can be improved for some families of meshes, as pointed out by Richter [78], that showed a  $\mathcal{O}(h^{k+1})$  optimal bound for the transport problem. This result was recently extended by Cockburn *et al.* [24], while Peterson [73] showed that the estimate  $\mathcal{O}(h^{k+1/2})$  is sharp for general families of quasi-uniform meshes.



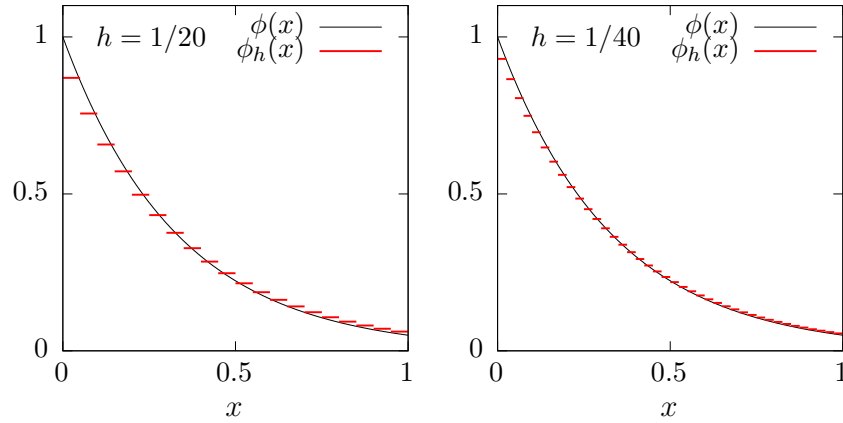


Figure 4.2: The discontinuous Galerkin method for the transport problem when  $k = 0$  and  $d = 1$ .

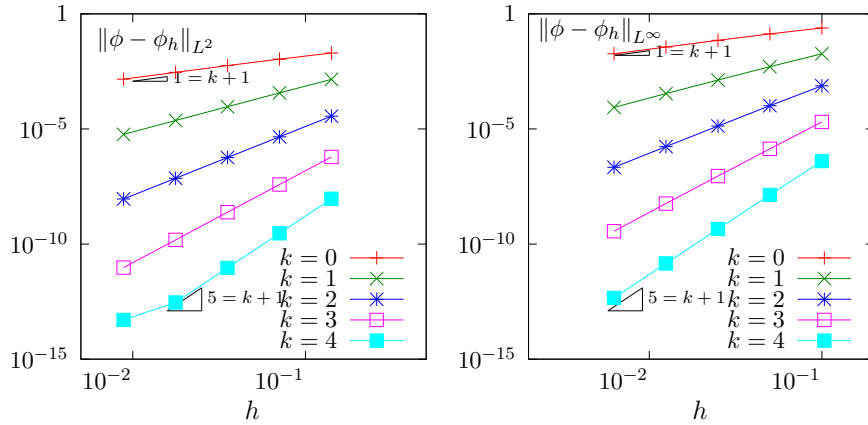


Figure 4.3: The discontinuous Galerkin method for the transport problem: convergence when  $d = 2$ .

### 4.1.2 Nonlinear scalar hyperbolic problems

The aim of this paragraph is to study the discontinuous Galerkin discretization of scalar nonlinear hyperbolic equations. This section presents the general framework and discretization tools while the next section illustrates the method for the Burgers equation.

#### Problem setting

A time-dependent nonlinear hyperbolic problem writes in general form [31, p. 99]:

(P): find  $u$ , defined in  $]0, T[ \times \Omega$ , such that

$$\frac{\partial u}{\partial t} + \operatorname{div} \mathbf{f}(u) = 0 \text{ in } ]0, T[ \times \Omega \tag{4.1a}$$

$$u(t=0) = u_0 \text{ in } \Omega \tag{4.1b}$$

$$\mathbf{f}(u) \cdot \mathbf{n} = \Phi(\mathbf{n}; u, g) \text{ on } ]0, T[ \times \partial\Omega \tag{4.1c}$$

where  $T > 0$ ,  $\Omega \subset \mathbb{R}^d$ ,  $d = 1, 2, 3$  and the initial condition  $\mathbf{u}_0$  being known. As usual,  $\mathbf{n}$  denotes the outward unit normal on the boundary  $\partial\Omega$ . The function  $\mathbf{f} : \mathbb{R} \rightarrow \mathbb{R}^d$  is also known and supposed to be continuously differentiable. The initial data  $u_0$ , defined in  $\Omega$ , and the boundary one,  $g$ ,

defined on  $\partial\Omega$  are given. The function  $\Phi$ , called the Godunov flux associated to  $\mathbf{f}$ , is defined, for all  $\boldsymbol{\nu} \in \mathbb{R}^d$  and  $a, b \in \mathbb{R}$ , by

$$\Phi(\boldsymbol{\nu}; a, b) = \begin{cases} \min_{v \in [a, b]} \mathbf{f}(v) \cdot \boldsymbol{\nu} & \text{when } a \leq b \\ \max_{v \in [b, a]} \mathbf{f}(v) \cdot \boldsymbol{\nu} & \text{otherwise} \end{cases} \quad (4.1d)$$

### Space discretization

In this section, we consider first the semi-discretization with respect to space while the problem remains continuous with respect to time. The semi-discrete problem writes in variational form [31, p. 100]:

(P)<sub>h</sub>: find  $u_h \in C^1([0, T], X_h)$  such that

$$\int_{\Omega} \frac{\partial u_h}{\partial t} v_h \, dx - \int_{\Omega} \mathbf{f}(u_h) \cdot \nabla_h v_h \, dx + \sum_{S \in \mathcal{S}_h^{(i)}} \Phi(\mathbf{n}; u_h^-, u_h^+) [[v_h]] \, ds + \int_{\partial\Omega} \Phi(\mathbf{n}; u_h, g) v_h \, ds = 0, \quad \forall v_h \in X_h$$

$$u_h(t=0) = \pi_h(u_0)$$

where  $\pi_h$  denotes the Lagrange interpolation operator on  $X_h$  and others notations has been introduced in the previous section.

For convenience, we introduce the discrete operator  $G_h$ , defined for all  $u_h, v_h \in X_h$  by

$$\int_{\Omega} G_h(u_h) v_h \, dx = - \int_{\Omega} \mathbf{f}(u_h) \cdot \nabla_h v_h \, dx + \sum_{S \in \mathcal{S}_h^{(i)}} \int_S \Phi(\mathbf{n}; u_h^-, u_h^+) [[v_h]] \, ds + \int_{\partial\Omega} \Phi(\mathbf{n}; u_h, g) v_h \, ds \quad (4.2)$$

For a given  $u_h \in X_h$ , we also define the linear form  $g_h$  as

$$g(v_h) = \int_{\Omega} G_h(u_h) v_h \, dx$$

As the matrix  $M$ , representing the  $L^2$  scalar product in  $X_h$ , is block-diagonal, it can be easily inverted at the element level, and for a given  $u_h \in X_h$ , we have  $G(u_h) = M^{-1}g_h$ . Then, the problem writes equivalently as a set of coupled nonlinear ordinary differential equations.

(P)<sub>h</sub>: find  $u_h \in C^1([0, T], X_h)$  such that

$$\frac{\partial u_h}{\partial t} + G_h(u_h) = 0$$

### Time discretization

Let  $\Delta t > 0$  be the time step. The previous nonlinear ordinary differential equations are discretized by using a specific explicit Runge-Kutta with intermediates states [42, 43, 104]. This specific variant of the usual Runge-Kutta scheme, called *strong stability preserving*, is suitable for avoiding possible spurious oscillations of the approximate solution when the exact solution has a poor regularity. Let  $u_h^n$  denotes the approximation of  $u_h$  at time  $t_n = n\Delta t$ ,  $n \geq 0$ . Then  $u_h^{n+1}$  is defined by recurrence:

$$\begin{aligned} u_h^{n,0} &= u_h^n \\ u_h^{n,i} &= \sum_{j=0}^{i-1} \alpha_{i,j} u_h^{n,j} - \Delta t \beta_{i,j} G_h(u_h^{n,j}), \quad 1 \leq i \leq p \\ u_h^{n+1} &= u_h^{n,p} \end{aligned}$$

where the coefficients satisfy  $\alpha_{i,j} \geq 0$  and  $\beta_{i,j} \geq 0$  for all  $1 \leq i \leq p$  and  $0 \leq j \leq i-1$ , and  $\sum_{j=0}^{i-1} \alpha_{i,j} = 1$  for all  $1 \leq i \leq p$ . Notice that when  $p = 1$  this scheme coincides with the usual

explicit Euler scheme. For a general  $p \geq 1$  order scheme, there are  $p - 1$  intermediate states  $u_h^{n,i}$ ,  $i = 1 \dots p - 1$ . Computation of the coefficients  $\alpha_{i,j}$  and  $\beta_{i,j}$  can be founded in [42, 43, 104] and are grouped in file ‘runge\_kutta\_ssp.icc’ of the examples directory for convenience.

### 4.1.3 Slope limiters

Slope limiters are required when the solution develops discontinuities: this is a very classical feature of most solutions of nonlinear hyperbolic equations. A preliminary version of the slope limiter proposed by Cocburn et al. [23, p. 208] is implemented in **Rheolef**: this preliminary version only supports the  $d = 1$  dimension and  $k = 1$  polynomial degree. Recall that the  $k = 0$  case do not need any limiter. More general implementation will support the  $d = 2, 3$  and  $k \geq 2$  cases in the future. The details of the limiter implementation is presented in this section: the impatient reader, who is interested by applications, can jump to the next section, devoted to the Burgers equation.

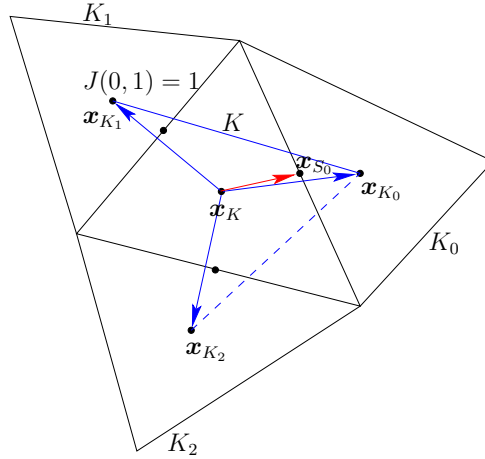


Figure 4.4: Limiter: the neighbors elements and the middle edge points.

Fig. 4.4 shows the  $d+1$  neighbor elements  $K_i$ ,  $i = 0 \dots d$  around an element  $d$ . Let  $S_i = \partial K \cap \partial K_i$ ,  $i = 0 \dots d$  be the  $i$ -th side of  $K$ . We denote by  $\mathbf{x}_K$ ,  $\mathbf{x}_{K_i}$  and  $\mathbf{x}_{S_i}$  the barycenters of these elements and sides,  $i = 0 \dots d$ . When  $d = 2$ , the barycenter  $\mathbf{x}_{S_i}$  of the edge belongs to the interior of a triangle  $(\mathbf{x}_K, \mathbf{x}_{K_i}, \mathbf{x}_{K_{J_{i,1}}})$  for exactly one of the two possible  $J_{i,1} \neq i$  and  $0 \leq J_{i,1} \leq d$ . When  $d = 3$ , the barycenter  $\mathbf{x}_{S_i}$  of the face belongs to the interior of a tetrahedron  $(\mathbf{x}_K, \mathbf{x}_{K_i}, \mathbf{x}_{K_{J_{i,1}}}, \mathbf{x}_{K_{J_{i,2}}})$  for exactly one pair  $(J_{i,1}, J_{i,2})$ , up to a permutation, of the three possible pairs  $J_{i,1}, J_{i,2} \neq i$  and  $0 \leq J_{i,1}, J_{i,2} \leq d$ . Let us denote  $J_{i,0} = i$ . Then, the vector  $\overrightarrow{\mathbf{x}_K \mathbf{x}_{S_i}}$  decompose on the basis  $(\overrightarrow{\mathbf{x}_K \mathbf{x}_{K_{J_{i,k}}}})_{0 \leq k \leq d-1}$  as

$$\overrightarrow{\mathbf{x}_K \mathbf{x}_{S_i}} = \sum_{k=0}^{d-1} \alpha_{i,k} \overrightarrow{\mathbf{x}_K \mathbf{x}_{K_{J_{i,k}}}} \quad (4.3)$$

where  $\alpha_{i,k} \geq 0$ ,  $k = 0 \dots d - 1$ . Let us consider now the patch  $\omega_K$  composed of  $K$  and its  $d$  neighbors:

$$\omega_K = K \cup K_0 \cup \dots \cup K_d$$

For any affine function  $\xi \in P_1(\omega_K)$  over this patch, let us denote

$$\begin{aligned} \delta_{K,i}(\xi) &= \sum_{k=0}^{d-1} \alpha_{i,k} \left( \xi(\mathbf{x}_{K_{J_{i,k}}}) - \xi(\mathbf{x}_K) \right), \quad i = 0 \dots d - 1 \\ &= \xi(\mathbf{x}_{S_i}) - \xi(\mathbf{x}_K) \quad \text{from (4.3)} \end{aligned}$$

In other terms,  $\delta_{K,i}(\xi)$  represents the departure of the value of  $\xi$  at  $\mathbf{x}_{S_i}$  from its average  $\xi(\mathbf{x}_K)$  on the element  $K$ .

Let now  $(\varphi_i)_{0 \leq i \leq d-1}$  denote the Lagrangian basis in  $K$  associated to the set of nodes  $(\mathbf{x}_{S_i})_{0 \leq i \leq d-1}$ :

$$\begin{aligned} \varphi_i(\mathbf{x}_{S_j}) &= \delta_{i,j}, \quad 0 \leq i, j \leq d-1 \\ \sum_{i=0}^{d-1} \varphi_i(\mathbf{x}) &= 1, \quad \forall \mathbf{x} \in K \end{aligned}$$

The affine function  $\xi \in P_1(\omega_K)$  expresses on this basis as

$$\xi(\mathbf{x}) = \xi(\mathbf{x}_K) + \sum_{i=0}^{d-1} \delta_{K,i}(\xi) \varphi_i(\mathbf{x}), \quad \forall \mathbf{x} \in K$$

Let now  $u_h \in \mathbb{P}_{1d}(\mathcal{T}_h)$ . On any element  $K \in \mathcal{T}_h$ , let us introduce its average value:

$$\bar{u}_K = \frac{1}{\text{meas}(K)} \int_K u_h(\mathbf{x}) \, dx$$

and its departure from its average value:

$$\tilde{u}_K(\mathbf{x}) = u_h|_K(\mathbf{x}) - \bar{u}_K, \quad \forall \mathbf{x} \in K$$

Notice that  $u_h \notin P_1(\omega_K)$ . Let us extend  $\delta_{K,i}$  to  $u_h$  as

$$\delta_{K,i}(u_h) = \sum_{k=0}^{d-1} \alpha_{i,k} (\bar{u}_{K_{J_{i,k}}} - \bar{u}_K), \quad i = 0 \dots d-1$$

Since  $u_h \notin P_1(\omega_K)$ , we have  $\tilde{u}_K(\mathbf{x}_{K_{J_{i,k}}}) \neq \delta_{K,i}(u_h)$  in general. The idea is then to capture oscillations by controlling the departure of the values  $\tilde{u}_K(\mathbf{x}_{K_{J_{i,k}}})$  from the values  $\delta_{K,i}(u_h)$ . Thus, associate to  $u_h \in \mathbb{P}_{1d}(\mathcal{T}_h)$  the quantities

$$\Delta_{K,i}(u_h) = \min_{\text{mod}_{\text{TVB}}} (\tilde{u}_K(\mathbf{x}_{K_{J_{i,k}}}), \theta \delta_{K,i}(u_h))$$

for all  $i = 0 \dots d-1$  and where  $\theta \geq 1$  is a parameter of the limiter and

$$\min_{\text{mod}_{\text{TVB}}}(a, b) = \begin{cases} a & \text{when } |a| \leq Mh^2 \\ \min_{\text{mod}}(a, b) & \text{otherwise} \end{cases}$$

where  $M > 0$  is a tunable parameter which can be evaluated from the curvature of the initial datum at its extrema by setting

$$M = \sup_{\mathbf{x} \in \Omega, \nabla u_0(\mathbf{x})=0} |\nabla \otimes \nabla u_0| \quad (4.4)$$

Introduced in [103], the basic idea is to deactivate the limiter when space derivatives are of order  $h^2$ . This improves the limiter behavior near smooth local extrema. The minmod function is defined by

$$\min_{\text{mod}}(a, b) = \begin{cases} \text{sgn}(a) \min(|a|, |b|) & \text{when } \text{sgn}(a) = \text{sgn}(b) \\ 0 & \text{otherwise} \end{cases}$$

Then, for all  $i = 0 \dots d-1$  we define

$$r_K(u_h) = \frac{\sum_{j=0}^{d-1} \max(0, -\Delta_{K,j}(u_h))}{\sum_{j=0}^{d-1} \max(0, \Delta_{K,j}(u_h))} \geq 0$$

$$\begin{aligned} \hat{\Delta}_{K,i}(u_h) &= \min(1, r_K(u_h)) \max(0, \Delta_{K,i}(u_h)) \\ &\quad - \min(1, 1/r_K(u_h)) \max(0, -\Delta_{K,i}(u_h)), \quad i = 0 \dots d-1 \text{ when } r_K(u_h) \neq 0 \end{aligned}$$

Finally, the limited function  $\Lambda_h(u_h)$  is defined element by element for all element  $K \in \mathcal{T}_h$  for all  $\mathbf{x} \in K$  by

$$\Lambda_h(u_h)|_K(\mathbf{x}) = \begin{cases} \bar{u}_K + \sum_{i=1}^{d-1} \Delta_{K,i}(u_h) \varphi_i(\mathbf{x}) & \text{when } r_K(u_h) = 0 \\ \bar{u}_K + \sum_{i=1}^{d-1} \hat{\Delta}_{K,i}(u_h) \varphi_i(\mathbf{x}) & \text{otherwise} \end{cases}$$

Notice that there are two types of computations involved in the limiter: one part is independent of  $u_h$  and depends only upon the mesh:  $J_{i,k}$  and  $\alpha_{i,k}$  on each element. It can be computed one time for all. The other part depends upon the values of  $u_h$ .

Notice that the limiter preserves the average value of  $u_h$  on each element  $K$  and also the functions that are globally affine on the patch  $\omega_K$ . Also we have, inside each element  $K$  and for all side index  $i = 0 \dots d-1$ :

$$|\Lambda_h(u_h)|_K(\mathbf{x}_{S_i}) - \bar{u}_K| \leq \max(|\Delta_{K,i}(u_h)|, |\hat{\Delta}_{K,i}(u_h)|) \leq |\Delta_{K,i}(u_h)| \leq |u_h|_K(\mathbf{x}_{S_i}) - \bar{u}_K|$$

It means that, inside each element, the gradient of the  $P_1$  limited function is no larger than that of the original one.

The limiter on an element close to the boundary should takes into account the inflow condition. In [25], the modifications are described.

#### 4.1.4 Example: the Burgers equation

As an illustration, let us consider now the test with the one-dimensional ( $d = 1$ ) Burgers equation for a propagating slant step (see e.g. [19, p. 87]) in  $\Omega = ]0, 1[$ . We have  $\mathbf{f}(u) = u^2/2$ , for all  $u \in \mathbb{R}$ . In that case, the Godunov flux (4.1d), introduced page 145, can be computed explicitly for any  $\nu = (\nu_0) \in \mathbb{R}^d$  and  $a, b \in \mathbb{R}$ :

$$\Phi(\nu; a, b) = \begin{cases} \nu_0 \min(a^2, b^2) / 2 & \text{when } \nu_0 \geq 0 \text{ and } a \leq b \text{ or } \nu_0 \leq 0 \text{ and } a \geq b \\ \nu_0 \max(a^2, b^2) / 2 & \text{otherwise} \end{cases}$$

Example file 4.2: burgers.icc

```
1 point f (const Float& u) { return point (sqr(u)/2); }
```

Example file 4.3: burgers\_flux\_godunov.icc

```
1 Float phi (const point& nu, Float a, Float b) {
2   if ((nu[0] >= 0 && a <= b) || (nu[0] <= 0 && a >= b))
3     return nu[0]*min(sqr(a),sqr(b))/2;
4   else
5     return nu[0]*max(sqr(a),sqr(b))/2;
6 }
```

#### Computing an exact solution

An exact solution is useful for testing numerical methods. The computation of such an exact solution for the one dimensional Burgers equation is described by Hartens *et al.* [45]. The authors consider first the problem with a periodic boundary condition:

(P): find  $u : ]0, T[ \times ]-1, 1[ \rightarrow \mathbb{R}$  such that

$$\begin{aligned} \frac{\partial u}{\partial t} + \frac{\partial}{\partial x} \left( \frac{u^2}{2} \right) &= 0 \text{ in } ]0, T[ \times ]-1, 1[ \\ u(t=0, x) &= \alpha + \beta \sin(\pi x + \gamma), \text{ a.e. } x \in ]-1, 1[ \\ u(t, x=-1) &= u(t, x=1) \text{ a.e. } t \in ]0, T[ \end{aligned}$$

where  $\alpha, \beta$  and  $\gamma$  are real parameters. Let us denote  $w$  the solution of the problem when  $\beta = 1$  and  $\alpha = \gamma = 0$ ; i.e. with the initial condition  $w(t=0, x) = \sin(\pi x)$ , a.e.  $x \in ]-1, 1[$ . For any  $x \in [0, 1[$  and  $t > 0$ , the solution  $\bar{w} = w(t, x)$  satisfies the characteristic relation

$$\bar{w} = \sin(\pi(x - \bar{w}t))$$

This nonlinear relation can be solved by a Newton algorithm. Then, for  $x \in ]-1, 0[$ , the solution is completed by symmetry:  $w(t, x) = -w(t, -x)$ . Finally, the general solution for any  $\alpha, \beta$  and  $\gamma = 0$  writes  $u(t, x) = \alpha + w(\beta t, x - \alpha t + \gamma)$ . File ‘harten.icc’ implements this approach.

Example file 4.4: harten.icc

```

1 #include "harten0.icc"
2 struct harten {
3     Float operator() (const point& x) const {
4         Float x0 = x[0] - a*t + c;
5         Float shift = -2*floor((x0+1)/2);
6         Float xs = x0 + shift;
7         check_macro (xs >= -1 && xs <= 1, "invalid xs=" << xs);
8         return a + b*h0 (point (xs));
9     }
10    harten (Float t1=0, Float a1=1, Float b1=0.5, Float c1=0):
11        h0(b1*t1, t(t1), a(a1), b(b1), c(c1)) {}
12    Float M() const { Float pi = acos(-1.0); return sqr(pi)*b; }
13    Float min() const { return a-b; }
14    Float max() const { return a+b; }
15 protected:
16    harten0 h0;
17    Float t, a, b, c;
18 };
19 using u_init = harten;
20 using g = harten;

```

Example file 4.5: harten\_show.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "harten.icc"
5 int main(int argc, char**argv) {
6     environment rheolef (argc, argv);
7     geo omega (argv[1]);
8     space Xh (omega, argv[2]);
9     size_t nmax = (argc > 3) ? atoi(argv[3]) : 1000;
10    Float tf = (argc > 4) ? atof(argv[4]) : 2.5;
11    Float a = (argc > 5) ? atof(argv[5]) : 1;
12    Float b = (argc > 6) ? atof(argv[6]) : 0.5;
13    Float c = (argc > 7) ? atof(argv[7]) : 0;
14    branch even("t", "u");
15    for (size_t n = 0; n <= nmax; ++n) {
16        Float t = n*tf/nmax;
17        field pi_h_u = interpolate (Xh, harten(t, a, b, c));
18        dout << even(t, pi_h_u);
19    }
20 }

```

The included file ‘harten0.icc’ is not shown here but is available in the example directory.

### Comments

Notice that the constant  $M$ , used by the limiter in (4.4), can be explicitly computed for this solution:  $M = \beta\pi^2$ .

The animation of this exact solution is performed by the following commands:

```

make harten_show
mkgeo_grid -e 2000 -a -1 -b 1 > line2.geo

```

```
./harten_show line2 P1 1000 2.5 > line2-exact.branch
branch line2-exact -gnuplot
```

Fig. 4.5 shows the solution  $u$  for  $\alpha = 1, \beta = 1/2$  and  $\gamma = 0$ . It is regular until  $t = 2/\pi$  (Fig. 4.5.c) and then develops a shock for  $t > 2/\pi$  (Fig. 4.5.d). After its apparition, this shock interacts with the expansion wave in  $] -1, 1[$ : this brings about a fast decay of the solution (Figs. 4.5.e and f). Fig. 4.5 plots also a numerical solution: its computation is the aim of the next section.

## Numerical resolution

When replacing the periodic boundary condition with a inflow one, associated with the boundary data  $g$ , we choose  $g$  to be the value of the exact solution of the problem with periodic boundary conditions:  $g(t, x) = \alpha + w(\beta t, x - \alpha t)$  for  $x \in \{-1, 1\}$ .

Example file 4.6: burgers\_dg.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "harten.icc"
5 #include "burgers.icc"
6 #include "burgers_flux_godunov.icc"
7 #include "runge_kutta_ssp.icc"
8 int main(int argc, char**argv) {
9     environment rheolef (argc, argv);
10    geo omega (argv[1]);
11    space Xh (omega, argv[2]);
12    Float cfl = 1;
13    limiter_option_type lopt;
14    size_t nmax = (argc > 3) ? atoi(argv[3]) : numeric_limits<size_t>::max();
15    Float tf = (argc > 4) ? atof(argv[4]) : 2.5;
16    size_t p = (argc > 5) ? atoi(argv[5]) : pmax;
17    lopt.M = (argc > 6) ? atoi(argv[6]) : u_init().M();
18    if (nmax == numeric_limits<size_t>::max()) {
19        nmax = floor(1+tf/(cfl*omega.hmin()));
20    }
21    Float delta_t = tf/nmax;
22    integrate_option foft;
23    foft.invert = true;
24    trial u (Xh); test v (Xh);
25    form inv_m = integrate (u*v, foft);
26    vector<field> uh(p+1, field(Xh,0));
27    uh[0] = interpolate (Xh, u_init());
28    branch even("t","u");
29    dout << catchmark("delta_t") << delta_t << endl
30        << even(0,uh[0]);
31    for (size_t n = 1; n <= nmax; ++n) {
32        for (size_t i = 1; i <= p; ++i) {
33            uh[i] = 0;
34            for (size_t j = 0; j < i; ++j) {
35                field lh =
36                    - integrate (dot(compose(f,uh[j]),grad_h(v)))
37                    + integrate ("internal_sides",
38                        compose (phi, normal(), inner(uh[j]), outer(uh[j]))*jump(v))
39                    + integrate ("boundary",
40                        compose (phi, normal(), uh[j], g(n*delta_t))*v);
41                uh[i] += alpha[p][i][j]*uh[j] - delta_t*beta[p][i][j]*(inv_m*lh);
42            }
43            uh[i] = limiter(uh[i], g(n*delta_t)(point(-1)), lopt);
44        }
45        uh[0] = uh[p];
46        dout << even(n*delta_t,uh[0]);
47    }
48 }
```

### Comments

The Runge-Kutta time discretization combined with the discontinuous Galerkin space discretization is implemented for this test case.

The  $P_0$  approximation is performed by the following commands:

```
make burgers_dg
mkgeo_grid -e 200 -a -1 -b 1 > line2-200.geo
./burgers_dg line2-200.geo P0 1000 2.5 > line2-200-P0.branch
branch line2-200-P0.branch -gnuplot
```

The two last commands compute the  $P_0$  approximation of the solution, as shown on Fig. 4.5. Observe the robust behavior of the solution at the vicinity of the chock. By replacing P0 by P1d in the previous commands, we obtain the  $P_1$ -discontinuous approximation.

```
./burgers_dg line2-200.geo P1d 1000 2.5 > line2-200-P1d.branch
branch line2-200-P1d.branch -gnuplot
```

Fig. 4.6 plots the error vs  $h$  for  $k = 0$  and  $k = 1$ . Fig. 4.6.a plots in a time interval  $[0, T]$  with  $T = 1/\pi$ , before the chock that occurs at  $t = 2/\pi$ . In that interval, the solution is regular and the error approximation behaves as  $\mathcal{O}(h^{k+1})$ . The time interval has been chosen sufficiently small for the error to depend only upon  $h$ . Fig. 4.6.b plots in a larger time interval  $[0, T]$  with  $T = 5/2$ , that includes the chock. Observe that the error behaves as  $\mathcal{O}(h)$  for both  $k = 0$  and 1. This is optimal when  $k = 0$  but not when  $k = 1$ . This is due to the loss of regularity of the exact solution that presents a discontinuity; A similar observation can be found in [114], table 4.1.



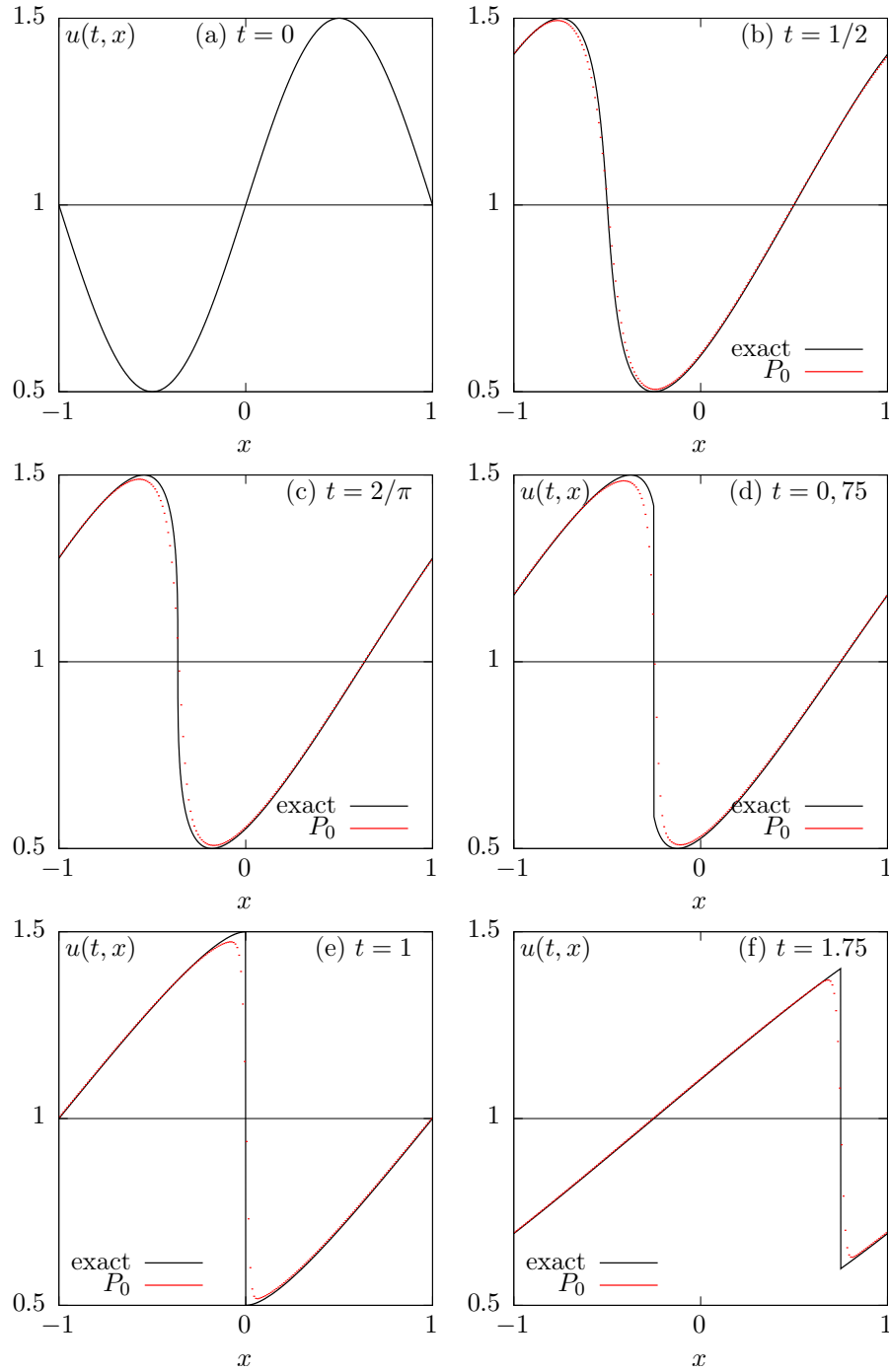


Figure 4.5: Harten's exact solution of the Burgers equation ( $\alpha = 1, \beta = 1/2, \gamma = 0$ ). Comparison with the  $P_0$  approximation ( $h = 1/100$ , RK-SSP(3)).

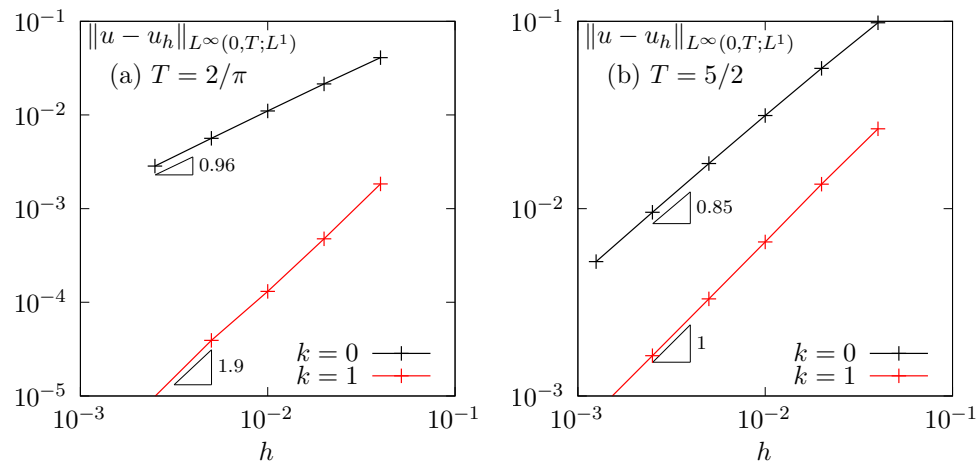


Figure 4.6: Burgers equation: error between the  $P_0$  approximation and the exact solution of the Harten's problem ( $\alpha = 1, \beta = 1/2, \gamma = 0$ ): (a) before chock, with  $T = 1/\pi$ ; (b) after chock, with  $T = 5/2$ .

## 4.2 Scalar second-order problems

### 4.2.1 The Poisson problem with Dirichlet boundary conditions

The Poisson problem with non-homogeneous Dirichlet boundary conditions has been already introduced in volume 1, section 1.1.12, page 22:

(P): find  $u$ , defined in  $\Omega$ , such that

$$\begin{aligned} -\Delta u &= f \text{ in } \Omega \\ u &= g \text{ on } \partial\Omega \end{aligned}$$

where  $f$  and  $g$  are given.

The *discontinuous* finite element space is defined by:

$$X_h = \{v_h \in L^2(\Omega); v_h|_K \in P_k, \forall K \in \mathcal{T}_h\}$$

where  $k \geq 1$  is the polynomial degree. As elements of  $X_h$  do not belong to  $H^1(\Omega)$ , due to discontinuities at inter-elements, we introduce the broken Sobolev space:

$$H^1(\mathcal{T}_h) = \{v \in L^2(\Omega); v|_K \in H^1(K), \forall K \in \mathcal{T}_h\}$$

such that  $X_h \subset H^1(\mathcal{T}_h)$ . We introduce the following bilinear form  $a_h(.,.)$  and linear form  $l_h(.,.)$ , defined for all  $u, v \in H^1(\mathcal{T}_h)$  by (see e.g. [31, p. 125 and 127], eqn. (4.12)):

$$\begin{aligned} a_h(u, v) &= \int_{\Omega} \nabla_h u \cdot \nabla_h v \, dx + \sum_{S \in \mathcal{S}_h} \int_S (\eta_s \llbracket u \rrbracket \llbracket v \rrbracket - \{\!\{ \nabla_h u \cdot \mathbf{n} \}\!\} \llbracket v \rrbracket - \llbracket u \rrbracket \{\!\{ \nabla_h v \cdot \mathbf{n} \}\!\}) \, ds \quad (4.5) \\ l_h(v) &= \int_{\Omega} f u \, dx + \int_{\partial\Omega} (g v - g \nabla_h v \cdot \mathbf{n}) \, ds \quad (4.6) \end{aligned}$$

The last term involves a sum over  $\mathcal{S}_h$ , the set of *all sides* of the mesh  $\mathcal{T}_h$ , i.e. the internal sides and the boundary sides. By convenience, the definition of the jump and average are extended to all boundary sides as  $\llbracket u \rrbracket = \{\!\{ u \}\!\} = u$ . Notice that, as for the previous transport problem, the Dirichlet boundary condition  $u = g$  is weakly imposed on  $\partial\Omega$  via the integrals on the boundary. Finally,  $\eta_s > 0$  is a stabilization parameter on a side  $S$ . The stabilization term associated to  $\eta_s$  is present in order to achieve coercivity: it penalize interface and boundary jumps. A common choice is  $\eta_s = \beta h_s^{-1}$  where  $\beta > 0$  is a constant and  $h_s$  is a local length scale associated to the current side  $S$ . One drawback to this choice is that it requires the end user to specify the numerical constant  $\beta$ . From one hand, if the value of this parameter is not sufficiently large, the form  $a_h(.,.)$  is not coercive and the approximate solution develops instabilities and do not converge [34]. From other hand, if the value of this parameter is too large, it affects the overall efficiency of the iterative solver of the linear system: the spectral condition number of the matrix associated to  $a_h(.,.)$  grows linearly with this parameter [20]. An explicit choice of penalty parameter is proposed in [101]:  $\eta_s = \beta \varpi_s$  where  $\beta = (k+1)(k+d)/d$  and

$$\varpi_s = \begin{cases} \frac{\text{meas}(\partial K)}{\text{meas}(K)} & \text{when } S = K \cap \partial\Omega \text{ is a boundary side} \\ \max\left(\frac{\text{meas}(\partial K_0)}{\text{meas}(K_0)}, \frac{\text{meas}(\partial K_1)}{\text{meas}(K_1)}\right) & \text{when } S = K_0 \cap K_1 \text{ is an internal side} \end{cases}$$

Notice that  $\varpi_s$  scales as  $h_s^{-1}$ . Now, the computation of the penalty parameter is fully automatic and the convergence of the method is always guaranteed to converge. Moreover, this choice has been founded to be sharp and it preserves the optimal efficiency of the iterative solvers. Finally, the discrete variational formulation writes:

(FV) $_h$ : find  $u_h \in X_h$  such that

$$a_h(u_h, v_h) = l_h(v_h), \forall v_h \in X_h$$

The following code implement this problem in the **Rheolef** environment.

Example file 4.7: dirichlet\_dg.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "cosinusprod_laplace.icc"
5 int main(int argc, char**argv) {
6     environment rheolef (argc, argv);
7     geo omega (argv[1]);
8     space Xh (omega, argv[2]);
9     size_t d = omega.dimension();
10    size_t k = Xh.degree();
11    Float beta = (k+1)*(k+d)/d;
12    trial u (Xh); test v (Xh);
13    form a = integrate (dot(grad_h(u),grad_h(v)))
14              + integrate ("sides", beta*penalty()*jump(u)*jump(v)
15                          - jump(u)*average(dot(grad_h(v),normal()))
16                          - jump(v)*average(dot(grad_h(u),normal())));
17    field lh = integrate (f(d)*v)
18              + integrate ("boundary", beta*penalty()*g(d)*v
19                          - g(d)*dot(grad_h(v),normal()));
20    solver sa (a.uu());
21    field uh(Xh);
22    uh.set_u() = sa.solve(lh.u());
23    dout << uh;
24 }

```

### Comments

The `penalty()` pseudo-function implements the computation of  $\varpi_s$  in **Rheolef**. The right-hand side  $f$  and  $g$  are given by (1.4), volume 1, page 23. In that case, the exact solution is known. Running the one-dimensional case writes:

```

make dirichlet_dg
mkgeo_grid -e 10 > line.geo
./dirichlet_dg line P1d | field -
./dirichlet_dg line P2d | field -

```

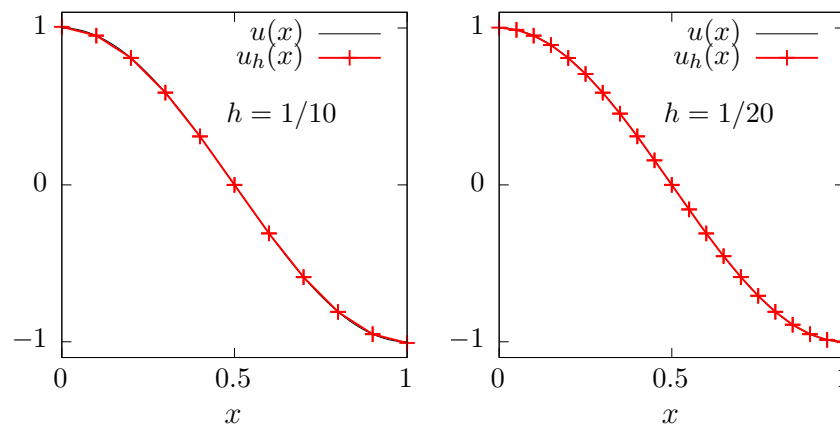


Figure 4.7: The discontinuous Galerkin method for the Poisson problem when  $k = 1$  and  $d = 1$ .

Fig. 4.7 plots the one-dimensional solution when  $k = 1$  for two meshes. Observe that the jumps at inter-element nodes decreases very fast with mesh refinement and are no more perceptible on the

plots. Recall that the Dirichlet boundary conditions at  $x = 0$  and  $x = 1$  is only weakly imposed: the corresponding jump at the boundary is also not perceptible.

The two-dimensional case writes:

```
mkgeo_grid -t 10 > square.geo
./dirichlet_dg square P1d | field -elevation -
./dirichlet_dg square P2d | field -elevation -
```

and the three-dimensional one

```
mkgeo_grid -T 10 > cube.geo
./dirichlet_dg cube P1d | field -
./dirichlet_dg cube P2d | field -
```

### Error analysis

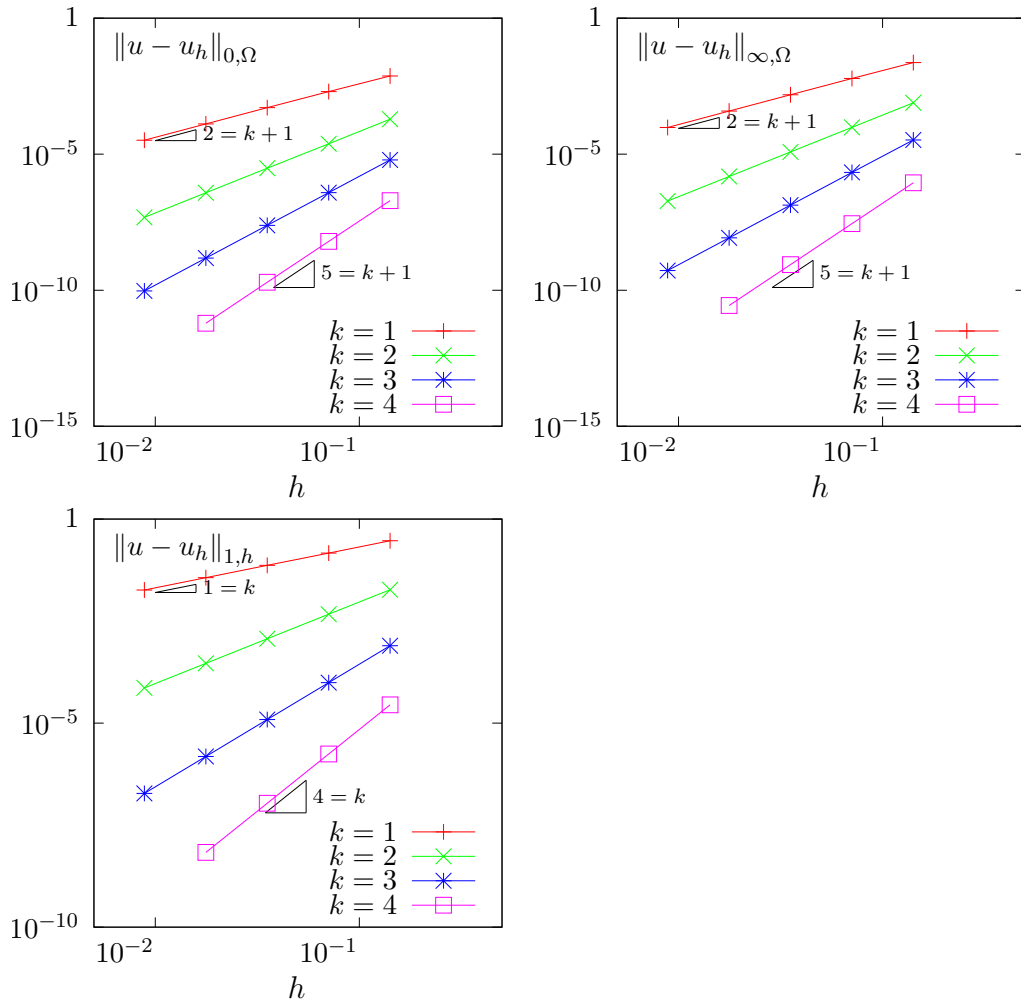


Figure 4.8: The discontinuous Galerkin method for the Poisson problem: convergence when  $d = 2$ .

The space  $H^1(\mathcal{T}_h)$  is equipped with the norm  $\|\cdot\|_{1,h}$ , defined for all  $v \in H^1(\mathcal{T}_h)$  by [31, p. 128]:

$$\|v\|_{1,h}^2 = \|\nabla_h v\|_{0,\Omega}^2 + \sum_{S \in \mathcal{S}_h} \int_S h_s^{-1} \llbracket v \rrbracket^2 ds$$

The code `cosinusprod_error_dg.cc` compute the error in these norms. This code it is not listed here but is available in the **Rheolef** example directory. The computation of the error writes:

```
make cosinusprod_error_dg
./dirichlet_dg_square P1d | cosinusprod_error_dg
```

Fig. 4.8 plots the error  $u - u_h$  in  $L^2$ ,  $L^\infty$  and the  $\|\cdot\|_{1,h}$  norms. The  $L^2$  and  $L^\infty$  error norms behave as  $\mathcal{O}(h^{k+1})$  for all  $k \geq 0$ , while the  $\|\cdot\|_{1,h}$  one behaves as  $\mathcal{O}(h^k)$ , which is optimal.

### 4.2.2 The Helmholtz problem with Neumann boundary conditions

The Poisson problem with non-homogeneous Neumann boundary conditions has been already introduced in volume 1, section 1.2, page 27:

(P): find  $u$ , defined in  $\Omega$ , such that

$$\begin{aligned} u - \Delta u &= f \text{ in } \Omega \\ \frac{\partial u}{\partial n} &= g \text{ on } \partial\Omega \end{aligned}$$

where  $f$  and  $g$  are given. We introduce the following bilinear form  $a_h(\cdot, \cdot)$  and linear for  $l_h(\cdot)$ , defined for all  $u, v \in H^1(\mathcal{T}_h)$  by (see e.g. [31, p. 127], eqn. (4.16)):

$$a_h(u, v) = \int_{\Omega} (uv + \nabla_h u \cdot \nabla_h v) \, dx \quad (4.7)$$

$$+ \sum_{S \in \mathcal{S}_h^{(i)}} \int_S (\beta \varpi_s [u][v] - \{\{\nabla_h u \cdot \mathbf{n}\}\} [v] - [u] \{\{\nabla_h v \cdot \mathbf{n}\}\}) \, ds \quad (4.8)$$

$$l_h(v) = \int_{\Omega} f u \, dx + \int_{\partial\Omega} g v \, ds \quad (4.9)$$

Let us comment the changes between these forms and those used for the Poisson problem with Dirichlet boundary conditions. The Poisson operator  $-\Delta$  has been replaced by the Helmholtz one  $I - \Delta$  in order to have an unique solution. Remark also that the sum is performed in (4.5) for all internal sides in  $\mathcal{S}_h^{(i)}$ , while, in (4.5), for Dirichlet boundary conditions, it was for all sides in  $\mathcal{S}_h$ , i.e. for both boundary and internal sides. Also, the right-hand-side linear form  $l_h(\cdot)$  do no more involves any sum over sides.

Finally, the discrete variational formulation writes:

(FV)<sub>h</sub>: find  $u_h \in X_h$  such that

$$a_h(u_h, v_h) = l_h(v_h), \quad \forall v_h \in X_h$$

The following code implement this problem in the **Rheolef** environment.

Example file 4.8: neumann\_dg.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "sinusprod_helmholtz.icc"
5 int main(int argc, char**argv) {
6     environment rheolef (argc, argv);
7     geo omega (argv[1]);
8     space Xh (omega, argv[2]);
9     size_t d = omega.dimension();
10    size_t k = Xh.degree();
11    Float beta = (k+1)*(k+d)/d;
12    trial u (Xh); test v (Xh);
13    form a = integrate (u*v + dot(grad_h(u),grad_h(v)))
14            + integrate ("internal_sides",
15                        beta*penalty()*jump(u)*jump(v)
16                        - jump(u)*average(dot(grad_h(v),normal()))
17                        - jump(v)*average(dot(grad_h(u),normal())));
18    field lh = integrate (f(d)*v) + integrate ("boundary", g(d)*v);
19    solver sa (a.uu());
20    field uh(Xh);
21    uh.set_u() = sa.solve(lh.u());
22    dout << uh;
23 }

```

### Comments

The right-hand side  $f$  and  $g$  are given by (1.5), volume 1, page 23. In that case, the exact solution is known. Running the program is obtained from the non-homogeneous Dirichlet case, by replacing `dirichlet_dg` by `neumann_dg`.

### 4.2.3 Nonlinear scalar hyperbolic problems with diffusion

A time-dependent nonlinear second order problem with nonlinear first order dominated terms problem writes:

(P): find  $u$ , defined in  $]0, T[ \times \Omega$ , such that

$$\frac{\partial u}{\partial t} + \operatorname{div} \mathbf{f}(u) - \varepsilon \Delta u = 0 \quad \text{in } ]0, T[ \times \Omega \quad (4.10a)$$

$$u(t=0) = u_0 \quad \text{in } \Omega \quad (4.10b)$$

$$u = g \quad \text{on } ]0, T[ \times \partial\Omega \quad (4.10c)$$

where  $\varepsilon > 0$ ,  $T > 0$ ,  $\Omega \subset \mathbb{R}^d$ ,  $d = 1, 2, 3$  and the initial condition  $\mathbf{u}_0$  being known. The function  $\mathbf{f} : \mathbb{R} \rightarrow \mathbb{R}^d$  is also known and supposed to be continuously differentiable. The initial data  $u_0$ , defined in  $\Omega$ , and the boundary one,  $g$ , defined on  $\partial\Omega$  are given.

Comparing (4.10a)-(4.10c) with the non-diffusive case (4.1a)-(4.1c) page 144, the second order term has been added in (4.10a) and the upstream boundary condition has been replaced by a Dirichlet one (4.10c).

### 4.2.4 Example: the Burgers equation with diffusion

#### Problem statement and its exact solution

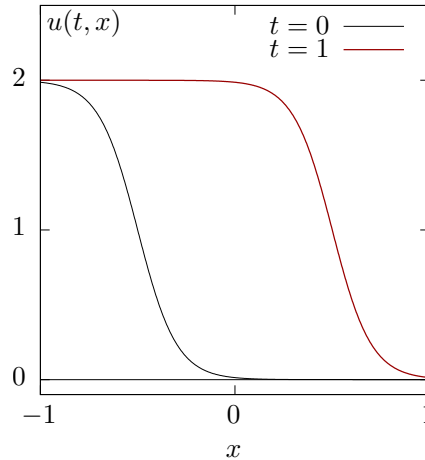


Figure 4.9: An exact solution for the Burgers equation with diffusion ( $\varepsilon = 10^{-1}$ ,  $x_0 = -1/2$ ).

Our model problem in this chapter is the one-dimensional Burgers equation. It was introduced in section 4.1.4, page 148 with the choice  $\mathbf{f}(u) = u^2/2$ , for all  $u \in \mathbb{R}$ . Equation (4.10a) admits an exact solution

$$u(t, x) = 1 - \tanh\left(\frac{x - x_0 - t}{2\varepsilon}\right) \quad (4.11)$$



Example file 4.9: burgers\_diffusion\_exact.icc

```

1 struct u_exact {
2   Float operator() (const point& x) const {
3     return 1 - tanh((x[0]-x0-t)/(2*epsilon)); }
4   u_exact (Float e1, Float t1=0) : epsilon(e1), t(t1), x0(-0.5) {}
5   Float M() const { return 0; }
6   Float epsilon, t, x0;
7 };
8 using u_init = u_exact;
9 using g = u_exact;

```

The solution is represented on Fig. 4.9. Here  $x_0$  represents the position of the front at  $t = 0$  and  $\varepsilon$  is a characteristic width of the front. The initial and boundary conditions are chosen such that  $u(t, x)$  is the solution of (4.10a)-(4.10c).

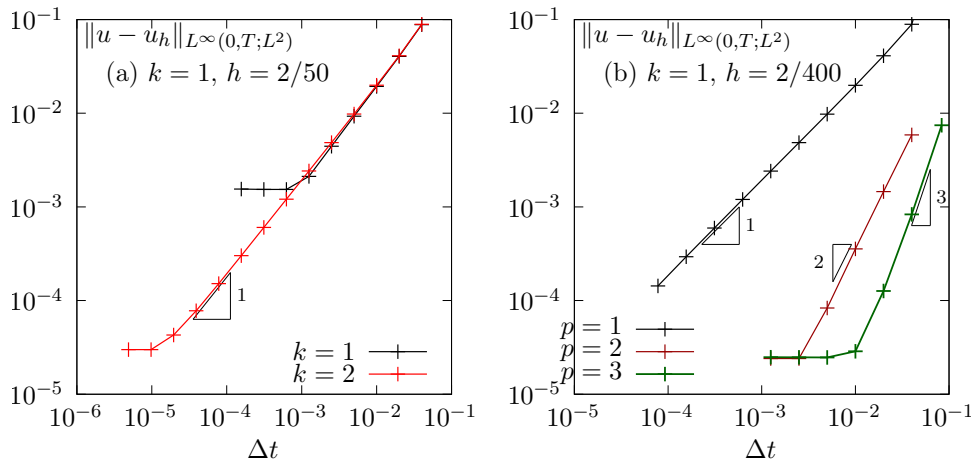


Figure 4.10: Convergence of the first order semi-implicit scheme for the Burgers equation with diffusion ( $\varepsilon = 0.1$ ,  $T = 1$ ). (a) first order semi-implicit scheme ; (b) Runge-Kutta semi-implicit scheme with  $p = 3$ .

Fig. 4.10.a plots the error versus  $\Delta t$  for the semi-implicit scheme when  $k = 1$  and  $2$ , and for  $h = 2/50$ . The time step for which the error becomes independent upon  $\Delta t$  and depends only upon  $h$  is of about  $\Delta t = 10^{-3}$  when  $k = 1$  and of about  $10^{-5}$  when  $k = 2$ . This approach is clearly inefficient for high order polynomial  $k$  and a higher order time scheme is required.

Fig. 4.10.b plots the error versus  $\Delta t$  for the Runge-Kutta semi-implicit scheme with  $p = 3$ ,  $k = 1$  and  $h = 2/200$ . The scheme is clearly only first-order, which is still unexpected. More work is required...

### Space discretization

The *discontinuous* finite element space is defined by:

$$X_h = \{v_h \in L^2(\Omega); v_h|_K \in P_k, \forall K \in \mathcal{T}_h\}$$

where  $k \geq 1$  is the polynomial degree. As elements of  $X_h$  do not belong to  $H^1(\Omega)$ , due to discontinuities at inter-elements, we introduce the broken Sobolev space:

$$H^1(\mathcal{T}_h) = \{v \in L^2(\Omega); v|_K \in H^1(K), \forall K \in \mathcal{T}_h\}$$

such that  $X_h \subset H^1(\mathcal{T}_h)$ . As for the Dirichlet problem, introduce the following bilinear form  $a_h(\cdot, \cdot)$  and linear for  $l_h(\cdot)$ , defined for all  $u, v \in H^1(\mathcal{T}_h)$  by (see e.g. [31, p. 125 and 127], eqn. (4.12)):

$$\begin{aligned} a_h(u, v) &= \int_{\Omega} \nabla_h u \cdot \nabla_h v \, dx + \sum_{S \in \mathcal{S}_h} \int_S (\eta_s \llbracket u \rrbracket \llbracket v \rrbracket - \{ \nabla_h u \cdot \mathbf{n} \} \llbracket v \rrbracket - \llbracket u \rrbracket \{ \nabla_h v \cdot \mathbf{n} \}) \, ds \\ l_h(v) &= \int_{\partial\Omega} (\eta_s g v - g \nabla_h v \cdot \mathbf{n}) \, ds \end{aligned} \quad (4.13)$$

The semi-discrete problem writes in variational form [31, p. 100]:

(P)<sub>h</sub>: find  $u_h \in C^1([0, T], X_h)$  such that

$$\begin{aligned} \int_{\Omega} \frac{\partial u_h}{\partial t} v_h \, dx + \int_{\Omega} G_h(u_h) v_h \, dx + \varepsilon a_h(u_h, v_h) &= \varepsilon l_h(v_h), \quad \forall v_h \in X_h \\ u_h(t=0) &= \pi_h(u_0) \end{aligned}$$

where  $G_h$  has been introduced in (4.2), page 145.

### Time discretization

Explicit Runge-Kutta scheme is possible for this problem but it leads to an excessive Courant-Friedrichs-Levy condition for the time step  $\Delta t$ , that is required to be lower than an upper bound that varies in  $\mathcal{O}(h^2)$ . The idea here is to continue to explicit the first order nonlinear terms and implicit the linear second order terms. Semi-implicit second order Runge-Kutta scheme was first introduced in 1997 by Ascher, Ruuth and Spiteri [8] and then extended in 2001 to third and fourth order by Calvo, de Frutos and Novo [18]. In 2015, Wang, Shu and Zhang [109, 110] applied it in the context of the discontinuous Galerkin method. The finite dimensional problem can be rewritten as

(P)<sub>h</sub>: find  $u_h \in C^1([0, T], X_h)$  such that

$$\begin{aligned} \frac{\partial u_h}{\partial t} + G_h(t, u_h) + A_h(t, u_h) &= 0, \quad \forall t \in ]0, T[ \\ u_h(t=0) &= \pi_h(u_0) \end{aligned}$$

where  $G_h$  has been introduced in (4.2), page 145 and  $A_h$  denotes the diffusive term. The semi-implicit Runge-Kutta scheme with  $p \geq 0$  intermediate steps writes at time step  $t_n$ :

$$u_h^{n,0} = u_h^n \quad (4.14a)$$

$$u_h^{n,i} = u_h^n - \Delta t \sum_{j=1}^i \alpha_{i,j} A_h(t_{n,j}, u_h^{n,j}) - \Delta t \sum_{j=0}^{i-1} \tilde{\alpha}_{i,j} G_h(t_{n,j}, u_h^{n,j}), \quad i = 1, \dots, p \quad (4.14b)$$

$$u_h^{n,p} = u_h^n - \Delta t \sum_{i=1}^p \beta_i A_h(t_{n,i}, u_h^{n,i}) - \Delta t \sum_{i=0}^p \tilde{\beta}_i G_h(t_{n,i}, u_h^{n,i}) \quad (4.14c)$$

where  $(u_h^{n,i})_{1 \leq i \leq p}$  are the  $p \geq 1$  intermediate states,  $t_{n,i} = t_n + \gamma_i \Delta t$ ,  $\gamma_i = \sum_{j=1}^i \alpha_{i,j} = \sum_{j=0}^{i-1} \tilde{\alpha}_{i,j}$ , and  $(\alpha_{i,j})_{0 \leq i, j \leq p}$ ,  $(\tilde{\alpha}_{i,j})_{0 \leq i, j \leq p}$ ,  $(\beta_i)_{0 \leq i \leq p}$  and  $(\tilde{\beta}_i)_{0 \leq i \leq p}$  are the coefficients of the scheme [8, 18, 110]. At each time step, have to solve  $p$  linear systems. From (4.14b) we get for all  $i = 1, \dots, p$ :

$$(I + \Delta t \alpha_{i,i} A_h(t_{n,i})) u_h^{n,i} = u_h^n - \Delta t \sum_{j=1}^{i-1} \alpha_{i,j} A_h(t_{n,j}, u_h^{n,j}) - \Delta t \sum_{j=0}^{i-1} \tilde{\alpha}_{i,j} G_h(t_{n,j}, u_h^{n,j})$$

Notice that when the matrix coefficients of  $A_h(t, \cdot)$  are independent of  $t$ , the matrix involved on the right-hand-side of the previous equation can be factored one time for all.

Example file 4.10: burgers\_diffusion\_dg.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "burgers.icc"
5 #include "burgers_flux_godunov.icc"
6 #include "runge_kutta_semiimplicit.icc"
7 #include "burgers_diffusion_exact.icc"
8 #undef NEUMANN
9 #include "burgers_diffusion_operators.icc"
10 int main(int argc, char**argv) {
11     environment rheolef (argc, argv);
12     geo omega (argv[1]);
13     space Xh (omega, argv[2]);
14     size_t k = Xh.degree();
15     Float epsilon = (argc > 3) ? atof(argv[3]) : 0.1;
16     size_t nmax = (argc > 4) ? atoi(argv[4]) : 500;
17     Float tf = (argc > 5) ? atof(argv[5]) : 1;
18     size_t p = (argc > 6) ? atoi(argv[6]) : min(k+1, rk::pmax);
19     Float delta_t = tf/nmax;
20     size_t d = omega.dimension();
21     Float beta = (k+1)*(k+d)/d;
22     trial u (Xh); test v (Xh);
23     form m = integrate (u*v);
24     integrate_option fopt;
25     fopt.invert = true;
26     form inv_m = integrate (u*v, fopt);
27     form a = epsilon*(
28         integrate (dot(grad_h(u), grad_h(v)))
29 #ifdef NEUMANN
30         + integrate ("internal_sides",
31 #else // NEUMANN
32         + integrate ("sides",
33 #endif // NEUMANN
34         beta*penalty()*jump(u)*jump(v)
35         - jump(u)*average(dot(grad_h(v), normal()))
36         - jump(v)*average(dot(grad_h(u), normal())));
37     vector<solver> sc (p+1);
38     for (size_t i = 1; i <= p; ++i) {
39         form ci = m + delta_t*rk::alpha[p][i][i]*a;
40         sc[i] = solver(ci.uu());
41     }
42     vector<field> uh(p+1, field(Xh,0));
43     uh[0] = interpolate (Xh, u_init(epsilon));
44     branch even("t", "u");
45     dout << catchmark("epsilon") << epsilon << endl
46         << even(0, uh[0]);
47     for (size_t n = 0; n < nmax; ++n) {
48         Float tn = n*delta_t;
49         Float t = tn + delta_t;
50         field uh_next = uh[0] - delta_t*rk::tilde_beta[p][0]*(inv_m*gh(epsilon, tn, uh[0], v));
51         for (size_t i = 1; i <= p; ++i) {
52             Float ti = tn + rk::gamma[p][i]*delta_t;
53             field rhs = m*uh[0] - delta_t*rk::tilde_alpha[p][i][0]*gh(epsilon, tn, uh[0], v);
54             for (size_t j = 1; j <= i-1; ++j) {
55                 Float tj = tn + rk::gamma[p][j]*delta_t;
56                 rhs -= delta_t*( rk::alpha[p][i][j]*(a*uh[j] - lh(epsilon, tj, v))
57                     + rk::tilde_alpha[p][i][j]*gh(epsilon, tj, uh[j], v));
58             }
59             rhs += delta_t*rk::alpha[p][i][i]*lh (epsilon, ti, v);
60             uh[i].set_u() = sc[i].solve (rhs.u());
61             uh_next -= delta_t*(inv_m*( rk::beta[p][i]*(a*uh[i] - lh(epsilon, ti, v))
62                 + rk::tilde_beta[p][i]*gh(epsilon, ti, uh[i], v)));
63         }
64         uh_next = limiter(uh_next);
65         dout << even(tn+delta_t, uh_next);
66         uh[0] = uh_next;
67     }
68 }

```

Example file 4.11: burgers\_diffusion\_operators.icc

```

1 field lh (Float epsilon, Float t, const test& v) {
2 #ifdef NEUMANN
3     return field (v.get_vf_space(), 0);
4 #else // NEUMANN
5     size_t d = v.get_vf_space().get_geo().dimension();
6     size_t k = v.get_vf_space().degree();
7     Float beta = (k+1)*(k+d)/d;
8     return epsilon*integrate ("boundary",
9         beta*penalty()*g(epsilon,t)*v
10        - g(epsilon,t)*dot(grad_h(v),normal()));
11 #endif // NEUMANN
12 }
13 field gh (Float epsilon, Float t, const field& uh, const test& v) {
14     return - integrate (dot(compose(f,uh),grad_h(v)))
15        + integrate ("internal_sides",
16            compose (phi, normal(), inner(uh), outer(uh))*jump(v))
17        + integrate ("boundary",
18            compose (phi, normal(), uh, g(epsilon,t))*v);
19 }

```

The included file ‘runge\_kutta\_semiimplicit.icc’ is not shown here but is available in the example directory.

### Running the program

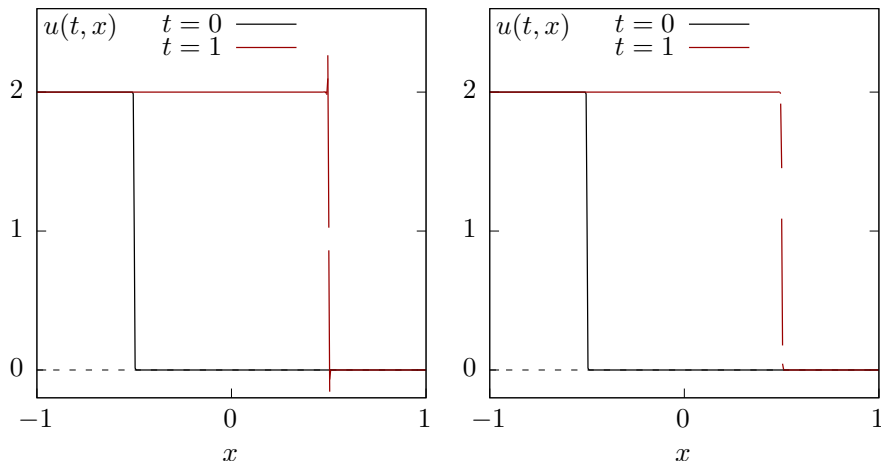


Figure 4.11: Burgers equation with a small diffusion ( $\varepsilon = 10^{-3}$ ). Third order in time semi-implicit scheme with  $P_{1d}$  element. (left) without limiter ; (right) with limiter.

Running the program writes with  $h = 2/400$  and  $\varepsilon = 10^{-2}$  writes:

```

make burgers_diffusion_dg
mkgeo_grid -e 400 -a -1 -b 1 > line.geo
./burgers_diffusion_dg line P1d 0.01 1000 1 3 > line.branch
branch -gnuplot line.branch -umin -0.1 -umax 2.1

```

Decreasing  $\varepsilon = 10^{-3}$  leads to a sharper solution:

```

./burgers_diffusion_dg line P1d 0.001 1000 1 3 > line.branch
branch -gnuplot line.branch -umin -0.1 -umax 2.1

```

As mentioned in [110], the time step should be chosen smaller when  $\varepsilon$  decreases. The result is shown on Fig. 4.11.left. Observe the oscillations near the smoothed shock when there is no limiter

while the value goes outside  $[0, 2]$ . Conversely, with a limiter (see Fig. 4.11.right) the approximate solution is decreasing and there is no more oscillations: the values remains in the range  $[0 : 2]$ .

## 4.3 Fluids and solids computations revisited

### 4.3.1 The linear elasticity problem

The elasticity problem (2.2) has been introduced in volume 1, section 2.1.1, page 41.

(P): find  $\mathbf{u}$  such that

$$\begin{aligned} -\operatorname{div}(\lambda \operatorname{div}(\mathbf{u}).I + 2D(\mathbf{u})) &= \mathbf{f} \text{ in } \Omega \\ \mathbf{u} &= \mathbf{g} \text{ on } \partial\Omega \end{aligned}$$

where  $\lambda \geq -1$  is a constant and  $\mathbf{f}, \mathbf{g}$  given. This problem is a natural extension to vector-valued field of the Poisson problem with Dirichlet boundary conditions.

The variational formulation writes:

(FV)<sub>h</sub>: find  $\mathbf{u} \in \mathbf{V}(\mathbf{g})$  such that

$$a(\mathbf{u}, \mathbf{v}) = l_h(\mathbf{v}), \quad \forall \mathbf{v} \in \mathbf{V}(0)$$

where

$$\begin{aligned} \mathbf{V}(\mathbf{g}) &= \{\mathbf{v} \in H^1(\Omega)^d; \mathbf{v} = \mathbf{g} \text{ on } \partial\Omega\} \\ a(\mathbf{u}, \mathbf{v}) &= \int_{\Omega} (\lambda \operatorname{div}(\mathbf{u}) \operatorname{div}(\mathbf{v}) + 2D(\mathbf{u}) : D(\mathbf{v})) \, dx \\ l(\mathbf{v}) &= \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, dx \end{aligned}$$

The discrete variational formulation writes:

(FV)<sub>h</sub>: find  $u_h \in \mathbf{X}_h$  such that

$$a_h(u_h, v_h) = l_h(v_h), \quad \forall v_h \in \mathbf{X}_h$$

where

$$\begin{aligned} \mathbf{X}_h &= \{\mathbf{v}_h \in L^2(\Omega)^d; \mathbf{v}_h|_K \in P_k^d, \forall K \in \mathcal{T}_h\} \\ a_h(\mathbf{u}, \mathbf{v}) &= \int_{\Omega} (\lambda \operatorname{div}_h(\mathbf{u}) \operatorname{div}_h(\mathbf{v}) + 2D_h(\mathbf{u}) : D_h(\mathbf{v})) \, dx \\ &\quad + \sum_{S \in \mathcal{S}_h} \int_S (\beta \varpi_s \llbracket \mathbf{u} \rrbracket \cdot \llbracket \mathbf{v} \rrbracket - \llbracket \mathbf{u} \rrbracket \cdot \{\{\lambda \operatorname{div}_h(\mathbf{v}) \mathbf{n} + 2D_h(\mathbf{v}) \mathbf{n}\}\} - \llbracket \mathbf{v} \rrbracket \cdot \{\{\lambda \operatorname{div}_h(\mathbf{u}) \mathbf{n} + 2D_h(\mathbf{u}) \mathbf{n}\}\}) \, ds \\ l_h(\mathbf{v}) &= \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, dx + \int_{\partial\Omega} \mathbf{g} \cdot (\beta \varpi_s \mathbf{v} - \lambda \operatorname{div}_h(\mathbf{v}) \mathbf{n} - 2D_h(\mathbf{v}) \mathbf{n}) \, ds \end{aligned}$$

where  $k \geq 1$  is the polynomial degree in  $\mathbf{X}_h$ .

Example file 4.12: elasticity\_taylor\_dg.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "taylor.icc"
5 int main(int argc, char**argv) {
6     environment rheolef (argc, argv);
7     geo omega (argv[1]);
8     space Xh (omega, argv[2], "vector");
9     Float lambda = (argc > 3) ? atof(argv[3]) : 1;
10    size_t d = omega.dimension();
11    size_t k = Xh.degree();
12    Float beta = (k+1)*(k+d)/d;
13    trial u (Xh); test v (Xh);
14    form a = integrate (lambda*div_h(u)*div_h(v) + 2*ddot(Dh(u),Dh(v)))
15            + integrate (omega.sides(),
16                        beta*penalty()*dot(jump(u),jump(v))
17                        - lambda*dot(jump(u),average(div_h(v)*normal()))
18                        - lambda*dot(jump(v),average(div_h(u)*normal()))
19                        - 2*dot(jump(u),average(Dh(v)*normal()))
20                        - 2*dot(jump(v),average(Dh(u)*normal())));
21    field lh = integrate (dot(f(),v))
22                + integrate (omega.boundary(),
23                            beta*penalty()*dot(g(),jump(v))
24                            - lambda*dot(g(),average(div_h(v)*normal()))
25                            - 2*dot(g(),average(Dh(v)*normal())));
26    solver sa (a.uu());
27    field uh(Xh);
28    uh.set_u() = sa.solve(lh.u());
29    dout << uh;
30 }

```

### Comments

The data are given when  $d = 2$  by:

$$\mathbf{g}(x) = \begin{pmatrix} -\cos(\pi x_0) \sin(\pi x_1) \\ \sin(\pi x_0) \cos(\pi x_1) \end{pmatrix} \quad \text{and} \quad \mathbf{f} = 2\pi^2 \mathbf{g} \quad (4.15)$$

This choice is convenient since the exact solution is known  $\mathbf{u} = \mathbf{g}$ . This benchmark solution was proposed in 1923 by Taylor [107] in the context of the Stokes problem. Notice that the solution is independent of  $\lambda$  since  $\text{div}(\mathbf{u}) = 0$ .

Example file 4.13: taylor.icc

```

1 struct g {
2     point operator() (const point& x) const {
3         return point(-cos(pi*x[0])*sin(pi*x[1]),
4                     sin(pi*x[0])*cos(pi*x[1])); }
5     g() : pi(acos(Float(-1.0))) {}
6     const Float pi;
7 };
8 struct f {
9     point operator() (const point& x) const { return 2*sqr(pi)*_g(x); }
10    f() : pi(acos(Float(-1.0))), _g() {}
11    const Float pi; g _g;
12 };

```

As the exact solution is known, the error can be computed. The code `elasticity_taylor_error_dg.cc` and its header file `taylor_exact.icc` compute the error in  $L^2$ ,  $L^\infty$  and energy norms. These files are not listed here but are available in the **Rheolef** example directory. The computation writes:

```

make elasticity_taylor_dg elasticity_taylor_error_dg
mkgeo_grid -t 10 > square.geo

```

```
./elasticity_taylor_dg square P1d | ./elasticity_taylor_error_dg
./elasticity_taylor_dg square P2d | ./elasticity_taylor_error_dg
```

### 4.3.2 The Stokes problem

Let us consider the Stokes problem for the driven cavity in  $\Omega = ]0, 1[^d$ ,  $d = 2, 3$ . The problem has been introduced in volume 1, section 2.1.4, page 51.

(P): find  $\mathbf{u}$  and  $p$ , defined in  $\Omega$ , such that

$$\begin{aligned} -\operatorname{div}(2D(\mathbf{u})) + \nabla p &= \mathbf{f} \text{ in } \Omega, \\ -\operatorname{div} \mathbf{u} &= 0 \text{ in } \Omega, \\ \mathbf{u} &= \mathbf{g} \text{ on } \partial\Omega \end{aligned}$$

where  $\mathbf{f}$  and  $\mathbf{g}$  are given. This problem is the extension to divergence free vector fields of the elasticity problem. The variational formulation writes:

$(VF)_h$  find  $\mathbf{u} \in \mathbf{V}(\mathbf{g})$  and  $p \in L^2(\Omega)$  such that:

$$\begin{aligned} a(\mathbf{u}, \mathbf{v}) + b(\mathbf{v}, p) &= l(\mathbf{v}), \quad \forall \mathbf{v} \in \mathbf{V}(0), \\ b(\mathbf{u}, q) &= 0, \quad \forall q \in L^2(\Omega) \end{aligned} \quad (4.16)$$

where

$$\begin{aligned} \mathbf{V}(\mathbf{g}) &= \{\mathbf{v} \in H^1(\Omega)^d; \mathbf{v} = \mathbf{g} \text{ on } \partial\Omega\} \\ a(\mathbf{u}, \mathbf{v}) &= \int_{\Omega} 2D(\mathbf{u}) : D(\mathbf{v}) \, dx \\ b(\mathbf{u}, q) &= - \int_{\Omega} \operatorname{div}(\mathbf{u}) q \, dx \\ l(\mathbf{v}) &= \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, dx \end{aligned}$$

The discrete variational formulation writes:

$(VF)_h$  find  $\mathbf{u}_h \in \mathbf{X}_h$  and  $p_h \in Q_h$  such that:

$$\begin{aligned} a_h(\mathbf{u}_h, \mathbf{v}_h) + b_h(\mathbf{v}_h, p_h) &= l_h(\mathbf{v}_h), \quad \forall \mathbf{v}_h \in \mathbf{X}_h, \\ b_h(\mathbf{u}_h, q_h) - c_h(p_h, q_h) &= k_h(q), \quad \forall q_h \in Q_h. \end{aligned} \quad (4.17)$$

The discontinuous finite element spaces are defined by:

$$\begin{aligned} \mathbf{X}_h &= \{\mathbf{v}_h \in L^2(\Omega)^d; \mathbf{v}_h|_K \in P_k^d, \forall K \in \mathcal{T}_h\} \\ Q_h &= \{q_h \in L^2(\Omega)^d; q_h|_K \in P_k^d, \forall K \in \mathcal{T}_h\} \end{aligned}$$

where  $k \geq 1$  is the polynomial degree. Notice that velocity and pressure are approximated by the same polynomial order. This method was introduced by [27] and some recent theoretical results can be founded in [30]. The forms are defined for all  $u, v \in H^1(\mathcal{T}_h)^d$  and  $q \in L^2(\Omega)$  by (see



e.g. [31, p. 249]):

$$\begin{aligned}
 a_h(\mathbf{u}, \mathbf{v}) &= \int_{\Omega} 2D_h(\mathbf{u}) : D_h(\mathbf{v}) \, dx \\
 &\quad + \sum_{S \in \mathcal{S}_h} \int_S (\beta \varpi_s [\mathbf{u}] \cdot [\mathbf{v}] - [\mathbf{u}] \cdot \{2D_h(\mathbf{v})\mathbf{n}\} - [\mathbf{v}] \cdot \{2D_h(\mathbf{u})\mathbf{n}\}) \, ds \\
 b_h(\mathbf{u}, q) &= \int_{\Omega} \mathbf{u} \cdot \nabla_h q \, dx - \sum_{S \in \mathcal{S}_h^{(i)}} \int_S \{ \mathbf{u} \} \cdot \mathbf{n} [q] \, ds \\
 c_h(p, q) &= \sum_{S \in \mathcal{S}_h^{(i)}} \int_S h_s [p] [q] \, ds \\
 l_h(\mathbf{v}) &= \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, dx + \int_{\partial\Omega} \mathbf{g} \cdot (\beta \varpi_s \mathbf{v} - 2D_h(\mathbf{v})\mathbf{n}) \, ds \\
 k_h(q) &= \int_{\partial\Omega} \mathbf{g} \cdot \mathbf{n} q \, ds
 \end{aligned}$$

The stabilization form  $c_h$  controls the pressure jump across internal sides. This stabilization term is necessary when using equal order polynomial approximation for velocity and pressure. The definition of the forms is grouped in a subroutine: it will be reused later for the Navier-Stokes problem.

Example file 4.14: stokes\_dirichlet\_dg.icc

```

1 void stokes_dirichlet_dg (const space& Xh, const space& Qh,
2   form& a, form& b, form& c, form& mp, field& lh, field& kh,
3   quadrature_option qopt = quadrature_option())
4 {
5   size_t k = Xh.degree();
6   size_t d = Xh.get_geo().dimension();
7   Float beta = (k+1)*(k+d)/d;
8   trial u (Xh), p (Qh);
9   test v (Xh), q (Qh);
10  a = integrate (2*ddot(Dh(u),Dh(v)), qopt)
11    + integrate ("sides", beta*penalty()*dot(jump(u), jump(v))
12      - 2*dot(jump(u), average(Dh(v)*normal()))
13      - 2*dot(jump(v), average(Dh(u)*normal())), qopt);
14  lh = integrate (dot(f(),v), qopt)
15    + integrate ("boundary", beta*penalty()*dot(g(),v)
16      - 2*dot(g(),Dh(v)*normal()), qopt);
17  b = integrate (dot(u,grad_h(q)), qopt)
18    + integrate ("internal_sides", - dot(average(u),normal())*jump(q), qopt);
19  kh = integrate ("boundary", dot(g(),normal())*q, qopt);
20  c = integrate ("internal_sides", h_local()*jump(p)*jump(q), qopt);
21  mp = integrate (p*q, qopt);
22 }

```

A simple test program writes:

Example file 4.15: stokes\_taylor\_dg.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "taylor.icc"
5 #include "stokes_dirichlet_dg.icc"
6 int main(int argc, char**argv) {
7     environment rheolef (argc, argv);
8     geo omega (argv[1]);
9     space Xh (omega, argv[2], "vector");
10    space Qh (omega, argv[2]);
11    form a, b, c, mp;
12    field lh, kh;
13    stokes_dirichlet_dg (Xh, Qh, a, b, c, mp, lh, kh);
14    field uh (Xh, 0), ph (Qh, 0);
15    solver_abtb stokes (a.uu(), b.uu(), c.uu(), mp.uu());
16    stokes.solve (lh.u(), kh.u(), uh.set_u(), ph.set_u());
17    dout << catchmark("u") << uh
18         << catchmark("p") << ph;
19 }

```

### Comments

The data are given when  $d = 2$  by (4.15). This choice is convenient since the exact solution is known  $\mathbf{u} = \mathbf{g}$  and  $p = 0$ . The code `stokes_taylor_error_dg.cc` compute the error in  $L^2$ ,  $L^\infty$  and energy norms. This code it is not listed here but is available in the **Rheolef** example directory. The computation writes:

```

make stokes_taylor_dg stokes_taylor_error_dg
mkgeo_grid -t 10 > square.geo
./stokes_taylor_dg square P1d | ./stokes_taylor_error_dg
./stokes_taylor_dg square P2d | ./stokes_taylor_error_dg

```

## 4.4 The stationary Navier-Stokes equations

### 4.4.1 Problem statement

The Navier-Stokes problem has been already introduced in volume 1, section 4.4 page 169. Here we consider the stationary version of this problem. Let  $Re \geq 0$  be the Reynolds number. The problem writes:

(P): find  $\mathbf{u}$  and  $p$ , defined in  $\Omega$ , such that

$$\begin{aligned}
 Re(\mathbf{u} \cdot \nabla) \mathbf{u} - \operatorname{div}(2D(\mathbf{u})) + \nabla p &= \mathbf{f} \text{ in } \Omega, \\
 -\operatorname{div} \mathbf{u} &= 0 \text{ in } \Omega, \\
 \mathbf{u} &= \mathbf{g} \text{ on } \partial\Omega
 \end{aligned}$$

Notice that, when  $Re > 0$ , the problem is nonlinear, due to the inertia term  $\mathbf{u} \cdot \nabla \mathbf{u}$ . When  $Re = 0$  the problem reduces to the linear Stokes problem, presented in the previous section/

The variationnal formulation of this nonlinear problem writes:

(FV): find  $\mathbf{u} \in \mathbf{V}(\mathbf{g})$  and  $p \in L^2(\Omega)$  such that

$$\begin{aligned}
 Ret(\mathbf{u}; \mathbf{u}, \mathbf{v}) + a(\mathbf{u}, \mathbf{v}) + b(\mathbf{v}, p) &= l(\mathbf{v}), \forall \mathbf{v} \in \mathbf{V}(0), \\
 b(\mathbf{u}, q) &= 0, \forall q \in L^2(\Omega)
 \end{aligned}$$

where the space  $\mathbf{V}(\mathbf{g})$  and forms  $a$ ,  $b$  and  $l$  are given as in the previous section 4.3.2 for the Stokes problem and the trilinear form  $t(\cdot; \cdot, \cdot)$  is given by:

$$t(\mathbf{w}; \mathbf{u}, \mathbf{v}) = \int_{\Omega} ((\mathbf{w} \cdot \nabla) \mathbf{u}) \cdot \mathbf{v} \, dx$$

#### 4.4.2 The discrete problem

Let

$$t(\mathbf{w}; \mathbf{u}, \mathbf{u}) = \int_{\Omega} (\mathbf{w} \cdot \nabla \mathbf{u}) \cdot \mathbf{u} \, dx$$

Observe that, for all  $\mathbf{u}, \mathbf{w} \in H^1(\Omega)^d$  we have

$$\begin{aligned} \int_{\Omega} (\mathbf{w} \cdot \nabla \mathbf{u}) \cdot \mathbf{u} \, dx &= \sum_{i,j=0}^{d-1} \int_{\Omega} u_i w_j \partial_j(u_i) \, dx \\ &= \sum_{i,j=0}^{d-1} - \int_{\Omega} u_i \partial_j(u_i w_j) \, dx + \int_{\partial\Omega} u_i^2 w_j n_j \, ds \\ &= \sum_{i,j=0}^{d-1} - \int_{\Omega} u_i \partial_j(u_i) w_j \, dx - \int_{\Omega} u_i^2 \partial_j(w_j) \, dx + \int_{\partial\Omega} u_i^2 w_j n_j \, ds \\ &= - \int_{\Omega} (\mathbf{w} \cdot \nabla \mathbf{u}) \cdot \mathbf{u} \, dx - \int_{\Omega} \operatorname{div}(\mathbf{w}) |\mathbf{u}|^2 \, dx + \int_{\partial\Omega} \mathbf{w} \cdot \mathbf{n} |\mathbf{u}|^2 \, ds \end{aligned} \quad (4.18)$$

Thus

$$t(\mathbf{w}; \mathbf{u}, \mathbf{u}) = \int_{\Omega} (\mathbf{w} \cdot \nabla \mathbf{u}) \cdot \mathbf{u} \, dx = -\frac{1}{2} \int_{\Omega} \operatorname{div}(\mathbf{w}) |\mathbf{u}|^2 \, dx + \frac{1}{2} \int_{\partial\Omega} \mathbf{w} \cdot \mathbf{n} |\mathbf{u}|^2 \, ds$$

When  $\operatorname{div}(\mathbf{w}) = 0$ , the trilinear form  $t(\cdot, \cdot, \cdot)$  reduces to a boundary term: it is formally skew-symmetric. The skew-symmetry of  $t$  is an important property: let  $(\mathbf{v}, q) = (\mathbf{u}, p)$  as test functions in  $(FV)$ . We obtain:

$$a(\mathbf{u}, \mathbf{u}) = l(\mathbf{u})$$

In other words, we obtain the same energy balance as for the Stokes flow and inertia do not contribute to the energy balance. This is an important property and we aim at obtaining the same one at the discrete level. As the discrete solution  $\mathbf{u}_h$  is not exactly divergence free, following Temam, we introduce the following modified trilinear form:

$$t^*(\mathbf{w}; \mathbf{u}, \mathbf{v}) = \int_{\Omega} \left( (\mathbf{w} \cdot \nabla \mathbf{u}) \cdot \mathbf{v} + \frac{1}{2} \operatorname{div}(\mathbf{w}) \mathbf{u} \cdot \mathbf{v} \right) dx - \frac{1}{2} \int_{\partial\Omega} (\mathbf{w} \cdot \mathbf{n}) \mathbf{u} \cdot \mathbf{v} \, ds, \quad \forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in H^1(\Omega)^d$$

This form integrates the non-vanishing terms and we have:

$$t^*(\mathbf{w}; \mathbf{u}, \mathbf{u}) = 0, \quad \forall \mathbf{u}, \mathbf{w} \in H^1(\Omega)^d$$

When the discrete solution is not exactly divergence free, it is better to use  $t^*$  than  $t$ .

The discontinuous finite element spaces  $\mathbf{X}_h$  and  $Q_h$  and forms  $a_h, b_h, c_h, l_h$  and  $k_h$  are defined as in the previous section. Let us introduce  $t_h^*$ , the following discrete trilinear form, defined for all  $\mathbf{u}_h, \mathbf{v}_h, \mathbf{w}_h \in \mathbf{X}_h$ :

$$t_h^*(\mathbf{w}_h; \mathbf{u}_h, \mathbf{v}_h) = \int_{\Omega} \left( (\mathbf{w}_h \cdot \nabla_h \mathbf{u}_h) \cdot \mathbf{v}_h + \frac{1}{2} \operatorname{div}_h(\mathbf{w}_h) \mathbf{u}_h \cdot \mathbf{v}_h \right) dx - \frac{1}{2} \int_{\partial\Omega} (\mathbf{w}_h \cdot \mathbf{n}) \mathbf{u}_h \cdot \mathbf{v}_h \, ds$$

Notice that  $t_h^*$  is similar to  $t^*$ : the gradient and divergence has been replaced by their broken counterpart in the first term. As  $\mathbf{X}_h \not\subset H^1(\Omega)^d$ , the skew-symmetry property is not expected to be true at the discrete level. Then

$$t_h^*(\mathbf{w}_h; \mathbf{u}_h, \mathbf{u}_h) = \sum_{K \in \mathcal{T}_h} \int_K \left( (\mathbf{w}_h \cdot \nabla \mathbf{u}_h) \cdot \mathbf{u}_h + \frac{1}{2} \operatorname{div}(\mathbf{w}_h) |\mathbf{u}_h|^2 \right) dx - \frac{1}{2} \int_{\partial\Omega} (\mathbf{w}_h \cdot \mathbf{n}) |\mathbf{u}_h|^2 \, ds$$

As the restriction of  $\mathbf{u}_h$  and  $\mathbf{w}_h$  to each  $K \in \mathcal{T}_h$  belongs to  $H^1(K)^d$ , we have, using a similar integration by part:

$$\int_K (\mathbf{w}_h \cdot \nabla \mathbf{u}_h) \cdot \mathbf{u}_h \, dx = -\frac{1}{2} \int_K \operatorname{div}(\mathbf{w}_h) |\mathbf{u}_h|^2 \, dx + \frac{1}{2} \int_{\partial K} (\mathbf{w}_h \cdot \mathbf{n}) |\mathbf{u}_h|^2 \, ds$$

Thus

$$t_h^*(\mathbf{w}_h; \mathbf{u}_h, \mathbf{u}_h) = \frac{1}{2} \sum_{K \in \mathcal{T}_h} \int_{\partial K} (\mathbf{w}_h \cdot \mathbf{n}) |\mathbf{u}_h|^2 \, ds - \frac{1}{2} \int_{\partial \Omega} (\mathbf{w}_h \cdot \mathbf{n}) |\mathbf{u}_h|^2 \, ds$$

The terms on boundary sides vanish while those on internal sides can be grouped:

$$t_h^*(\mathbf{w}_h; \mathbf{u}_h, \mathbf{u}_h) = \frac{1}{2} \sum_{S \in \mathcal{S}_h^{(i)}} \int_S \llbracket |\mathbf{u}_h|^2 \mathbf{w}_h \rrbracket \cdot \mathbf{n} \, ds$$

The jump term  $\llbracket (\mathbf{u}_h \cdot \mathbf{v}_h) \mathbf{w}_h \rrbracket \cdot \mathbf{n}$  is not easily manageable and could be developed. A short computation shows that, for all scalar fields  $\phi, \varphi$  we have on any internal side:

$$\llbracket \phi \varphi \rrbracket = \llbracket \phi \rrbracket \{\varphi\} + \{\phi\} \llbracket \varphi \rrbracket \quad (4.19)$$

$$\{\phi \varphi\} = \{\phi\} \{\varphi\} + \frac{1}{4} \llbracket \phi \rrbracket \llbracket \varphi \rrbracket \quad (4.20)$$

Then

$$\begin{aligned} t_h^*(\mathbf{w}_h; \mathbf{u}_h, \mathbf{u}_h) &= \frac{1}{2} \sum_{S \in \mathcal{S}_h^{(i)}} \int_S (\{\mathbf{w}_h\} \cdot \mathbf{n} \llbracket |\mathbf{u}_h|^2 \rrbracket + \llbracket \mathbf{w}_h \rrbracket \cdot \mathbf{n} \{\!|\mathbf{u}_h|^2\!\}) \, ds \\ &= \sum_{S \in \mathcal{S}_h^{(i)}} \int_S \left( \{\mathbf{w}_h\} \cdot \mathbf{n} (\llbracket \mathbf{u}_h \rrbracket \cdot \{\!|\mathbf{u}_h|^2\!\}) + \frac{1}{2} \llbracket \mathbf{w}_h \rrbracket \cdot \mathbf{n} \{\!|\mathbf{u}_h|^2\!\} \right) \, ds \end{aligned}$$

Thus, as expected, the skew-symmetry property is no more satisfied at the discrete level, due to the jumps of the fields at the inter-element boundaries. Following the previous idea, we introduce the following modified discrete trilinear form:

$$\begin{aligned} t_h(\mathbf{w}_h; \mathbf{u}_h, \mathbf{v}_h) &= t_h^*(\mathbf{w}_h; \mathbf{u}_h, \mathbf{v}_h) - \sum_{S \in \mathcal{S}_h^{(i)}} \int_S \left( \{\mathbf{w}_h\} \cdot \mathbf{n} (\llbracket \mathbf{u}_h \rrbracket \cdot \{\!|\mathbf{v}_h|^2\!\}) + \frac{1}{2} \llbracket \mathbf{w}_h \rrbracket \cdot \mathbf{n} \{\!|\mathbf{u}_h \cdot \mathbf{v}_h|\!\} \right) \, ds \\ &= \int_{\Omega} \left( (\mathbf{w}_h \cdot \nabla \mathbf{u}_h) \cdot \mathbf{v}_h + \frac{1}{2} \operatorname{div}_h(\mathbf{w}_h) \mathbf{u}_h \cdot \mathbf{v}_h \right) \, dx - \frac{1}{2} \int_{\partial \Omega} (\mathbf{w}_h \cdot \mathbf{n}) \mathbf{u}_h \cdot \mathbf{v}_h \, ds \\ &\quad - \sum_{S \in \mathcal{S}_h^{(i)}} \int_S \left( \{\mathbf{w}_h\} \cdot \mathbf{n} (\llbracket \mathbf{u}_h \rrbracket \cdot \{\!|\mathbf{v}_h|^2\!\}) + \frac{1}{2} \llbracket \mathbf{w}_h \rrbracket \cdot \mathbf{n} \{\!|\mathbf{u}_h \cdot \mathbf{v}_h|\!\} \right) \, ds \quad (4.21) \end{aligned}$$

This expression has been proposed by Pietro and Ern [30, p. 22], eqn (72) (see also [31, p. 272], eqn (6.57)). The boundary term introduced in  $t_h$  may be compensated in the right-hand side:

$$l_h^*(\mathbf{v}) := l_h(\mathbf{v}) - \frac{Re}{2} \int_{\partial \Omega} (\mathbf{g} \cdot \mathbf{n}) \mathbf{g} \cdot \mathbf{v}_h \, ds$$

Notice that the boundary term introduced in  $t_h$  is compensated in the right-hand side  $l_h^*$ .

Example file 4.16: inertia.icc

```

1  template<class W, class U, class V>
2  form inertia (W w, U u, V v,
3    quadrature_option qopt = quadrature_option())
4  {
5    return
6    integrate (dot(grad_h(u)*w,v) + 0.5*div_h(w)*dot(u,v), qopt)
7    + integrate ("boundary", - 0.5*dot(w,normal())*dot(u,v), qopt)
8    + integrate ("internal_sides",
9      - dot(average(w),normal())*dot(jump(u),average(v))
10     - 0.5*dot(jump(w),normal())
11     *(dot(average(u),average(v)) + 0.25*dot(jump(u),jump(v))), qopt);
12 }
13 field inertia_fix_rhs (test v,
14   quadrature_option qopt = quadrature_option())
15 {
16   return integrate("boundary", - 0.5*dot(g(),normal())*dot(g(),v), qopt);
17 }

```

The discrete problem is

$(FV)_h$ : find  $\mathbf{u}_h \in \mathbf{X}_h$  and  $p \in Q_h$  such that

$$\begin{aligned}
 Ret_h(\mathbf{u}_h; \mathbf{u}_h, \mathbf{v}_h) + a_h(\mathbf{u}_h, \mathbf{v}_h) + b_h(\mathbf{v}_h, p_h) &= l_h^*(\mathbf{v}_h), \forall \mathbf{v}_h \in \mathbf{X}_h, \\
 b_h(\mathbf{u}_h, q_h) - c_h(p_h, q_h) &= k_h(q), \forall q_h \in Q_h
 \end{aligned} \tag{4.22}$$

The simplest approach for solving the discrete problem is to consider a fixed-point algorithm. The sequence  $(\mathbf{u}_h^{(k)})_{k \geq 0}$  is defined by recurrence as:

- $k = 0$ : let  $\mathbf{u}_h^{(0)} \in \mathbf{X}_h$  being known.
- $k \geq 0$ : let  $\mathbf{u}_h^{(k-1)} \in \mathbf{X}_h$  given. Find  $\mathbf{u}_h^{(k)} \in \mathbf{X}_h$  and  $p_h^{(k)} \in Q_h$  such that

$$\begin{aligned}
 Ret_h(\mathbf{u}_h^{(k-1)}; \mathbf{u}_h^{(k)}, \mathbf{v}_h) + a_h(\mathbf{u}_h^{(k)}, \mathbf{v}_h) + b_h(\mathbf{v}_h, p_h^{(k)}) &= l_h^*(\mathbf{v}_h), \forall \mathbf{v}_h \in \mathbf{X}_h, \\
 b_h(\mathbf{u}_h^{(k)}, q_h) - c_h(p_h^{(k)}, q_h) &= k_h(q), \forall q_h \in Q_h.
 \end{aligned}$$

At each step  $k \geq 0$ , this algorithm involves a linear subproblem of Stokes-type.

Example file 4.17: navier\_stokes\_taylor\_dg.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "taylor.icc"
5 #include "stokes_dirichlet_dg.icc"
6 #include "inertia.icc"
7 int main(int argc, char**argv) {
8     environment rheolef (argc, argv);
9     geo omega (argv[1]);
10    space Xh (omega, argv[2], "vector");
11    space Qh (omega, argv[2]);
12    Float Re = (argc > 3) ? atof(argv[3]) : 1;
13    size_t max_iter = (argc > 4) ? atoi(argv[4]) : 1;
14    form a, b, c, mp;
15    field lh, kh;
16    stokes_dirichlet_dg (Xh, Qh, a, b, c, mp, lh, kh);
17    field uh (Xh, 0), ph (Qh, 0);
18    solver_abtb stokes (a.uu(), b.uu(), c.uu(), mp.uu());
19    stokes.solve (lh.u(), kh.u(), uh.set_u(), ph.set_u());
20    trial u (Xh); test v (Xh);
21    form a1 = a + Re*inertia (uh, u, v);
22    lh += Re*inertia_fix_rhs (v);
23    derr << "#k r as" << endl;
24    for (size_t k = 0; k < max_iter; ++k) {
25        solver_abtb stokes (a1.uu(), b.uu(), c.uu(), mp.uu());
26        stokes.solve (lh.u(), kh.u(), uh.set_u(), ph.set_u());
27        form th = inertia (uh, u, v);
28        a1 = a + Re*th;
29        field rh = a1*uh + b.trans_mult(ph) - lh;
30        derr << k << " " << rh.max_abs() << " " << th(uh,uh) << endl;
31    }
32    dout << catchmark("Re") << Re << endl
33         << catchmark("u") << uh
34         << catchmark("p") << ph;
35 }

```

### Comments

The data are given when  $d = 2$  by (4.15). This choice is convenient since the exact solution is known  $\mathbf{u} = \mathbf{g}$  and  $p = -(Re/4)(\cos(2\pi x_0) + \cos(2\pi x_1))$ . The code `navier_stokes_taylor_error_dg.cc` compute the error in  $L^2$ ,  $L^\infty$  and energy norms. This code it is not listed here but is available in the `Rheolef` example directory. The computation writes:

```

make navier_stokes_taylor_dg navier_stokes_taylor_error_dg
./navier_stokes_taylor_dg square P1d 10 10 | ./navier_stokes_taylor_error_dg
./navier_stokes_taylor_dg square P2d 10 10 | ./navier_stokes_taylor_error_dg

```

### 4.4.3 A conservative variant

Remark the identity

$$\operatorname{div}(\mathbf{u} \otimes \mathbf{u}) = (\mathbf{u} \cdot \nabla) \mathbf{u} + \operatorname{div}(\mathbf{u}) \mathbf{u}$$

The momentum conservation can be rewritten in conservative form and the problem writes:

( $\tilde{P}$ ): find  $\mathbf{u}$  and  $p$ , defined in  $\Omega$ , such that

$$\begin{aligned} \operatorname{div}(Re \mathbf{u} \otimes \mathbf{u} - 2D(\mathbf{u})) + \nabla p &= \mathbf{f} \text{ in } \Omega, \\ - \operatorname{div} \mathbf{u} &= 0 \text{ in } \Omega, \\ \mathbf{u} &= \mathbf{g} \text{ on } \partial\Omega \end{aligned}$$

Notice the Green formulae (see volume 1, appendix A.1.2, page 229):

$$\int_{\Omega} \operatorname{div}(\mathbf{u} \otimes \mathbf{u}) \cdot \mathbf{v} \, dx = - \int_{\Omega} (\mathbf{u} \otimes \mathbf{u}) : \nabla \mathbf{v} \, dx + \int_{\partial\Omega} (\mathbf{u} \cdot \mathbf{n})(\mathbf{u} \cdot \mathbf{v}) \, ds$$

The variational formulation is:

( $\widetilde{FV}$ ): find  $\mathbf{u} \in \mathbf{V}(\mathbf{g})$  and  $p \in L^2(\Omega)$  such that

$$\begin{aligned} \operatorname{Re} \tilde{t}(\mathbf{u}; \mathbf{u}, \mathbf{v}) + a(\mathbf{u}, \mathbf{v}) + b(\mathbf{v}, p) &= \tilde{l}(\mathbf{v}), \quad \forall \mathbf{v} \in \mathbf{V}(0), \\ b(\mathbf{u}, q) &= 0, \quad \forall q \in L^2(\Omega) \end{aligned}$$

where the forms  $\tilde{t}$  and  $\tilde{l}_h$  are given by:

$$\begin{aligned} \tilde{t}(\mathbf{w}; \mathbf{u}, \mathbf{v}) &= - \int_{\Omega} (\mathbf{w} \otimes \mathbf{u}) : \nabla \mathbf{v} \, dx \\ \tilde{l}(\mathbf{v}) &= l(\mathbf{v}) - \operatorname{Re} \int_{\partial\Omega} (\mathbf{g} \cdot \mathbf{n}) (\mathbf{g} \cdot \mathbf{v}) \, ds \end{aligned}$$

Notice that the right-hand side  $\tilde{l}$  contains an additional term that compensates those coming from the integration by parts. Then, with  $\mathbf{v} = \mathbf{u}$ :

$$\begin{aligned} \tilde{t}(\mathbf{w}; \mathbf{u}, \mathbf{u}) &= - \int_{\Omega} (\mathbf{w} \otimes \mathbf{u}) : \nabla \mathbf{u} \, dx \\ &= \int_{\Omega} \operatorname{div}(\mathbf{w} \otimes \mathbf{u}) \cdot \mathbf{u} \, dx - \int_{\partial\Omega} (\mathbf{w} \otimes \mathbf{u}) : (\mathbf{u} \otimes \mathbf{n}) \, dx \\ &= \int_{\Omega} ((\mathbf{u} \cdot \nabla) \mathbf{w}) \cdot \mathbf{u} + \operatorname{div}(\mathbf{u}) (\mathbf{u} \cdot \mathbf{w}) \, dx - \int_{\partial\Omega} (\mathbf{u} \cdot \mathbf{n}) (\mathbf{u} \cdot \mathbf{w}) \, dx \end{aligned}$$

From an integration by part similar to (4.18):

$$\int_{\Omega} (\mathbf{u} \cdot \nabla \mathbf{w}) \cdot \mathbf{u} \, dx = - \int_{\Omega} (\mathbf{u} \cdot \nabla \mathbf{u}) \cdot \mathbf{w} \, dx - \int_{\Omega} \operatorname{div}(\mathbf{u}) (\mathbf{u} \cdot \mathbf{w}) \, dx + \int_{\partial\Omega} (\mathbf{u} \cdot \mathbf{n}) (\mathbf{u} \cdot \mathbf{w}) \, ds$$

The term  $(\mathbf{u} \cdot \nabla \mathbf{w}) \cdot \mathbf{u}$  do not reappear after the integration by parts: instead, it appears  $(\mathbf{u} \cdot \nabla \mathbf{u}) \cdot \mathbf{w}$ . Thus, the structure of the  $\tilde{t}$  trilinear form do not permit a general skew-symmetry property as it was the case for  $t$ . It requires the three arguments to be the same:

$$\tilde{t}(\mathbf{u}; \mathbf{u}, \mathbf{u}) = \int_{\Omega} ((\mathbf{u} \cdot \nabla) \mathbf{u}) \cdot \mathbf{u} + \operatorname{div}(\mathbf{u}) |\mathbf{u}|^2 \, dx - \int_{\partial\Omega} (\mathbf{u} \cdot \mathbf{n}) |\mathbf{u}|^2 \, dx$$

Using (4.18) with  $\mathbf{w} = \mathbf{u}$  leads to:

$$\int_{\Omega} ((\mathbf{u} \cdot \nabla) \mathbf{u}) \cdot \mathbf{u} \, dx = -\frac{1}{2} \int_{\Omega} \operatorname{div}(\mathbf{u}) |\mathbf{u}|^2 \, dx + \frac{1}{2} \int_{\partial\Omega} (\mathbf{u} \cdot \mathbf{n}) |\mathbf{u}|^2 \, ds \quad (4.23)$$

Then

$$\tilde{t}(\mathbf{u}; \mathbf{u}, \mathbf{u}) = \frac{1}{2} \int_{\Omega} \operatorname{div}(\mathbf{u}) |\mathbf{u}|^2 \, dx - \frac{1}{2} \int_{\partial\Omega} (\mathbf{u} \cdot \mathbf{n}) |\mathbf{u}|^2 \, ds$$

When working with velocities that are not divergence-free, a possible modification of the trilinear form  $\tilde{t}$  is to consider

$$\begin{aligned} \tilde{t}^*(\mathbf{w}; \mathbf{u}, \mathbf{v}) &= \tilde{t}(\mathbf{w}; \mathbf{u}, \mathbf{v}) - \frac{1}{2} \int_{\Omega} \operatorname{div}(\mathbf{v}) (\mathbf{u} \cdot \mathbf{w}) \, dx + \frac{1}{2} \int_{\partial\Omega} (\mathbf{v} \cdot \mathbf{n}) (\mathbf{u} \cdot \mathbf{w}) \, ds \\ &= - \int_{\Omega} \left( (\mathbf{w} \otimes \mathbf{u}) : D(\mathbf{v}) + \frac{1}{2} \operatorname{div}(\mathbf{v}) (\mathbf{u} \cdot \mathbf{w}) \right) \, dx + \frac{1}{2} \int_{\partial\Omega} (\mathbf{v} \cdot \mathbf{n}) (\mathbf{u} \cdot \mathbf{w}) \, ds \end{aligned}$$

Then we have

$$\tilde{t}^*(\mathbf{u}; \mathbf{u}, \mathbf{u}) = 0, \quad \forall \mathbf{u} \in H^1(\Omega)^d$$

The new variational formulation is:

$(\widetilde{FV})^*$ : find  $\mathbf{u} \in \mathbf{V}(\mathbf{g})$  and  $\tilde{p} \in L^2(\Omega)$  such that

$$\begin{aligned} Re \tilde{t}^*(\mathbf{u}; \mathbf{u}, \mathbf{v}) + a(\mathbf{u}, \mathbf{v}) + b(\mathbf{v}, \tilde{p}) &= \tilde{l}(\mathbf{v}), \quad \forall \mathbf{v} \in \mathbf{V}(0), \\ b(\mathbf{u}, q) &= 0, \quad \forall q \in L^2(\Omega) \end{aligned}$$

One can easily check that when  $(\mathbf{u}, \tilde{p})$  is a solution of  $(\widetilde{FV})^*$ , then  $(\mathbf{u}, p)$  is a solution of  $(\widetilde{FV})$  with  $p = \tilde{p} + Re|\mathbf{u}|/2$ . The apparition of the kinetic energy term  $Re|\mathbf{u}|/2$  in the modified pressure field  $\tilde{p}$  is due to the introduction of the  $\text{div}(\mathbf{v})(\mathbf{u} \cdot \mathbf{w})$  term in the trilinear form  $\tilde{t}^*$ .

At the discrete level, let us define for all  $\mathbf{u}_h, \mathbf{v}_h, \mathbf{w}_h \in \mathbf{X}_h$ :

$$\begin{aligned} \tilde{t}_h^*(\mathbf{w}_h; \mathbf{u}_h, \mathbf{v}_h) &= - \int_{\Omega} \left( (\mathbf{w}_h \otimes \mathbf{u}_h) : \nabla_h \mathbf{v}_h + \frac{1}{2} \text{div}_h(\mathbf{v}_h)(\mathbf{u}_h \cdot \mathbf{w}_h) \right) dx \\ &\quad + \frac{1}{2} \int_{\partial\Omega} (\mathbf{v}_h \cdot \mathbf{n})(\mathbf{u}_h \cdot \mathbf{w}_h) ds \end{aligned}$$

Notice that  $\tilde{t}_h^*$  is similar to  $\tilde{t}^*$ : the gradient and divergence has been replaced by their broken counterpart in the first term. As  $\mathbf{X}_h \not\subset H^1(\Omega)^d$ , the skew-symmetry property is not expected to be true at the discrete level. Then

$$\tilde{t}_h^*(\mathbf{u}_h; \mathbf{u}_h, \mathbf{u}_h) = - \int_{\Omega} \left( (\mathbf{u}_h \otimes \mathbf{u}_h) : \nabla_h \mathbf{u}_h + \frac{1}{2} \text{div}_h(\mathbf{u}_h)|\mathbf{u}_h|^2 \right) dx + \frac{1}{2} \int_{\partial\Omega} (\mathbf{u}_h \cdot \mathbf{n})|\mathbf{u}_h|^2 ds$$

Next, using (4.23) in each  $K$ , and then developing thanks to (4.19)-(4.20), we get

$$\begin{aligned} \tilde{t}_h^*(\mathbf{u}_h; \mathbf{u}_h, \mathbf{u}_h) &= \frac{1}{2} \int_{\partial\Omega} (\mathbf{u}_h \cdot \mathbf{n})|\mathbf{u}_h|^2 ds - \frac{1}{2} \sum_{K \in \mathcal{T}_h} \int_{\partial K} (\mathbf{u}_h \cdot \mathbf{n})|\mathbf{u}_h|^2 ds \\ &= -\frac{1}{2} \sum_{S \in \mathcal{S}_h^{(i)}} \int_S \llbracket (\mathbf{u}_h \cdot \mathbf{n})|\mathbf{u}_h|^2 \rrbracket ds \\ &= -\frac{1}{2} \sum_{S \in \mathcal{S}_h^{(i)}} \int_S \left( (\llbracket \mathbf{u}_h \rrbracket \cdot \mathbf{n}) \llbracket |\mathbf{u}_h|^2 \rrbracket + (\llbracket \mathbf{u}_h \rrbracket \cdot \mathbf{n}) \{ |\mathbf{u}_h|^2 \} \right) ds \\ &= - \sum_{S \in \mathcal{S}_h^{(i)}} \int_S \left( (\llbracket \mathbf{u}_h \rrbracket \cdot \mathbf{n}) (\llbracket \mathbf{u}_h \rrbracket \cdot \llbracket \mathbf{u}_h \rrbracket) + \frac{1}{2} (\llbracket \mathbf{u}_h \rrbracket \cdot \mathbf{n}) \{ |\mathbf{u}_h|^2 \} \right) ds \end{aligned}$$

The idea is to integrate this term in the definition of a discrete  $\tilde{t}_h$ . One of the possibilities is

$$\begin{aligned} \tilde{t}_h(\mathbf{w}_h; \mathbf{u}_h, \mathbf{v}_h) &= \tilde{t}_h^*(\mathbf{w}_h; \mathbf{u}_h, \mathbf{v}_h) + \sum_{S \in \mathcal{S}_h^{(i)}} \int_S \left( (\llbracket \mathbf{u}_h \rrbracket \cdot \mathbf{n}) (\llbracket \mathbf{w}_h \rrbracket \cdot \llbracket \mathbf{v}_h \rrbracket) + \frac{1}{2} \{ \mathbf{u}_h \cdot \mathbf{w}_h \} (\llbracket \mathbf{v}_h \rrbracket \cdot \mathbf{n}) \right) ds \\ &= - \int_{\Omega} \left( (\mathbf{w}_h \otimes \mathbf{u}_h) : \nabla_h \mathbf{v}_h + \frac{1}{2} \text{div}_h(\mathbf{v}_h)(\mathbf{u}_h \cdot \mathbf{w}_h) \right) dx \\ &\quad + \frac{1}{2} \int_{\partial\Omega} (\mathbf{v}_h \cdot \mathbf{n})(\mathbf{u}_h \cdot \mathbf{w}_h) ds \\ &\quad + \sum_{S \in \mathcal{S}_h^{(i)}} \int_S \left( (\llbracket \mathbf{u}_h \rrbracket \cdot \mathbf{n}) (\llbracket \mathbf{w}_h \rrbracket \cdot \llbracket \mathbf{v}_h \rrbracket) + \frac{1}{2} \{ \mathbf{u}_h \cdot \mathbf{w}_h \} (\llbracket \mathbf{v}_h \rrbracket \cdot \mathbf{n}) \right) ds \end{aligned} \quad (4.24)$$

This expression was proposed by [30, p. 21], eqn (73) (see also [31, p. 282]) following and original idea introduced in [26].



Example file 4.18: inertia\_cks.icc

```

1 form inertia (field w, trial u, test v,
2   quadrature_option qopt = quadrature_option())
3 {
4   return
5     integrate (- dot(trans(grad_h(v))*w,u) - 0.5*div_h(v)*dot(u,w), qopt)
6   + integrate ("internal_sides",
7     dot(average(u),normal())*dot(jump(v),average(w))
8     + 0.5*dot(jump(v),normal())
9     *(dot(average(u),average(w)) + 0.25*dot(jump(u),jump(w))), qopt)
10  + integrate ("boundary", 0.5*dot(v,normal())*dot(u,w), qopt);
11 }
12 field inertia_fix_rhs (test v,
13   quadrature_option qopt = quadrature_option())
14 {
15   return integrate("boundary", -dot(g(),normal())*dot(g(),v), qopt);
16 }

```

The discrete problem is

$(\widetilde{FV})_h$ : find  $\mathbf{u}_h \in \mathbf{X}_h$  and  $\tilde{p} \in Q_h$  such that

$$\begin{aligned} \text{Re } \tilde{t}_h(\mathbf{u}_h; \mathbf{u}_h, \mathbf{v}_h) + a_h(\mathbf{u}_h, \mathbf{v}_h) + b_h(\mathbf{v}_h, \tilde{p}_h) &= \tilde{l}_h^*(\mathbf{v}_h), \forall \mathbf{v}_h \in \mathbf{X}_h, \\ b_h(\mathbf{u}_h, q_h) - c_h(p_h, q_h) &= k_h(q), \forall q_h \in Q_h \end{aligned}$$

A simple test program is obtained by replacing in `navier_stokes_taylor_dg.cc` the include `inertia.icc` by `inertia_cks.icc`. The compilation and run are similar.

#### 4.4.4 Newton solver

The discrete problems  $(FV)_h$  can be put in a compact form:

$$F(\mathbf{u}_h, p_h) = 0$$

where  $F$  is defined in variational form:

$$\langle F(\mathbf{u}_h, p_h), (\mathbf{v}_h, q_h) \rangle = \left( \begin{array}{cccc} \text{Re } t_h(\mathbf{u}_h; \mathbf{u}_h, \mathbf{v}_h) & + & a_h(\mathbf{u}_h, \mathbf{v}_h) & + & b_h(\mathbf{v}_h, p_h) & - & l_h^*(\mathbf{v}_h) \\ & & b_h(\mathbf{u}_h, q_h) & - & c_h(p_h, q_h) & - & k_h(q) \end{array} \right)$$

for all  $(\mathbf{v}_h, q_h) \in \mathbf{X}_h \times Q_h$ . Notices that, after some minor modifications in the definition of  $F$ , this method could also applies for the locally conservative formulation  $(\widetilde{FV})_h$ . The previous formulation is simply the variational expression of  $F(\mathbf{u}_h, p_h) = 0$ . The Newton method defines the sequence  $(\mathbf{u}_h^{(k)})_{k \geq 0}$  by recurrence as:

- $k = 0$ : let  $\mathbf{u}_h^{(0)} \in \mathbf{X}_h$  being known.
- $k \geq 0$ : let  $\mathbf{u}_h^{(k-1)} \in \mathbf{X}_h$  given. Find  $\delta \mathbf{u}_h \in \mathbf{X}_h$  and  $\delta p_h \in Q_h$  such that

$$F'(\mathbf{u}_h^{(k-1)}, p_h^{(k-1)}) \cdot (\delta \mathbf{u}_h, \delta p_h) = -F(\mathbf{u}_h^{(k-1)}, p_h^{(k-1)})$$

and then defines

$$\mathbf{u}_h^{(k)} = \mathbf{u}_h^{(k-1)} + \delta \mathbf{u}_h \quad \text{and} \quad p_h^{(k)} = p_h^{(k-1)} + \delta p_h$$

At each step  $k \geq 0$ , this algorithm involves a linear subproblem involving the jacobian  $F'$  that is defined by its variational form:

$$\begin{aligned} &\langle F'(\mathbf{u}_h^{(k-1)}, p_h^{(k-1)}) \cdot (\delta \mathbf{u}_h, \delta p_h), (\mathbf{v}_h, q_h) \rangle \\ &= \left( \begin{array}{cccc} \text{Re } (t_h(\delta \mathbf{u}_h; \mathbf{u}_h, \mathbf{v}_h) + t_h(\mathbf{u}_h; \delta \mathbf{u}_h, \mathbf{v}_h)) & + & a_h(\delta \mathbf{u}_h, \mathbf{v}_h) & + & b_h(\mathbf{v}_h, \delta p_h) \\ & & b_h(\delta \mathbf{u}_h, q_h) & - & c_h(\delta p_h, q_h) \end{array} \right) \end{aligned}$$

Example file 4.19: navier\_stokes\_taylor\_newton\_dg.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "taylor.icc"
5 #include "stokes_dirichlet_dg.icc"
6 #include "inertia.icc"
7 #include "navier_stokes_dg.h"
8 int main(int argc, char**argv) {
9     environment rheolef (argc, argv);
10    Float eps = numeric_limits<Float>::epsilon();
11    geo omega (argv[1]);
12    string approx = (argc > 2) ? argv[2] : "P1d";
13    Float Re = (argc > 3) ? atof(argv[3]) : 100;
14    Float tol = (argc > 4) ? atof(argv[4]) : eps;
15    size_t max_iter = (argc > 5) ? atoi(argv[5]) : 100;
16    string restart = (argc > 6) ? argv[6] : "";
17    navier_stokes_dg F (Re, omega, approx);
18    navier_stokes_dg::value_type xh = F.initial (restart);
19    int status = damped_newton (F, xh, tol, max_iter, &derr);
20    dout << catchmark("Re") << Re << endl
21         << catchmark("u") << xh[0]
22         << catchmark("p") << xh[1];
23    return status;
24 }

```

## Comments

The implementation of the Newton method follows the generic approach introduced in volume 1, section 3.2.3, page 114. For that purpose we define a class `navier_stokes_dg`.

Example file 4.20: navier\_stokes\_dg.h

```

1 struct navier_stokes_dg {
2     typedef Float float_type;
3     typedef valarray<field> value_type;
4     navier_stokes_dg (Float Re, const geo& omega, string approx);
5     value_type initial (string restart) const;
6     value_type residue (const value_type& uh) const;
7     void update_derivative (const value_type& uh) const;
8     value_type derivative_solve (const value_type& mrh) const;
9     value_type derivative_trans_mult (const value_type& mrh) const;
10    Float space_norm (const value_type& uh) const;
11    Float dual_space_norm (const value_type& mrh) const;
12    Float Re;
13    space Xh, Qh;
14    quadrature_option qopt;
15    form a0, b, c, mu, mp;
16    field lh0, lh, kh;
17    solver smu, smp;
18    mutable form a1;
19    mutable solver_abtb stokes1;
20 };
21 #include "navier_stokes_dg1.icc"
22 #include "navier_stokes_dg2.icc"

```

The member functions of the class are defined in two separate files.

Example file 4.21: navier\_stokes\_dg1.icc

```

1  navier_stokes_dg::navier_stokes_dg (
2  Float Re1, const geo& omega, string approx)
3  : Re(Re1), Xh(), Qh(), qopt(), a0(), b(), c(), mu(), mp(), lh0(), lh(), kh(),
4  smu(), smp(), a1(), stokes1()
5  {
6  Xh = space (omega, approx, "vector");
7  Qh = space (omega, approx);
8  qopt.set_family(quadrature_option::gauss);
9  qopt.set_order(2*Xh.degree()+1);
10 stokes_dirichlet_dg (Xh, Qh, a0, b, c, mp, lh0, kh, qopt);
11 trial u (Xh); test v (Xh);
12 lh = lh0 + Re*inertia_fix_rhs (v, qopt);
13 mu = integrate (dot(u,v), qopt);
14 smu = solver(mu.uu());
15 smp = solver(mp.uu());
16 }
17 navier_stokes_dg::value_type
18 navier_stokes_dg::initial (string restart) const {
19 value_type xh(2);
20 xh[0] = field (Xh, 0);
21 xh[1] = field (Qh, 0);
22 Float Re0 = 0;
23 if (restart == "") {
24 solver_abtb stokes0 (a0.uu(), b.uu(), c.uu(), mp.uu());
25 stokes0.solve (lh0.u(), kh.u(), xh[0].set_u(), xh[1].set_u());
26 } else {
27 idiststream in (restart);
28 in >> catchmark("Re") >> Re0
29 >> catchmark("u") >> xh[0]
30 >> catchmark("p") >> xh[1];
31 check_macro (xh[1].get_space() == Qh, "unexpected "
32 << xh[0].get_space().stamp() << " approximation in file \""
33 << restart << "\" (" << Xh.stamp() << " expected)");
34 }
35 derr << "# continuation: from Re=" << Re0 << " to " << Re << endl;
36 return xh;
37 }
38 navier_stokes_dg::value_type
39 navier_stokes_dg::residue (const value_type& xh) const {
40 trial u (Xh); test v (Xh);
41 form a = a0 + Re*inertia(xh[0], u, v, qopt);
42 value_type mrh(2);
43 mrh[0] = a*xh[0] + b.trans_mult(xh[1]) - lh;
44 mrh[1] = b*xh[0] - c*xh[1] - kh;
45 return mrh;
46 }
47 void navier_stokes_dg::update_derivative (const value_type& xh) const {
48 trial u (Xh); test v (Xh);
49 a1 = a0 + Re*(inertia(xh[0], u, v, qopt) + inertia(u, xh[0], v, qopt));
50 stokes1 = solver_abtb (a1.uu(), b.uu(), c.uu(), mp.uu());
51 }
52 navier_stokes_dg::value_type
53 navier_stokes_dg::derivative_solve (const value_type& mrh) const {
54 value_type delta_xh(2);
55 delta_xh[0] = field (Xh, 0);
56 delta_xh[1] = field (Qh, 0);
57 stokes1.solve (mrh[0].u(), mrh[1].u(),
58 delta_xh[0].set_u(), delta_xh[1].set_u());
59 return delta_xh;
60 }
61 navier_stokes_dg::value_type
62 navier_stokes_dg::derivative_trans_mult (const value_type& mrh) const {
63 value_type rh(2);
64 rh[0] = field (Xh);
65 rh[1] = field (Qh);
66 rh[0].set_u() = smu.solve(mrh[0].u());
67 rh[1].set_u() = smp.solve(mrh[1].u());
68 value_type mgh(2);
69 mgh[0] = a1.trans_mult(rh[0]) + b.trans_mult(rh[1]);
70 mgh[1] = b*rh[0] - c*rh[1];
71 return mgh;
72 }

```

Example file 4.22: navier\_stokes\_dg2.icc

```

1 Float navier_stokes_dg::space_norm (const value_type& xh) const {
2   return sqrt (mu(xh[0],xh[0]) + mp(xh[1],xh[1]));
3 }
4 Float navier_stokes_dg::dual_space_norm (const value_type& mrh) const {
5   value_type rh(2);
6   rh[0] = field (Xh,0);
7   rh[1] = field (Qh,0);
8   rh[0].set_u() = smu.solve(mrh[0].u());
9   rh[1].set_u() = smp.solve(mrh[1].u());
10  return sqrt (dual(rh[0],mrh[0]) + dual(rh[1],mrh[1]));
11 }

```

```

make navier_stokes_taylor_newton_dg navier_stokes_taylor_error_dg
./navier_stokes_taylor_newton_dg square P2d 1000 | ./navier_stokes_taylor_error_dg

```

#### 4.4.5 Application to the driven cavity benchmark

Example file 4.23: cavity\_dg.icc

```

1 struct g {
2   point operator() (const point& x) const {
3     return point((abs(1-x[1]) < 1e-7) ? 1 : 0, 0, 0); }
4 };
5 struct f {
6   point operator() (const point& x) const { return point(0,0,0); }
7 };

```

The program `navier_stokes_cavity_newton_dg.cc` is obtained by replacing in `navier_stokes_taylor_newton_dg.cc` the include `taylor.icc` by `cavity_dg.icc` that defines the boundary conditions. The compilation and run are similar.

```

make navier_stokes_cavity_newton_dg streamf_cavity
./navier_stokes_cavity_newton_dg square P1d 500 > square.field
field square.field -proj -field | ./streamf_cavity | \
field -bw -n-iso-negative 10 -

```

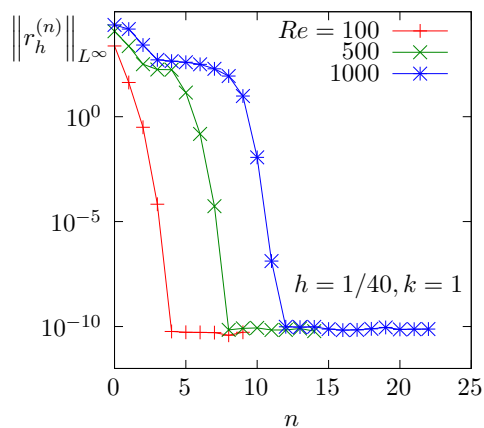


Figure 4.12: The discontinuous Galerkin method for the Navier-Stokes problem on the driven cavity benchmark when  $k = 1$  and  $d = 2$ : convergence of the damped Newton algorithm.

### 4.4.6 Upwinding

The skew symmetry property is generalized to the requirement that  $t_h$  be non-dissipative (see [31, p. 282], eqn (6.68)):

$$t_h(\mathbf{w}_h; \mathbf{u}_h, \mathbf{u}_h) \geq 0, \quad \forall \mathbf{w}_h, \mathbf{u}_h \in \mathbf{X}_h$$

A way to satisfy this property is to add an *upwinding* term in  $t_h$ :

$$\check{t}_h(\mathbf{w}_h; \mathbf{u}_h, \mathbf{v}_h) := t_h(\mathbf{w}_h; \mathbf{u}_h, \mathbf{v}_h) + s_h(\mathbf{w}_h; \mathbf{u}_h, \mathbf{v}_h)$$

with

$$s_h(\mathbf{w}_h; \mathbf{u}_h, \mathbf{v}_h) = \frac{1}{2} \sum_{S \in \mathcal{S}_h^{(i)}} \int_S |\{\{\mathbf{w}_h\}\} \cdot \mathbf{n}| (\llbracket \mathbf{u}_h \rrbracket \cdot \llbracket \mathbf{v}_h \rrbracket) ds$$

We aim at using a Newton method. We replace  $t_h$  by its extension  $\check{t}_h$  containing the upwind terms in the definition of  $F$ , and then we compute its jacobian  $F'$ . As the absolute value is not differentiable, the functions  $s_h$ ,  $\check{t}_h$  and then  $F$  are also not differentiable with respect to  $\mathbf{w}_h$ . Nevertheless, the absolute value is convex and we can use some concets of the sudifferential calculus. Let us introduce the multi-valued sign function:

$$\text{sgn}(x) = \begin{cases} \{1\} & \text{when } x > 0 \\ [-1, 1] & \text{when } x = 0 \\ \{-1\} & \text{when } x < 0 \end{cases}$$

Then, the subdifferential of the absolute value function is  $\text{sgn}(x)$  and for all  $\delta \mathbf{w}_h, \mathbf{w}_h, \mathbf{u}_h, \mathbf{v}_h \in \mathbf{X}_h$ , we define a generalization of the partial derivative as

$$\frac{\partial s_h}{\partial \mathbf{w}_h}(\mathbf{w}_h; \mathbf{u}_h, \mathbf{v}_h) \cdot (\delta \mathbf{w}_h) = \frac{1}{2} \sum_{S \in \mathcal{S}_h^{(i)}} \int_S \text{sgn}(\{\{\mathbf{w}_h\}\} \cdot \mathbf{n}) (\{\{\delta \mathbf{w}_h\}\} \cdot \mathbf{n}) (\llbracket \mathbf{u}_h \rrbracket \cdot \llbracket \mathbf{v}_h \rrbracket) ds$$

Example file 4.24: inertia\_upw.icc

```

1 #include "sgn.icc"
2 form inertia_upw (field w, trial u, test v,
3   quadrature_option qopt = quadrature_option())
4 {
5   return integrate ("internal_sides",
6     0.5*abs(dot(average(w), normal()))*dot(jump(u), jump(v)));
7 }
8 form d_inertia_upw (field w, trial dw, field u, test v,
9   quadrature_option qopt = quadrature_option())
10 {
11   return integrate ("internal_sides",
12     0.5*compose (sgn, dot(average(w), normal())))
13     *dot(average(dw), normal())*dot(jump(u), jump(v)));
14 }

```

A multi-valued jacobian  $F'$  is then defined:

$$\begin{aligned} & \langle F' \left( \mathbf{u}_h^{(k-1)}, p_h^{(k-1)} \right) \cdot (\delta \mathbf{u}_h, \delta p_h), (\mathbf{v}_h, q_h) \rangle \\ &= \text{Re} \left( \begin{array}{c} t_h(\delta \mathbf{u}_h; \mathbf{u}_h, \mathbf{v}_h) + t_h(\mathbf{u}_h; \delta \mathbf{u}_h, \mathbf{v}_h) + \frac{\partial s_h}{\partial \mathbf{w}_h}(\mathbf{u}_h; \mathbf{u}_h, \mathbf{v}_h) \cdot (\delta \mathbf{u}_h) + s_h(\mathbf{u}_h; \delta \mathbf{u}_h, \mathbf{v}_h) \\ 0 \end{array} \right) \\ &+ \begin{pmatrix} a_h(\delta \mathbf{u}_h, \mathbf{v}_h) & + & b_h(\mathbf{v}_h, \delta p_h) \\ b_h(\delta \mathbf{u}_h, q_h) & - & c_h(\delta p_h, q_h) \end{pmatrix} \end{aligned}$$

We are able to extend the Newton method to the  $F$  function that allows a multi-valued subdifferential  $F'$ . During iterations, we can choose any of the available directions in the subdifferential. One the possibilities is then to replace the multi-valued sign function by a single-value one:

$$\widetilde{\text{sgn}}(x) = \begin{cases} 1 & \text{when } x \geq 0 \\ -1 & \text{when } x < 0 \end{cases}$$

Example file 4.25: `sgn.icc`

```
1 Float sgn (Float x) { return (x >= 0) ? 1 : -1; }
```

Example file 4.26: `navier_stokes_upw_dg.h`

```
1 #include "navier_stokes_dg.h"
2 struct navier_stokes_upw_dg: navier_stokes_dg {
3     typedef Float float_type;
4     typedef navier_stokes_dg::value_type value_type;
5     navier_stokes_upw_dg (Float Re, const geo& omega, string approx);
6     value_type residue (const value_type& uh) const;
7     void update_derivative (const value_type& uh) const;
8 };
9 #include "navier_stokes_upw_dg.icc"
```

Example file 4.27: `navier_stokes_upw_dg.icc`

```
1 #include "inertia_upw.icc"
2 navier_stokes_upw_dg::navier_stokes_upw_dg (
3     Float Re1, const geo& omega, string approx)
4     : navier_stokes_dg (Re1, omega, approx) {}
5
6 navier_stokes_upw_dg::value_type
7 navier_stokes_upw_dg::residue (const value_type& xh) const {
8     trial u (Xh); test v (Xh);
9     form a = a0 + Re*( inertia (xh[0], u, v, qopt)
10        + inertia_upw (xh[0], u, v, qopt));
11     value_type mrh(2);
12     mrh[0] = a*xh[0] + b.trans_mult(xh[1]) - lh;
13     mrh[1] = b*xh[0] - c*xh[1] - kh;
14     return mrh;
15 }
16 void navier_stokes_upw_dg::update_derivative (const value_type& xh) const {
17     trial du (Xh); test v (Xh);
18     a1 = a0 + Re*( inertia (xh[0], du, v, qopt)
19        + inertia_upw (xh[0], du, v, qopt)
20        + inertia (du, xh[0], v, qopt)
21        + d_inertia_upw (xh[0], du, xh[0], v, qopt));
22     stokes1 = solver_abtb (a1.uu(), b.uu(), c.uu(), mp.uu());
23 }
```

The program `navier_stokes_cavity_newton_upw_dg.cc` is obtained by replacing in `navier_stokes_taylor_newton_dg.cc` the string `navier_stokes_dg` by `navier_stokes_upw_dg` (two occurrences: in the includes and then in the definition of  $F$ ). Also replace the include `taylor.icc` by `cavity_dg.icc` that defines the boundary conditions. The compilation and run are similar.

```
make navier_stokes_cavity_newton_upw_dg streamf_cavity
mkgeo_grid -t 80 > square.geo
./navier_stokes_cavity_newton_upw_dg square P1d 500 1e-15 100 > square-500.field
field square-500.field -proj -field | ./streamf_cavity | \
    field -bw -n-iso 30 -n-iso-negative 20 -
```

Computations for higher Reynolds numbers are performed by continuation. Starting from a previous computation at  $Re = 500$ , we compute it at  $Re = 1000$  as:

```
./navier_stokes_cavity_newton_upw_dg square P1d 1000 1e-15 100 square-500.field \
```

```
> square-1000.field
field square-1000.field -proj -field | ./streamf_cavity | \
field -bw -n-iso 30 -n-iso-negative 20 -
```

Then, for  $Re = 1500$ :

```
./navier_stokes_cavity_newton_upw_dg square P1d 1500 1e-15 100 square-1000.field \
> square-1500.field
field square-1500.field -proj -field | ./streamf_cavity | \
field -bw -n-iso 30 -n-iso-negative 20 -
```

By this way, computations of solutions can be performed until  $Re = 25\,000$  without problems. Note that, from  $Re \approx 10\,000$ , these solutions are no more stable with respect to time [92, chap. 6], but are valid solutions of the stationary Navier-Stokes problem and can be interpreted as time-averaged solutions.

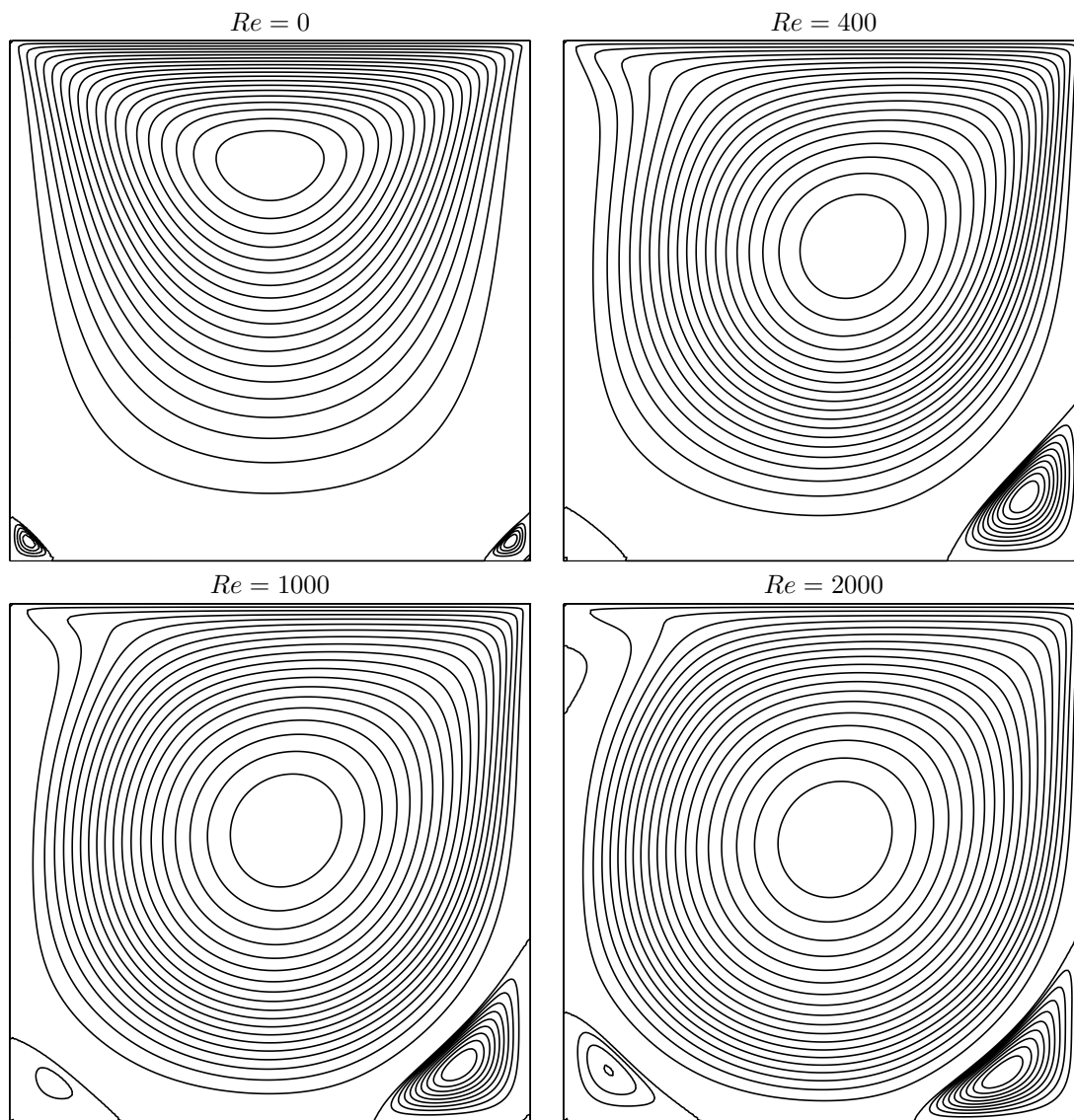


Figure 4.13: The discontinuous Galerkin method for the Navier-Stokes problem on the driven cavity benchmark when  $k = 1$  ( $80 \times 80$  grid): stream function isovalues for various  $Re$ .



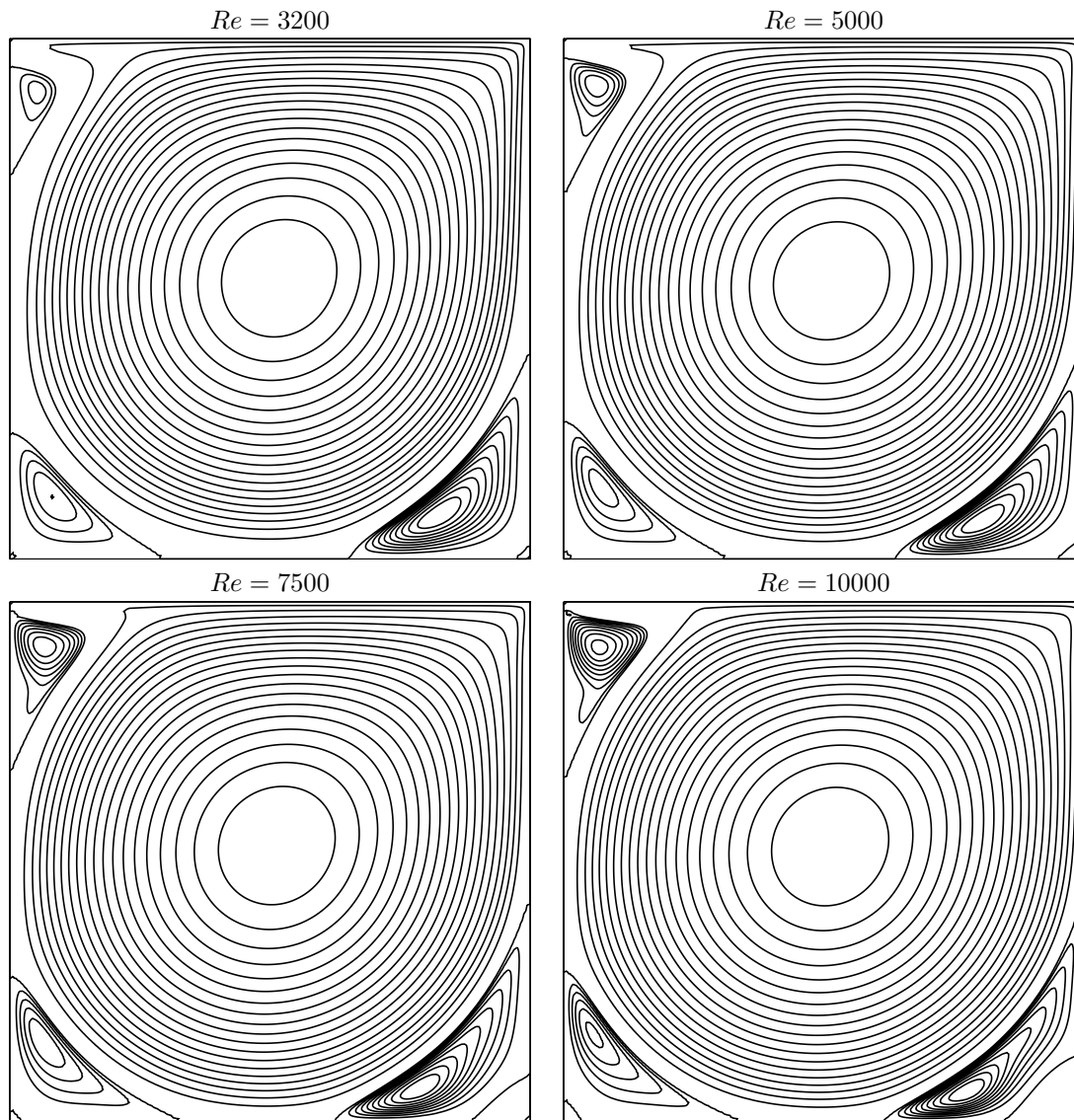


Figure 4.14: The discontinuous Galerkin method for the Navier-Stokes problem on the driven cavity benchmark when  $k = 1$  ( $80 \times 80$  grid): stream function isovalues for various  $Re$  (cont.).

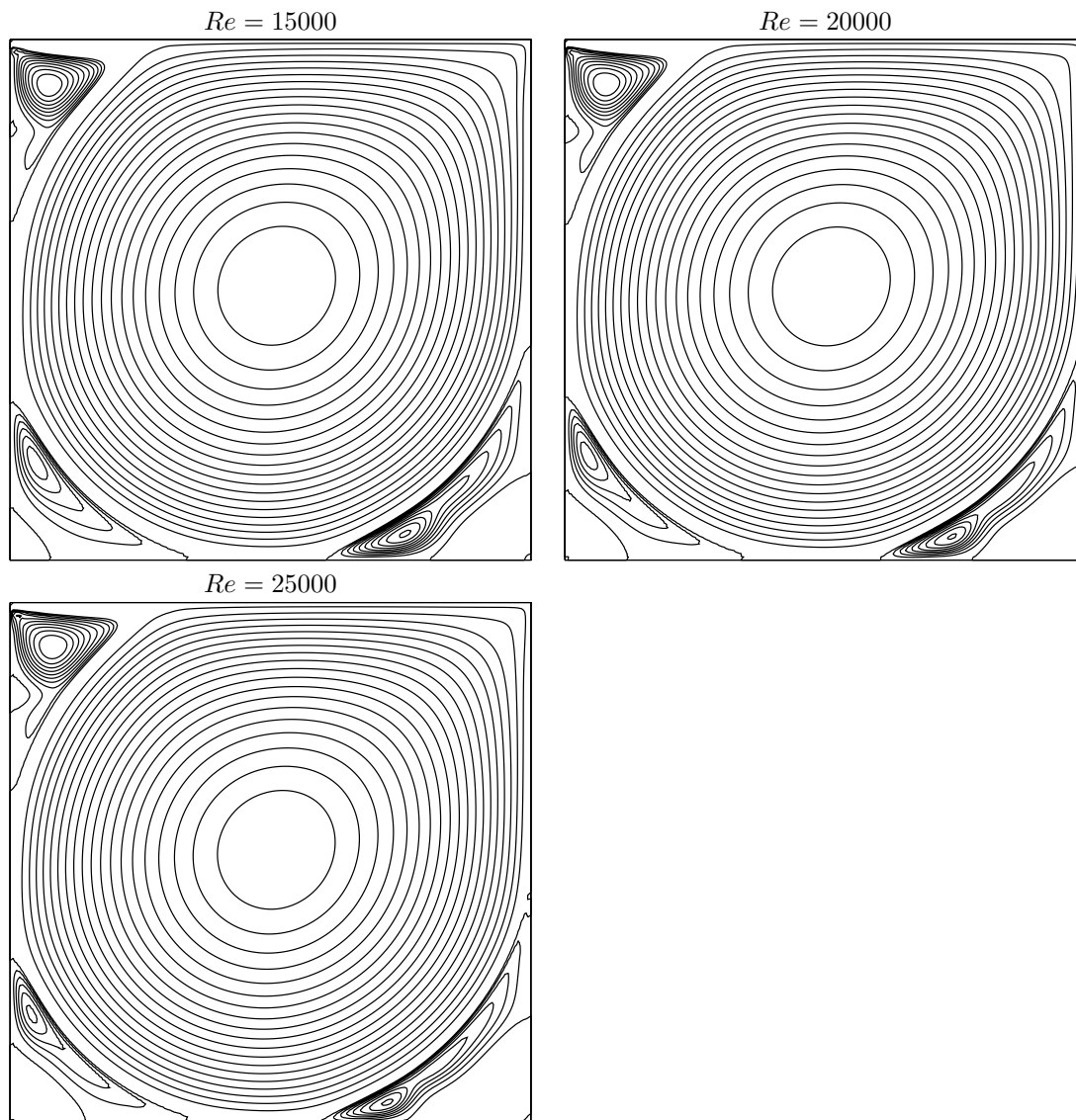


Figure 4.15: The discontinuous Galerkin method for the Navier-Stokes problem on the driven cavity benchmark when  $k = 1$  ( $80 \times 80$  grid): stream function isovalues for various  $Re$  (cont.).



# Chapter 5

## [New] Complex fluids

This part presents in details the practical computational aspects of numerical modeling with complex fluids. Most of the examples involve only few lines of code: the concision and readability of codes written with **Rheolef** is certainly a major key-point of this environment. The theoretical background for complex fluids and associated numerical methods can be found in [94]. We start with yield slip boundary condition as a preliminary problem. Slip at the wall occurs in many applications with complex fluids. This problem is solved both by augmented Lagrangian and Newton methods. Then, viscoplastic fluids are introduced and an augmented Lagrangian method is presented. A preliminary for viscoelastic fluids problems is the linear tensor transport equation: it is solved by a discontinuous Galerkin method. Finally, viscoelastic fluids problems are solved by an operator splitting algorithm, the  $\theta$ -scheme.

### 5.1 [New] Yield slip at the wall

#### 5.1.1 Problem statement

The problem of a Newtonian fluid with yield slip at the wall and flowing in a pipe [81] writes:

(P): find  $u$ , defined in  $\Omega$ , such that

$$-\Delta u = f \quad \text{in } \Omega \tag{5.1a}$$

$$\left. \begin{array}{l} \left| \frac{\partial u}{\partial n} \right| \leq S \quad \text{when } u = 0 \\ \frac{\partial u}{\partial n} = -C_f |u|^{-1+n} u - S \frac{u}{|u|} \quad \text{otherwise} \end{array} \right\} \quad \text{on } \partial\Omega \tag{5.1b}$$

Here,  $S \geq 0$  and  $C_f > 0$  are respectively the yield slip and the friction coefficient while  $n > 0$  is a power-law index. The computational domain  $\Omega$  represents the cross-section of the pipe and  $u$  is the velocity component along the axis of the pipe. The right-hand side  $f$  is a given constant, and without loss of generality, we can suppose  $f = 1$ : the parameters are  $S, C_f$  and  $n$ . When  $S = 0$  and  $n = 1$ , the problem reduces to a Poisson problem with homogeneous Robin boundary condition that depend upon  $C_f$ .

#### 5.1.2 The augmented Lagrangian algorithm

##### Principle of the algorithm

The problem writes as a minimization one:

$$\min_{u \in H^1(\Omega)} J(u)$$

where

$$J(u) = \frac{C_f}{1+n} \int_{\partial\Omega} |u|^{1+n} ds + S \int_{\partial\Omega} |u| ds + \frac{1}{2} \int_{\Omega} |\nabla u|^2 dx - \int_{\Omega} f u dx$$

This problem is solved by using an augmented Lagrangian algorithm. The auxiliary variable  $\gamma = u|_{\partial\Omega}$  is introduced together with the Lagrangian multiplier  $\lambda$  associated to the constraints  $u|_{\partial\Omega} - \gamma = 0$ . For all  $r > 0$ , let:

$$\begin{aligned} L((u, \gamma); \lambda) &= \frac{C_f}{1+n} \int_{\partial\Omega} |\gamma|^{1+n} ds + S \int_{\partial\Omega} |\gamma| ds + \frac{1}{2} \int_{\Omega} |\nabla u|^2 dx - \int_{\Omega} f u dx \\ &+ \int_{\partial\Omega} \lambda (u - \gamma) ds + \frac{r}{2} \int_{\partial\Omega} |u - \gamma|^2 ds \end{aligned}$$

An Uzawa-like minimization algorithm writes:

- $k = 0$ : let  $\lambda^{(0)}$  and  $\gamma^{(0)}$  arbitrarily chosen.
- $k \geq 0$ : let  $\lambda^{(k)}$  and  $\gamma^{(k)}$  being known.

$$\begin{aligned} u^{(k+1)} &:= \arg \min_{v \in H^1(\Omega)} L((v, \gamma^{(k)}); \lambda^{(k)}) \\ \gamma^{(k+1)} &:= \arg \min_{\delta \in L^\infty(\partial\Omega)} L((u^{(k+1)}, \delta); \lambda^{(k)}) \\ \lambda^{(k+1)} &:= \lambda^{(k)} + \rho (u^{(k+1)} - \gamma^{(k+1)}) \quad \text{on } \partial\Omega \end{aligned}$$

The descent step  $\rho$  is chosen as  $\rho = r$  for sufficiently large  $r$ . The Lagrangian  $L$  is quadratic in  $u$  and thus the computation of  $u^{(k+1)}$  reduces to a linear problem. The non-linearity is treated when computing  $\gamma^{(k+1)}$ . This operation is performed point-by-point on  $\partial\Omega$  by minimizing:

$$\gamma := \arg \min_{\delta \in \mathbb{R}} \frac{C_f |\delta|^{1+n}}{1+n} + \frac{r |\delta|^2}{2} + S |\delta| - \xi \delta$$

where  $\xi = \lambda^{(k)} + r u^{(k+1)}$  is given. This problem is convex and its solution is unique. The solution has the form:

$$\gamma = P_{n,r}(\xi) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{when } |\xi| \leq S \\ \phi_{n,r}(|\xi| - S) \operatorname{sgn}(\xi) & \text{otherwise} \end{cases} \quad (5.2)$$

where  $\phi_{n,r}(x) = f_{n,r}^{-1}(x)$  and  $f_{n,r}$  is defined for all  $y > 0$  by:

$$f_{n,r}(y) = C_f y^n + r y \quad (5.3)$$

#### Example file 5.1: projection.icc

```

1 #include "phi.icc"
2 // p(x) = phi(|x|-a)*sgn(x)
3 struct projection {
4     Float operator() (const Float& x) const {
5         if (fabs(x) <= a) return 0;
6         return (x > 0) ? _phi(x-a) : -_phi(-x-a);
7     }
8     projection (Float a1, Float n=1, Float c=1, Float r=0)
9         : a(a1), _phi(n,c,r) {}
10    Float a;
11    phi _phi;
12 };

```

Observe that  $f'_{n,r}(y) = nC_f y^{-1+n} + r > 0$  when  $r > 0$ : the function is strictly increasing and is thus invertible. When  $n = 1$  then  $f_{n,r}$  is linear and  $\phi_{1,r}(x) = x/(C_f + r)$ . When  $n = 1/2$  this

problem reduces to a second order polynomial equation and the solution is also explicit:

$$\phi_{\frac{1}{2},r}(x) = \frac{1}{4} \left( \left\{ 1 + \frac{4rx}{C_f^2} \right\}^{1/2} - 1 \right)^2$$

When  $r = 0$ , for any  $n > 0$  the solution is  $\phi_{n,0}(x) = (x/C_f)^{\frac{1}{n}}$ . In general, when  $n > 0$  and  $r > 0$ , the solution is no more explicit. We consider the Newton method:

- $i = 0$ : Let  $y_0$  being given.
- $i \geq 0$ : Suppose  $y_i$  being known and compute  $y_{i+1} = y_i - (f_{n,r}(y_i) - x)/f'_{n,r}(y_i)$ .

Example file 5.2: phi.icc

```

1 struct phi {
2   phi (Float n1=2, Float c1=1, Float r1=0) : n(n1), c(c1), r(r1) {}
3   Float operator() (const Float& x) const {
4     if (x <= 0) return 0;
5     if (n == 1) return x/(c+r);
6     if (r == 0) return pow(x/c,1/n);
7     Float y = x/(c+r);
8     const Float tol = numeric_limits<Float>::epsilon();
9     for (size_t i = 0; true; ++i) {
10      Float ry = f(y)-x;
11      Float dy = -ry/df_dy(y);
12      if (fabs(ry) <= tol && fabs(dy) <= tol) break;
13      if (i >= max_iter) break;
14      if (y+dy > 0) {
15        y += dy;
16      } else {
17        y /= 2;
18        check_macro (1+y != y, "phi: machine precision problem");
19      }
20    }
21    return y;
22  }
23  Float derivative (const Float& x) const {
24    Float phi_x = operator()(x);
25    return 1/(r + n*c*pow(phi_x,-1+n));
26  }
27 protected:
28  Float f(Float y) const { return c*pow(y,n) + r*y; }
29  Float df_dy(Float y) const { return n*c*pow(y,-1+n) + r; }
30  Float n,c,r;
31  static const size_t max_iter = 100;
32 };

```

In the present implementation, in order to avoid too large steps, the Newton step is damped when  $y_{i+1}$  becomes negative.

The Uzawa algorithm writes:

- $k = 0$ : let  $\lambda^{(0)}$  and  $\gamma^{(0)}$  arbitrarily chosen.
- $k \geq 0$ : let  $\lambda^{(k)}$  and  $\gamma^{(k)}$  being known, find  $u^{(k+1)}$ , defined in  $\Omega$ , such that

$$\begin{aligned} -\Delta u^{(k+1)} &= f \text{ in } \Omega \\ \frac{\partial u^{(k+1)}}{\partial n} + r u^{(k+1)} &= r \gamma^{(k)} - \lambda^{(k)} \text{ on } \partial\Omega \end{aligned}$$

and then compute explicitly  $\gamma^{(k+1)}$  and  $\lambda^{(k+1)}$ :

$$\begin{aligned} \gamma^{(k+1)} &:= P_{n,r} \left( \lambda^{(k)} + r u^{(k+1)} \right) \text{ on } \partial\Omega \\ \lambda^{(k+1)} &:= \lambda^{(k)} + r \left( u^{(k+1)} - \gamma^{(k+1)} \right) \text{ on } \partial\Omega \end{aligned}$$

This algorithm reduces the nonlinear problem to a sequence of linear and completely standard Poisson problems with a Robin boundary condition and some explicit computations. At convergence,  $\lambda = -\frac{\partial u}{\partial n}$  and  $\gamma = u$  on  $\partial\Omega$ .

Notice that the solution satisfies the following variational formulation:

$$\int_{\Omega} \nabla u \cdot \nabla v \, dx + \int_{\partial\Omega} v \lambda \, ds = \int_{\Omega} f v \, dx, \quad \forall v \in H^1(\Omega)$$

$$\int_{\partial\Omega} u \gamma \, ds - \int_{\partial\Omega} P_{n,0}(\lambda) \gamma \, ds = 0, \quad \forall \gamma \in L^\infty(\partial\Omega)$$

This formulation is the base of the computation of the residual test used for the stopping criteria.

Example file 5.3: yield\_slip\_augmented\_lagrangian.cc

```

1 #include "rheolef.h"
2 using namespace std;
3 using namespace rheolef;
4 #include "yield_slip_augmented_lagrangian.icc"
5 #include "poisson_robin.icc"
6 int main(int argc, char**argv) {
7     environment rheolef (argc,argv);
8     dlog << noverbose;
9     geo omega (argv[1]);
10    string approx = (argc > 2) ? argv[2] : "P1";
11    Float S = (argc > 3) ? atof(argv[3]) : 0.6;
12    Float n = (argc > 4) ? atof(argv[4]) : 1;
13    Float Cf = (argc > 5) ? atof(argv[5]) : 1;
14    Float r = (argc > 6) ? atof(argv[6]) : 1;
15    Float tol = 1e3*numeric_limits<Float>::epsilon();
16    size_t max_iter = 100000;
17    space Xh (omega, approx);
18    test v (Xh);
19    field lh = integrate(v);
20    field uh = poisson_robin (Cf, omega["boundary"], lh);
21    space Wh (omega["boundary"], Xh.get_approx());
22    field lambda_h = Cf*uh["boundary"];
23    int status = yield_slip_augmented_lagrangian(S, n, Cf, omega["boundary"],
24        lh, lambda_h, uh, tol, max_iter, r);
25    dout << setprecision(numeric_limits<Float>::digits10)
26        << catchmark("S") << S << endl
27        << catchmark("n") << n << endl
28        << catchmark("Cf") << Cf << endl
29        << catchmark("r") << r << endl
30        << catchmark("u") << uh
31        << catchmark("lambda") << lambda_h;
32    return status;
33 }

```

Example file 5.4: poisson\_robin.icc

```

1 field poisson_robin (Float Cf, const geo& boundary, const field& lh) {
2     const space& Xh = lh.get_space();
3     trial u (Xh); test v (Xh);
4     form a = integrate(dot(grad(u), grad(v))) + Cf*integrate(boundary, u*v);
5     solver sa (a.uu());
6     field uh (Xh);
7     uh.set_u() = sa.solve (lh.u());
8     return uh;
9 }

```

Example file 5.5: yield\_slip\_augmented\_lagrangian.icc

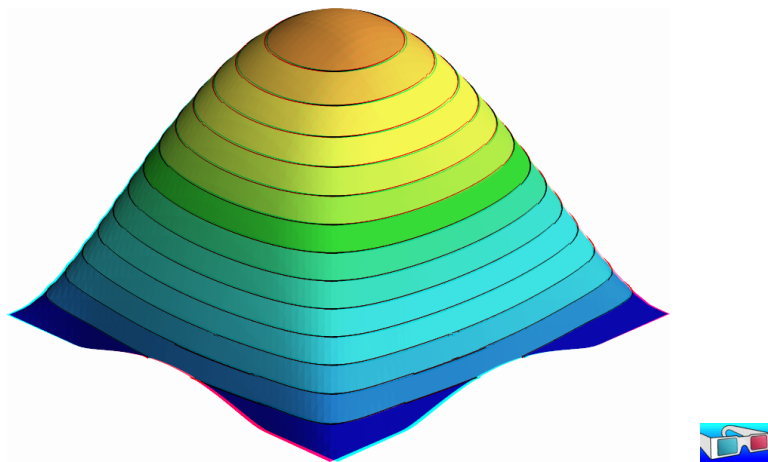
```

1 #include "projection.icc"
2 int yield_slip_augmented_lagrangian (Float S, Float n, Float Cf,
3   geo boundary, field lh, field& lambda_h, field& uh,
4   Float tol, size_t max_iter, Float r)
5 {
6   const space& Xh = uh.get_space();
7   const space& Wh = lambda_h.get_space();
8   trial u(Xh), lambda(Wh);
9   test v(Xh), mu(Wh);
10  form m = integrate (u*v),
11        a0 = integrate (dot(grad(u),grad(v))),
12        a = a0 + integrate (boundary, r*u*v),
13        mb = integrate (lambda*mu),
14        b = integrate (boundary, u*mu);
15  solver sa (a.uu()), sm (m.uu());
16  derr << "# k residue" << endl;
17  Float residue0 = 0;
18  for (size_t k = 0; true; ++k) {
19    field gamma_h = interpolate(Wh,
20      compose(projection(S,n,Cf,r), lambda_h + r*uh["boundary"]));
21    field delta_lambda_h = r*(uh["boundary"] - gamma_h);
22    lambda_h += delta_lambda_h;
23    Float residue = delta_lambda_h.max_abs();
24    derr << k << " " << residue << endl;
25    if (residue <= tol || k >= max_iter) return (residue <= tol) ? 0 : 1;
26    field rhs = lh + b.trans_mult(r*gamma_h - lambda_h);
27    uh.set_u() = sa.solve (rhs.u());
28  }
29 }

```

Observe also that the stopping criterion for breaking the loop bases on the max of the relative error for the  $\lambda_h$  variable. For this algorithm, this stopping criterion guaranties that all residual terms of the initial problem are also converging to zero, as it will be checked here. Moreover, this stopping criterion is very fast to compute while the full set of residual terms of the initial problem would take more computational time inside the loop.

### Running the program

Figure 5.1: The yield slip problem for  $S = 0.6$  and  $n = 1$ .

Assume that the previous code is contained in the file 'yield-slip-augmented-lagrangian.cc'. Compile the program as usual, using a Makefile suitable for rheolef (see [92, chap. 1]):



```

make yield_slip_augmented_lagrangian
mkgeo_grid -a -1 -b 1 -c -1 -d 1 -t 50 > square.geo
./yield_slip_augmented_lagrangian square.geo P1 0.6 1 > square.field

```

Also you can replace P1 by P2. The solution can be represented in elevation view (see Fig. 5.1):

```
field square.field -elevation -stereo
```

As analysed in [81], when  $S \leq 0.382$ , the fluid slips at the wall, when  $0.382 < S < 0.674$ , the fluid partially sticks and when  $S \geq 0.674$  the fluid completely sticks. Remark that the velocity is not zero along the boundary: there is a stick-slip transition point. The velocity along along the  $0x_0$

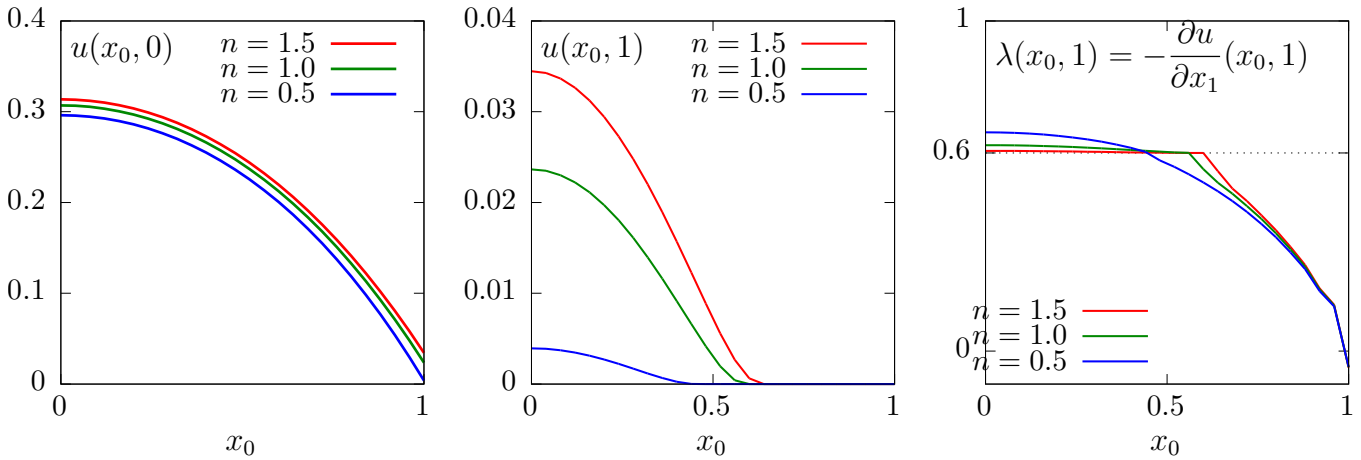


Figure 5.2: The yield slip problem for  $S = 0.6$ : cut of the velocity (left) along the  $0x_0$  axis ; (center) along the boundary ; (right) cut of the normal stress  $\lambda$  along the boundary.

axis and the top boundary are available as (see Fig. 5.2):

```

field square.field -normal 0 1 -origin 0 0 -cut -gnuplot
field square.field -domain top -elevation -gnuplot

```

The corresponding Lagrange multiplier  $\lambda$  on the boundary can also be viewed as:

```
field square.field -mark lambda -elevation
```

The file ‘yield\_slip\_residue.cc’ implement the computation of the full set of residual terms of the initial problem. This file it is not listed here but is available in the **Rheolef** example directory. The computation of residual terms is obtained by:

```

make yield_slip_residue
./yield_slip_residue < square.field

```

Observe that the residual terms of the initial problem are of about  $10^{-10}$ , as required by the stopping criterion. Fig. 5.3 plots the max of the relative error for the  $\lambda_h$  variable: this quantity is used as stopping criterion. Observe that it behaves asymptotically as  $1/k$  for larges meshes, with a final acceleration to machine precision. Note that these convergence properties could be dramatically improved by using a Newton method, as shown in the next section.

### 5.1.3 Newton algorithm

#### Reformulation of the problem

The idea of this algorithm first appears in [3] in the context of contact and friction problems. At convergence, the augmented Lagrangian method solve the following problem:

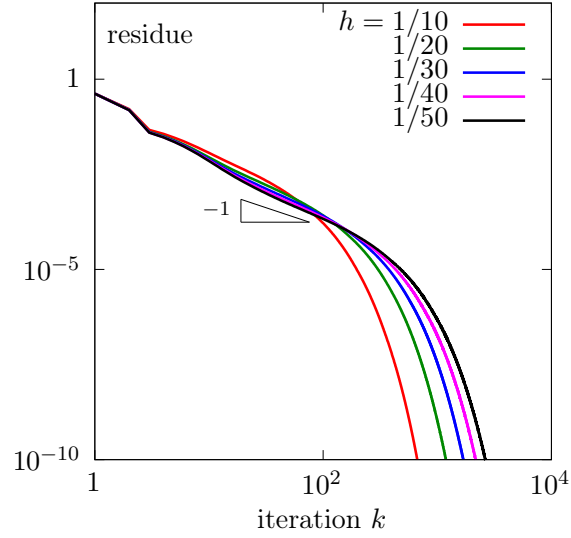


Figure 5.3: The convergence of the augmented Lagrangian algorithm for the yield slip problem with  $S = 0.6$  and  $n = 1$  and  $P_1$  polynomial approximation.

$(P)_r$ : find  $u$ , defined in  $\Omega$ , and  $\lambda$ , defined on  $\partial\Omega$ , such that

$$\begin{aligned} -\Delta u &= f \text{ in } \Omega \\ \frac{\partial u}{\partial n} + \lambda &= 0 \text{ on } \partial\Omega \\ u - P_{n,r}(\lambda + r u) &= 0 \text{ on } \partial\Omega \end{aligned}$$

The solution is independent upon  $r \in \mathbb{R}$  and this problem is equivalent to the original one. In order to diagonalize the non-linearity in  $P_{n,r}(\cdot)$ , let us introduce  $\beta = \lambda + r u$ . The problem becomes:

$(P)_r$ : find  $u$ , defined in  $\Omega$ , and  $\beta$ , defined on  $\partial\Omega$ , such that

$$\begin{aligned} -\Delta u &= f \text{ in } \Omega \\ \frac{\partial u}{\partial n} - r u + \beta &= 0 \text{ on } \partial\Omega \\ u - P_{n,r}(\beta) &= 0 \text{ on } \partial\Omega \end{aligned}$$

### Variational formulation

Consider the following forms:

$$\begin{aligned} m(u, v) &= \int_{\Omega} u v \, dx, \quad \forall u, v \in L^2(\Omega) \\ a(u, v) &= \int_{\Omega} \nabla u \cdot \nabla v \, dx - r \int_{\partial\Omega} u v \, ds, \quad \forall u, v \in H^1(\Omega) \\ b(v, \gamma) &= \int_{\Omega} v \gamma \, ds, \quad \forall \gamma \in L^2(\partial\Omega), \quad \forall v \in H^1(\Omega) \\ c(\beta, \gamma) &= \int_{\Omega} P_{n,r}(\beta) \gamma \, ds, \quad \forall \beta \in L^\infty(\partial\Omega), \quad \forall \gamma \in L^2(\partial\Omega) \end{aligned}$$

Remark that, since  $\Omega \subset \mathbb{R}^2$ , from the Sobolev imbedding theorem, if  $u \in H^1(\Omega)$  then  $u|_{\partial\Omega} \in L^\infty(\partial\Omega)$ . Then, all integrals have sense. The variational formulation writes:

(FV): find  $u \in H^1(\Omega)$  and  $\beta \in L^\infty(\partial\Omega)$  such that

$$\begin{aligned} a(u, v) + b(v, \beta) &= m(f, v), \quad \forall v \in H^1(\Omega) \\ b(u, \gamma) - c(\beta, \gamma) &= 0, \quad \forall \gamma \in L^2(\partial\Omega) \end{aligned}$$

Let  $M$ ,  $A$  and  $B$  the operators associated to forms  $m$ ,  $a$ ,  $b$  and  $c$ . The problem writes also as:

$$\begin{pmatrix} A & B^* \\ B & -C \end{pmatrix} \begin{pmatrix} u \\ \beta \end{pmatrix} = \begin{pmatrix} Mf \\ 0 \end{pmatrix}$$

where  $B^*$  denotes the formal adjoint of  $B$ . The bilinear form  $a$  is symmetric positive definite when  $r \in ]0, C_f[$ . Then  $A$  is non-singular and let  $A^{-1}$  denotes its inverse. The unknown  $u$  can be eliminated:  $u = A^{-1}(Mf - B^T\beta)$  and the problem reduces to:

$$\text{find } \beta \in L^\infty(\partial\Omega) \text{ such that } F(\beta) = 0$$

where

$$F(\beta) = C(\beta) + BA^{-1}B^*\beta - BA^{-1}Mf$$

This problem is a good candidate for a Newton method:

$$F'(\beta)\delta\beta = C'(\beta)\delta\beta + BA^{-1}B^*\delta\beta$$

where  $C'(\beta)$  is associated to the bilinear form:

$$c_1(\beta; \gamma, \delta) = \int_{\Omega} P'_{n,r}(\beta) \gamma \delta \, ds, \quad \forall \beta \in L^\infty(\partial\Omega), \forall \gamma, \delta \in L^2(\partial\Omega)$$

where, for all  $\xi \in \mathbb{R}$ :

$$P'_{n,r}(\xi) = \begin{cases} 0 & \text{when } |\xi| \leq S \\ \phi'_{n,r}(|\xi| - S) & \text{otherwise} \end{cases} \quad (5.4)$$

Recall that, for all  $\zeta > 0$ ,  $\phi_{n,r}(\zeta) = f_{n,r}^{-1}(\zeta)$  where  $f_{n,r}(y) = C_f y^n + r y$ , for all  $y > 0$ . Then

$$\phi'_{n,r}(\zeta) = \frac{1}{f'_{n,r}(f_{n,r}^{-1}(\zeta))} = \frac{1}{f'_{n,r}(\phi_{n,r}(\zeta))} = \frac{1}{r + n C_f \{\phi_{n,r}(\zeta)\}^{-1+n}} \quad (5.5)$$

When  $n = 1$  we simply have:  $\phi'_{1,r}(\zeta) = \frac{1}{C_f + r}$ . When  $r = 0$ , for any  $n > 0$  we have

$$\phi'_{n,0}(\zeta) = \frac{\zeta^{-1+1/n}}{n C_f^{1/n}}.$$

Example file 5.6: d\_projection\_dx.icc

```

1 #include "projection.icc"
2 struct d_projection_dx {
3   Float operator() (const Float& x) const {
4     if (fabs(x) <= a) return 0;
5     if (n == 1) return 1/(c + r);
6     if (r == 0) return pow(fabs(x)-a, -1+1/n)/(n*pow(c,1/n));
7     return 1/(r + n*c*pow(_phi(fabs(x)-a), -1+n));
8   }
9   d_projection_dx (Float a1, Float n1=1, Float c1=1, Float r1=0)
10    : a(a1), n(n1), c(c1), r(r1), _phi(n1,c1,r1) {}
11   Float a,n,c,r;
12   phi _phi;
13 };

```

Example file 5.7: yield\_slip\_damped\_newton.cc

```

1 #include "rheolef.h"
2 using namespace std;
3 using namespace rheolef;
4 #include "yield_slip.h"
5 int main(int argc, char**argv) {
6     environment rheolef (argc,argv);
7     geo omega (argv[1]);
8     string approx = (argc > 2) ? argv[2] : "P1";
9     Float S = (argc > 3) ? atof(argv[3]) : 0.6;
10    Float n = (argc > 4) ? atof(argv[4]) : 1;
11    Float Cf = (argc > 5) ? atof(argv[5]) : 1;
12    Float r = (argc > 6) ? atof(argv[6]) : 1;
13    domain boundary = omega["boundary"];
14    yield_slip F (S, n, Cf, r, omega, boundary, approx);
15    field beta_h = F.initial();
16    Float tol = 10*numeric_limits<Float>::epsilon();
17    size_t max_iter = 10000;
18    int status = damped_newton (F, beta_h, tol, max_iter, &derr);
19    field uh, lambda_h;
20    F.post (beta_h, uh, lambda_h);
21    dout << setprecision(numeric_limits<Float>::digits10)
22         << catchmark("S") << S << endl
23         << catchmark("n") << n << endl
24         << catchmark("Cf") << Cf << endl
25         << catchmark("r") << r << endl
26         << catchmark("u") << uh
27         << catchmark("lambda") << lambda_h;
28    return status;
29 }

```

Example file 5.8: yield\_slip.h

```

1 class yield_slip {
2 public:
3     typedef field value_type;
4     typedef Float float_type;
5     yield_slip (Float S, Float n, Float Cf, Float r,
6               const geo& omega, const geo& boundary, string approx = "P1");
7     field residue (const field& beta_h) const;
8     void update_derivative (const field& beta_h) const;
9     field derivative_solve (const field& mrh) const;
10    field derivative_trans_mult (const field& mrh) const;
11    Float space_norm (const field&) const;
12    Float dual_space_norm (const field&) const;
13    field initial () const;
14    void post (const field& beta_h, field& uh, field& lambda_h) const;
15 protected:
16    Float S, n, Cf, r;
17    geo boundary;
18    space Xh, Wh, Yh;
19    field lh, mkh;
20    form m, mb, a, b;
21    mutable form c1;
22    solver smb, sa;
23    mutable solver sA;
24 };
25 #include "yield_slip1.icc"
26 #include "yield_slip2.icc"

```

Example file 5.9: yield\_slip1.icc

```

1 #include "d_projection_dx.icc"
2 yield_slip::yield_slip (Float S1, Float n1, Float Cf1, Float r1,
3   const geo& omega, const geo& boundary1, string approx)
4   : S(S1), n(n1), Cf(Cf1), r(r1), boundary(boundary1), Xh(), Wh(), Yh(),
5     lh(), mkh(), m(), mb(), a(), b(), c1(), smb(), sa(), sA()
6   {
7     Xh = space (omega, approx);
8     Wh = space (boundary, approx);
9     Yh = Xh*Wh;
10    trial u (Xh), lambda(Wh);
11    test v (Xh), mu(Wh);
12    m = integrate(u*v);
13    mb = integrate(lambda*mu);
14    a = integrate(dot(grad(u),grad(v))) - r*integrate(boundary, u*v);
15    b = integrate(boundary, u*mu);
16    lh = integrate(v);
17    smb = solver (mb.uu());
18    sa = solver (a.uu());
19    field vh(Xh);
20    vh.set_u() = sa.solve (lh.u());
21    mkh = b*vh;
22  }
23  field yield_slip::residue (const field& beta_h) const {
24    field vh (Xh);
25    field rhs = b.trans_mult (beta_h);
26    vh.set_u() = sa.solve (rhs.u());
27    test mu (Wh);
28    field c0h = integrate(mu*compose(projection(S,n,Cf,r), beta_h));
29    field mrh = b*vh + c0h - mkh;
30    return mrh;
31  }
32  void yield_slip::update_derivative (const field& beta_h) const {
33    trial lambda (Wh); test mu (Wh);
34    c1 = integrate (lambda*mu*compose(d_projection_dx(S,n,Cf,r), beta_h));
35    csr<Float> A = { { a.uu(), trans(b.uu()) },
36                  { b.uu(), -c1.uu() } };
37    A.set_symmetry (c1.uu().is_symmetric());
38    sA = solver(A);
39  }

```

Example file 5.10: yield\_slip2.icc

```

1 #include "poisson_robin.icc"
2 field yield_slip::derivative_solve (const field& mrh) const {
3   field mryh(Yh, 0.);
4   mryh[1] = -mrh;
5   field delta_yh(Yh);
6   delta_yh.set_u() = sA.solve(mryh.u());
7   return delta_yh[1];
8 }
9 field yield_slip::derivative_trans_mult (const field& mrh) const {
10  field rh (Wh);
11  rh.set_u() = smb.solve(mrh.u());
12  field rhs = b.trans_mult(rh);
13  field delta_vh (Xh);
14  delta_vh.set_u() = sa.solve(rhs.u());
15  delta_vh.set_b() = 0;
16  field mgh = b*delta_vh + c1*rh;
17  field gh (Wh);
18  gh.set_u() = smb.solve(mgh.u());
19  return gh;
20 }
21 Float yield_slip::space_norm (const field& rh) const {
22   return sqrt (mb(rh,rh));
23 }
24 Float yield_slip::dual_space_norm (const field& mrh) const {
25   field rh (Wh,0);
26   rh.set_u() = smb.solve (mrh.u());
27   return sqrt (dual (mrh,rh));
28 }
29 field yield_slip::initial () const {
30   field uh = poisson_robin (Cf, boundary, lh);
31   return (Cf+r)*uh["boundary"];
32 }
33 void yield_slip::post (const field& beta_h, field& uh, field& lambda_h) const {
34   field rhs = lh - b.trans_mult(beta_h);
35   uh = field (Xh);
36   uh.set_u() = sa.solve(rhs.u());
37   lambda_h = beta_h - r*uh["boundary"];
38 }

```

## Running the program

```

make ./yield_slip_damped_newton
./yield_slip_damped_newton square.geo P1 0.6 1 > square.field
field square.field -elevation -stereo
field square.field -mark lambda -elevation

```

Observe on Fig. 5.4.a and 5.4.b that the convergence is super-linear and mesh-independent when  $n = 1/2$  and  $n = 0.9$ . For mesh-independent convergence of the Newton method, see e.g. the  $p$ -Laplacian example in [92]. When  $n = 0.9$ , observe that the convergence depends slightly upon the mesh for rough meshes while it becomes asymptotically mesh independent for fine meshes. When  $n \geq 1$ , the convergence starts to depend also upon the mesh (Fig. 5.4.bottom-left and 5.4.bottom-right). Recall that when  $n \geq 1$ , the problem becomes non-differentiable and the convergence of the Newton method is no more assured. Nevertheless, in that case, the convergence is clearly faster (about 100 times faster) than the corresponding one with the augmented Lagrangian algorithm on the same problem (see Fig. 5.3, page 193) and moreover there is no saturation of the residual terms on large meshes.

### 5.1.4 Error analysis

Assume that the previous code is contained in the file ‘yield-slip-augmented-lagrangian.cc’. When  $\Omega$  is the unit circle, the exact solution is known. In polar coordinates  $(r, \theta)$ , as the solution

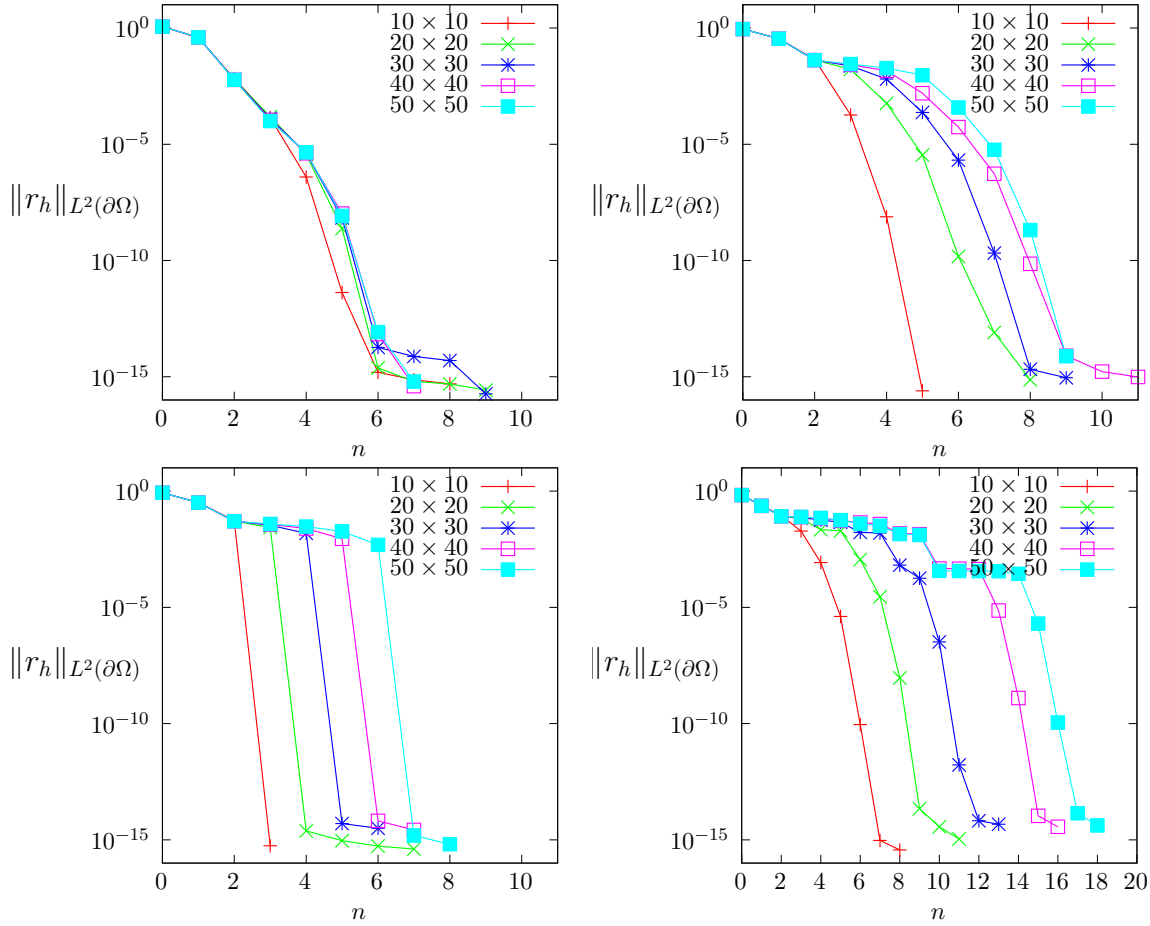


Figure 5.4: The convergence of the damped Newton algorithm for the yield slip problem ( $S = 0.6$ ): (top-left)  $n = 0.5$ ; (top-right)  $n = 0.9$ ; (bottom-left)  $n = 1$ ; (bottom-right)  $n = 1.5$ .

$u$  depends only of  $r$ , equation (5.1a) becomes:

$$-\frac{1}{r}\partial_r(r\partial_r u) = 1$$

and then, from the symmetry,  $u(r) = c - r^2/4$ , where  $c$  is a constant to determine from the boundary condition. On the boundary  $r = 1$  we have  $\partial_n u = \partial_r u = -1/2$ . When  $S \geq 1/2$  the fluid sticks at the wall and when  $S > 1/2$  we have from (5.1b):

$$-C_f u^n - S = \frac{1}{2}$$

From the previous expression  $u(r)$ , taken for  $r = 1$  we obtain the constant  $c$  and then:

$$u(r) = \frac{1-r^2}{4} + \left( \frac{\max(0, 1/2 - S)}{C_f} \right)^{1/n}, \quad 0 \leq r \leq 1$$

The error computation is implemented in the files ‘yield\_slip\_error.cc’ and ‘yield\_slip\_circle.icc’: it is not listed here but is available in the **Rheolef** example directory.

The error can be computed (see Fig. 5.5):

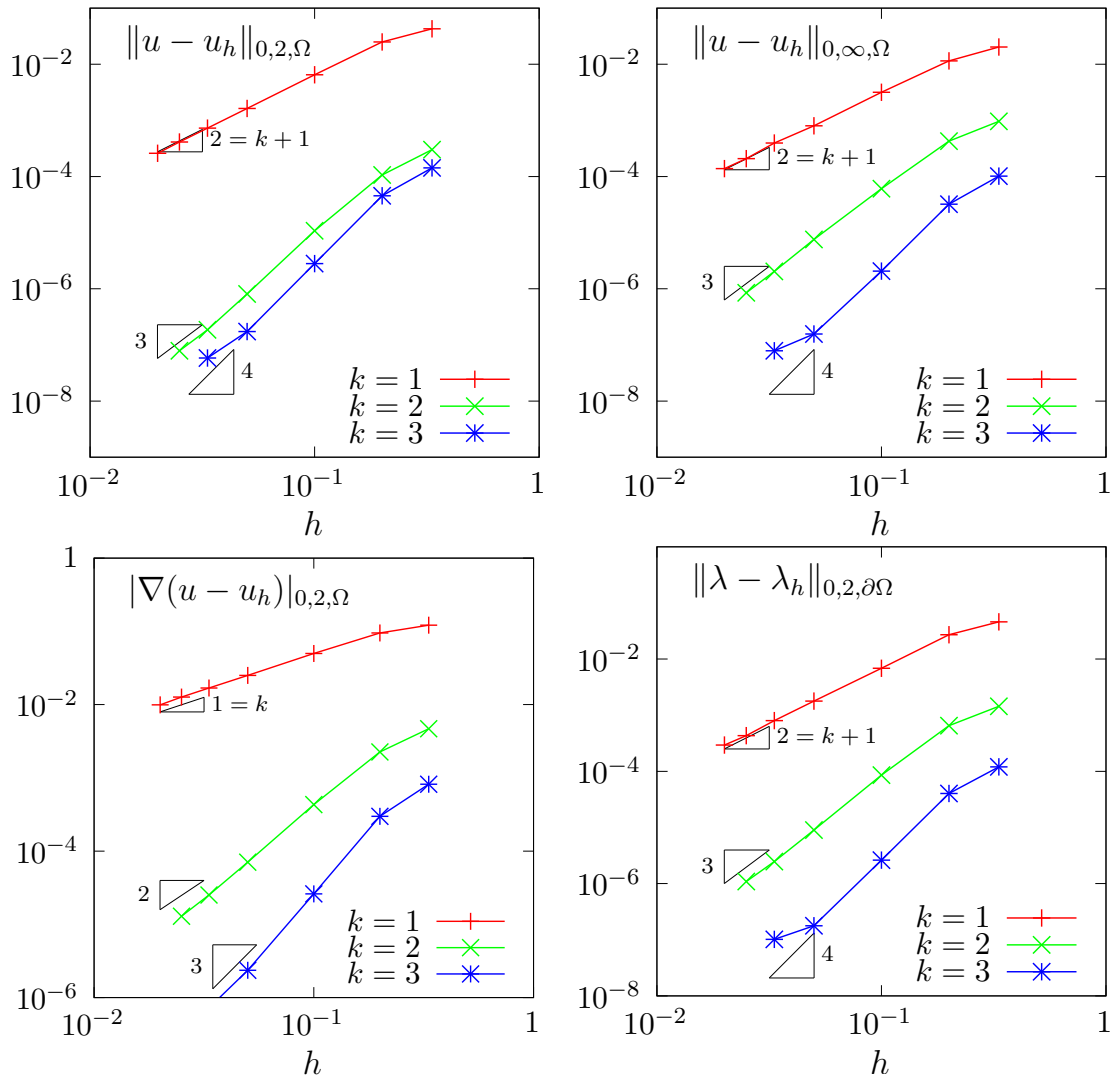


Figure 5.5: The yield slip problem: error analysis.

```

make yield_slip_error
mkgeo_ball -t 20 -order 2 > circle-20-P2.geo
./yield_slip_damped_newton circle-20-P2.geo P2 0.6 1 | ./yield_slip_error

```

It appears that the discrete formulation develops optimal convergence properties versus mesh refinement for  $k \geq 1$  in  $H^1$ ,  $L^2$  and  $L^\infty$  norms.



## 5.2 [New] Viscoplastic fluids

### 5.2.1 Problem statement

Viscoplastic fluids develops an yield stress behavior (see e.g. [94,97]. Mosolov and Miasnikov [58–60] first investigated the flow of a viscoplastic in a pipe with an arbitrarily cross section. Its numerical investigation by augmented Lagrangian methods was first performed in [82,96]. The Mosolov problem [96] writes:

(P): find  $\boldsymbol{\sigma}$  and  $u$ , defined in  $\Omega$ , such that

$$\operatorname{div} \boldsymbol{\sigma} = -f \quad \text{in } \Omega \quad (5.6a)$$

$$u = 0 \quad \text{on } \partial\Omega \quad (5.6b)$$

$$\left. \begin{array}{l} |\boldsymbol{\sigma}| \leq Bi \\ \boldsymbol{\sigma} = |\nabla u|^{-1+n} \nabla u + Bi \frac{\nabla u}{|\nabla u|} \end{array} \right\} \begin{array}{l} \text{when } \nabla u = 0 \\ \text{otherwise} \end{array} \quad \text{in } \Omega \quad (5.6c)$$

where  $Bi \geq 0$  is the Bingham number and  $n > 0$  is a power-law index. The computational domain  $\Omega$  represents the cross-section of the pipe. In the bidimensional case  $d = 2$  and when  $f$  is constant, this problem describes the stationary flow of an Herschel-Bulkley fluid in a general pipe section  $\Omega$ . Let  $Ox_2$  be the axis of the pipe and  $Ox_0x_1$  the plane of the section  $\Omega$ . The vector-valued field  $\boldsymbol{\sigma}$  represents the shear stress components  $(\sigma_{0,1}, \sigma_{0,2})$  while  $u$  is the axial component of velocity along  $Ox_3$ . When  $Bi = 0$ , the problem reduces to the nonlinear  $p$ -Laplacian problem. When  $n = 1$  the fluid is a Bingham fluid. When  $n = 1$  and  $Bi = 0$ , the problem reduces to the linear Poisson one with  $\boldsymbol{\sigma} = \nabla u$ .

### 5.2.2 The augmented Lagrangian algorithm

This problem writes as a minimization of an energy:

$$u = \arg \min_{v \in W_0^{1,p}(\Omega)} J(v) \quad (5.7a)$$

where

$$J(v) = \frac{1}{1+n} \int_{\Omega} |\nabla v|^{1+n} dx + Bi \int_{\Omega} |\nabla v| dx - \int_{\Omega} f v dx \quad (5.7b)$$

This problem is solved by using an augmented Lagrangian algorithm. The auxiliary variable  $\boldsymbol{\gamma} = \nabla u$  is introduced together with the Lagrangian multiplier  $\boldsymbol{\sigma}$  associated to the constraint  $\nabla u - \boldsymbol{\gamma} = 0$ . For all  $r > 0$ , let:

$$L((v, \boldsymbol{\gamma}); \boldsymbol{\sigma}) = \frac{1}{1+n} \int_{\Omega} |\boldsymbol{\gamma}|^{1+n} dx + Bi \int_{\Omega} |\boldsymbol{\gamma}| dx - \int_{\Omega} f v dx + \int_{\Omega} \boldsymbol{\sigma} \cdot (\nabla u - \boldsymbol{\gamma}) dx + \frac{r}{2} \int_{\Omega} |\nabla u - \boldsymbol{\gamma}|^2 dx$$

An Uzawa-like minimization algorithm writes:

- $k = 0$ : let  $\lambda^{(0)}$  and  $\boldsymbol{\gamma}^{(0)}$  arbitrarily chosen.
- $k \geq 0$ : let  $\lambda^{(k)}$  and  $\boldsymbol{\gamma}^{(k)}$  being known.

$$\begin{aligned} u^{(k+1)} &:= \arg \min_{v \in W^{1,p}(\Omega)} L((v, \boldsymbol{\gamma}^{(k)}); \boldsymbol{\sigma}^{(k)}) \\ \boldsymbol{\gamma}^{(k+1)} &:= \arg \min_{\boldsymbol{\delta} \in L^2(\Omega)^d} L((u^{(k+1)}, \boldsymbol{\delta}); \boldsymbol{\sigma}^{(k)}) \\ \boldsymbol{\sigma}^{(k+1)} &:= \boldsymbol{\sigma}^{(k)} + \rho \left( \nabla u^{(k+1)} - \boldsymbol{\gamma}^{(k+1)} \right) \quad \text{in } \Omega \end{aligned}$$

The descent step  $\rho$  is chosen as  $\rho = r$  for sufficiently large  $r$ . The Lagrangian  $L$  is quadratic in  $u$  and thus the computation of  $u^{(k+1)}$  reduces to a linear problem. The non-linearity is treated when computing  $\gamma^{(k+1)}$ . This operation is performed point-by-point in  $\Omega$  by minimizing:

$$\gamma := \arg \min_{\delta \in \mathbb{R}^d} \frac{|\delta|^{1+n}}{1+n} + \frac{r|\delta|^2}{2} + Bi|\delta| - \xi \cdot \delta$$

where  $\xi = \sigma^{(k)} + r \nabla u^{(k+1)}$  is given. This problem is convex and its solution is unique. The solution has the form:

$$\gamma = P_{n,r}(\xi) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{when } |\xi| \leq S \\ \phi_{n,r}(|\xi| - S) \frac{\xi}{|\xi|} & \text{otherwise} \end{cases} \quad (5.8)$$

where  $\phi_{n,r}(x) = f_{n,r}^{-1}(x)$  has been introduced in (5.2) page 188 in the context of the yield slip problem together with the scalar projector.

Example file 5.11: vector\_projection.icc

```

1 #include "phi.icc"
2 struct vector_projection {
3     Float operator() (const Float& x) const {
4         if (x <= a) return 0;
5         return _phi(x-a)/x;
6     }
7     vector_projection (Float a1, Float n=1, Float c=1, Float r=0)
8         : a(a1), _phi(n,c,r) {}
9     Float a;
10    phi _phi;
11 };

```

Finally, the Uzawa-like minimization algorithm [96] writes:

- $k = 0$ : let  $\sigma^{(0)}$  and  $\gamma^{(0)}$  arbitrarily chosen.
- $k \geq 0$ : let  $\sigma^{(k)}$  and  $\gamma^{(k)}$  being known, find  $u^{(k+1)}$  such that

$$\begin{aligned} -r\Delta u^{(k+1)} &= f + \operatorname{div}(\sigma^{(k)} - r\gamma^{(k)}) && \text{in } \Omega \\ u^{(k+1)} &= 0 && \text{on } \partial\Omega \end{aligned}$$

and then compute explicitly  $\gamma^{(k+1)}$  and  $\sigma^{(k+1)}$ :

$$\begin{aligned} \gamma^{(k+1)} &:= P_{n,r}(\sigma^{(k)} + r\nabla u^{(k+1)}) \\ \sigma^{(k+1)} &:= \sigma^{(k)} + r(\nabla u^{(k+1)} - \gamma^{(k+1)}) \end{aligned} \quad (5.9)$$

Here  $r > 0$  is a numerical parameter. This algorithm reduces the nonlinear problem to a sequence of linear and completely standard Poisson problems and some explicit computations. For convenience, this algorithm is implemented as the `solve` function member of a class:

Example file 5.12: mosolov\_augmented\_lagrangian1.icc

```

1 #include "vector_projection.icc"
2 int mosolov_augmented_lagrangian::solve (field& sigma_h, field& uh) const {
3     test v(Xh);
4     if (p_err) *p_err << "# k residue" << endl;
5     for (size_t k = 0; true; ++k) {
6         field grad_uh = inv_mt*(b*uh);
7         auto c = compose(vector_projection(Bi,n,1,r), norm(sigma_h+r*grad_uh));
8         field gamma_h = interpolate(Th, c*(sigma_h + r*grad_uh));
9         field delta_sigma_h = r*(grad_uh - gamma_h);
10        sigma_h += delta_sigma_h;
11        Float residue = delta_sigma_h.max_abs();
12        if (p_err) *p_err << k << " " << residue << endl;
13        if (residue <= tol || k >= max_iter) {
14            if (p_err) *p_err << endl << endl;
15            return (pow(residue,3) <= tol) ? 0 : 1;
16        }
17        field rhs = (1/r)*(lh - integrate(dot(sigma_h - r*gamma_h, grad(v))));
18        uh.set_u() = sa.solve (rhs.u() - a.ub()*uh.b());
19    }
20 }

```

For convenience, the order of the update of the three variables  $u$ ,  $\gamma$  and  $\sigma$  has been rotated: by this way, the algorithm starts with initial values for  $u$  and  $\sigma$ . instead of  $\gamma$  and  $\sigma$ . Observe that the projection step (5.9) is implemented by using the `interpolate` operator: this projection step interprets as point-wise at Lagrange nodes instead as a numerical resolution of the element-wise minimization problem. Note that, for the lowest order  $k = 1$ , these two approaches are strictly equivalent, while, when  $k \geq 2$ , the numerical solution obtained by this algorithm is no more solution of the discrete version of the saddle-point problem for the Lagrangian  $L$ . Nevertheless, the numerical solution is founded to converge to the exact solution of the initial problem (5.6a)-(5.6c): this will be checked here in a forthcoming section, dedicated to the error analysis. Observe also that the stopping criterion for breaking the loop bases on the max of the relative error for the  $\sigma_h$  variable. For this algorithm, this stopping criterion guaranties that all residual terms of the initial problem are also converging to zero, as it will be checked here. Moreover, this stopping criterion is very fast to compute while the full set of residual terms of the initial problem would take more computational time inside the loop.

The class declaration contains all model parameters, loop controls, form and space variables, together with some pre- and post-treatments:

Example file 5.13: mosolov\_augmented\_lagrangian.h

```

1 struct mosolov_augmented_lagrangian: adapt_option {
2     mosolov_augmented_lagrangian();
3     void reset (geo omega, string approx);
4     void initial (field& sigma_h, field& uh) const;
5     int solve (field& sigma_h, field& uh) const;
6     void put (odiststream& out, field& sigma_h, field& uh) const;
7     // data:
8     Float Bi, n, r, tol;
9     size_t max_iter;
10    odiststream* p_err;
11    mutable space Xh, Th;
12    mutable field lh;
13    mutable form a, b, inv_mt;
14    mutable solver sa;
15 };
16 #include "mosolov_augmented_lagrangian1.icc"
17 #include "mosolov_augmented_lagrangian2.icc"

```

Example file 5.14: mosolov\_augmented\_lagrangian2.icc

```

1 mosolov_augmented_lagrangian::mosolov_augmented_lagrangian()
2 : Bi(0), n(1), r(1), tol(1e-10), max_iter(1000000), p_err(&derr),
3   Xh(), Th(), lh(), a(), b(), inv_mt(), sa()
4 {}
5 void mosolov_augmented_lagrangian::reset (geo omega, string approx) {
6   Xh = space (omega, approx);
7   Xh.block ("boundary");
8   string grad_approx = "P" + itos(Xh.degree()-1) + "d";
9   Th = space (omega, grad_approx, "vector");
10  trial u (Xh), sigma(Th);
11  test v (Xh), tau (Th);
12  lh = integrate (2*v);
13  a = integrate (dot(grad(u),grad(v)));
14  b = integrate (dot(grad(u),tau));
15  integrate_option fopt;
16  fopt.invert = true;
17  inv_mt = integrate(dot(sigma,tau), fopt);
18  sa = solver(a.uu());
19 }
20 void
21 mosolov_augmented_lagrangian::initial (field& sigma_h, field& uh) const {
22   uh = field(Xh);
23   uh["boundary"] = 0;
24   uh.set_u() = sa.solve (lh.u() - a.uv()*uh.b());
25   test tau (Th);
26   field mt_grad_uh = integrate(dot(grad(uh),tau));
27   sigma_h = inv_mt*mt_grad_uh;
28 }
29 void mosolov_augmented_lagrangian::put (odiststream& out,
30   field& sigma_h, field& uh) const
31 {
32   out << catchmark("Bi") << Bi << endl
33   << catchmark("n") << n << endl
34   << catchmark("r") << r << endl
35   << catchmark("sigma") << sigma_h
36   << catchmark("u") << uh;
37 }

```

Example file 5.15: mosolov\_augmented\_lagrangian.cc

```

1 #include "rheolef.h"
2 using namespace std;
3 using namespace rheolef;
4 #include "mosolov_augmented_lagrangian.h"
5 int main(int argc, char**argv) {
6     environment rheolef (argc,argv);
7     mosolov_augmented_lagrangian pb;
8     geo omega (argv[1]);
9     string approx = (argc > 2) ? argv[2] : "P1";
10    pb.Bi = (argc > 3) ? atof(argv[3]) : 0.2;
11    pb.n = (argc > 4) ? atof(argv[4]) : 1;
12    size_t n_adapt = (argc > 5) ? atoi(argv[5]) : 0;
13    pb.max_iter = (argc > 6) ? atoi(argv[6]) : 10000;
14    pb.err = (argc > 7) ? atof(argv[7]) : 1e-4;
15    pb.r = 100;
16    pb.tol = 1e-10;
17    pb.hmin = 1e-4;
18    pb.hmax = 1e-1;
19    pb.ratio = 3;
20    pb.additional = "-AbsError";
21    field sigma_h, uh;
22    for (size_t i = 0; true; i++) {
23        pb.reset (omega, approx);
24        pb.initial (sigma_h, uh);
25        int status = pb.solve (sigma_h, uh);
26        odistream out (omega.name(), "field");
27        pb.put (out, sigma_h, uh);
28        if (i == n_adapt) break;
29        space T0h (sigma_h.get_geo(), sigma_h.get_approx());
30        field ch = interpolate (T0h, sqrt(abs(dot(sigma_h, grad(uh)))));
31        omega = adapt (ch, pb);
32        omega.save();
33    }
34 }

```

The main program read parameters from the command line and performs an optional mesh adaptation loop. This implementation supports any  $n > 0$ , any continuous piecewise polynomial  $P_k$ ,  $k \geq 1$  and also isoparametric approximations for curved boundaries.

### Running the program

Compile the program as usual:

```

make mosolov_augmented_lagrangian
mkgeo_grid -a -1 -b 1 -c -1 -d 1 -t 10 > square.geo
./mosolov_augmented_lagrangian square.geo P1 0.4 1
field -mark u square.field -elevation

```

Observe on Fig. 5.6.left the central region where the velocity is constant. A cut of the velocity field along the first bisector is obtained by:

```
field -mark u square.field -cut -origin 0 0 -normal 1 1 -gnuplot
```

Observe on Fig. 5.6.right the small regions with zero velocity, near the outer corner of the square pipe section. This region is really small but exists. This question will be revisited in the next section dedicated to auto-adaptive mesh refinement.

The file 'mosolov\_residue.cc' implement the computation of the full set of residual terms of the initial problem. This file it is not listed here but is available in the **Rheolef** example directory. The computation of residual terms is obtained by:

```

make mosolov_residue
zcat square.field.gz | ./mosolov_residue

```

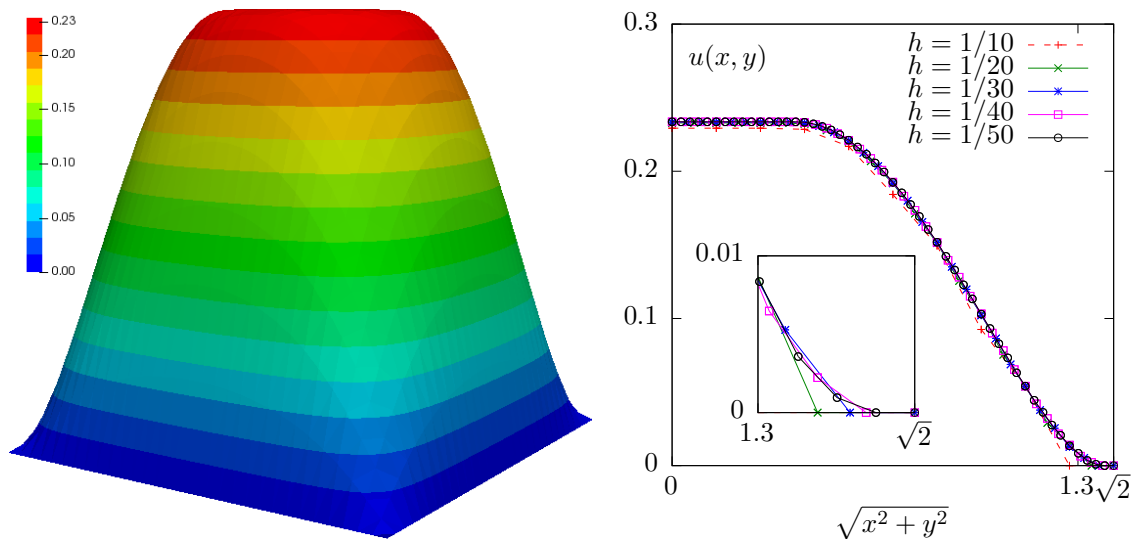


Figure 5.6: The augmented Lagrangian algorithm on the Mosolov problem with  $Bi = 0.4$  and  $n = 1$ : (left) the velocity field in elevation view ( $h = 1/30$ ); (right) velocity cut along the first bisector for various  $h$ .

Observe that the residual terms of problem (5.6a)-(5.6c) are of about  $10^{-10}$ , as required by the stopping criterion. Fig. 5.7 plots the max of the relative error for the  $\sigma_h$  variable: this quantity is used as stopping criterion. Observe that it behaves asymptotically as  $1/k$  for large iteration number  $k$ . Note that these convergence properties could be dramatically improved by using a Newton method, as shown in [95].

Finally, computation can be performed for any  $n > 0$ , any polynomial order  $k \geq 1$  and in a distributed environment for enhancing performances on larger meshes:

```
make mosolov_augmented_lagrangian
mkgeo_grid -a -1 -b 1 -c -1 -d 1 -t 40 > square-40.geo
mpirun -np 8 ./mosolov_augmented_lagrangian square-40.geo P2 0.4 0.5
```

### 5.2.3 Mesh adaptation

An important improvement can be obtained by using mesh adaptation, as shown in [80,82,96]: with a well chosen criterion, rigid regions, where the velocity is constant, can be accurately determined with reasonable mesh sizes. This is especially true for obtaining an accurate determination of the shape of the small regions with zero velocity in the outer corner of the pipe section. In order to reduce the computational time, we can reduce the pipe section flow to only a sector, thanks to symmetries (see Fig. 5.8):

```
mkgeo_sector
geo sector.geo
```

Then, the computation is run by indicating an adaptation loop with ten successive meshes:

```
make mosolov_augmented_lagrangian
mpirun -np 8 ./mosolov_augmented_lagrangian sector P2 0.5 1 10
```

Fig. 5.9 shows the evolution of the mesh size and the minimal edge length during the adaptation loop: observe the convergence of the meshes to an optimal one.

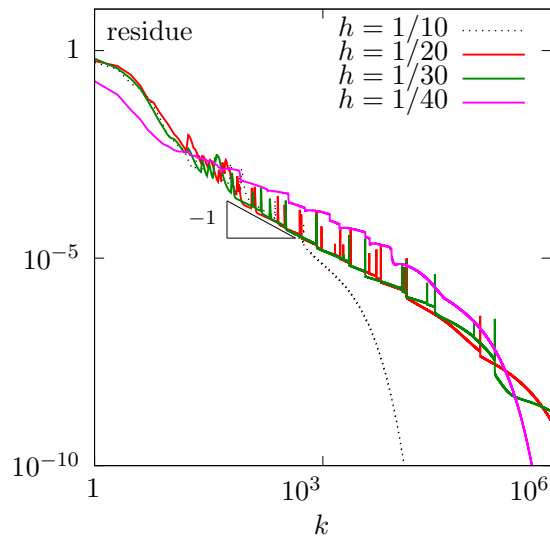


Figure 5.7: The augmented Lagrangian algorithm on the Mosolov problem with  $Bi = 0.4$  and  $n = 1$ : residue versus iteration  $k$  for various  $h$ .

Example file 5.16: mosolov\_yield\_surface.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 int main(int argc, char**argv) {
5     environment rheolef (argc,argv);
6     Float tol = (argc > 1) ? atof(argv[1]) : 1e-15;
7     Float Bi;
8     field sigma_h;
9     din >> catchmark("Bi") >> Bi
10    >> catchmark("sigma") >> sigma_h;
11    space Th = sigma_h.get_space();
12    space Th1 (Th.get_geo(), "P" + itos(4*(Th.degree()+1)) + "d");
13    dout << interpolate (Th1, norm(sigma_h)-Bi);
14 }

```

The yield surface is the zero isosurface of the level set function  $\phi(\mathbf{x} = |\sigma_h(\mathbf{x}) - Bi$ . Its visualization is easy to obtain by the following commands:

```

make mosolov_yield_surface
zcat sector-010.field.gz | ./mosolov_yield_surface | \
    field -proj P1 -n-iso 10 -n-iso-negative 5

```

The unyielded zones, associated to negative values, appear in cold colors. Conversely, the yielded ones are represented by warm colors. A combined representation of the solution can be obtained by the following command:

```
bash mkview_mosolov sector-010.field.gz
```

The shell script `mkview_mosolov` invokes `mosolov_yield_surface` and then directly builds the view shown on Fig. 5.8 in the `paraview` graphic render. Please, click to deselect the `show box` option for completing the view. A cut of the velocity field along the first bisector is obtained by:

```
field -mark u sector-010.field.gz -mark u -domain bisector -elevation -gnuplot
```

Observe on Fig. 5.8.bottom the good capture of the small regions with zero velocity, near the outer corner of the square pipe section. When compared with Fig. 5.6.right, the benefit of mesh adaptation appears clearly.

### 5.2.4 Error analysis

Theoretical error bounds for this problem can be found in [79]. In order to study the error between the numerical solution and the exact solution of the Mosolov problem, let us investigate a case for which the exact solution can be explicitly expressed. We consider the special case of a flow of a viscoplastic fluid in a circular pipe. The pressure is expressed by  $p(z) = -fz$ . The velocity has only one nonzero component along the  $0z$  axis, denoted as  $u(r)$  for simplicity. Conversely, the symmetric tensor  $\sigma$  has only one non-zero  $rz$  component, denoted as  $\sigma(r)$ . Thanks to the expression of the tensor-divergence operator in axisymmetric coordinates [12, p. 588], the problem reduces to :

(P) : find  $\sigma(r)$  and  $u(r)$ , defined in  $] - R, R[$  such that

$$\begin{cases} \sigma(r) = K|u'(r)|^{-1+n} u'(r) + \sigma_0 \frac{u'(r)}{|u'(r)|}, & \text{when } u'(r) \neq 0 \\ |\sigma(r)| \leq \sigma_0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} -\frac{1}{r}(r\sigma)' - f &= 0 & \text{in } ] - R, R[ \\ u(-R) = u(R) &= 0 \end{aligned}$$

Let  $\Sigma = fR/2$  be a representative stress and  $U$  a representative velocity such that  $K(U/R)^n = \Sigma$ . Then, we consider the following change of unknown:

$$r = R\tilde{r}, \quad u = U\tilde{u}, \quad \sigma = \Sigma\tilde{\sigma}$$

The system reduces to a problem with only two parameters  $n$  and the Bingham number  $Bi = 2\sigma_0/(fR)$ , that measures the ratio between the yield stress and the load. Since there is no more ambiguity, we omit the tildes :

(P) : find  $\sigma$  and  $u$ , defined in  $] - 1, 1[$  such that

$$\begin{cases} \sigma(r) = |u'(r)|^{-1+n} u'(r) + Bi \frac{u'(r)}{|u'(r)|}, & \text{when } u'(r) \neq 0 \\ |\sigma(r)| \leq Bi & \text{otherwise} \end{cases}$$

$$\begin{aligned} -\frac{1}{r}(r\sigma)' - 2 &= 0 & \text{in } ] - 1, 1[ \\ u(-1) = u(1) &= 0 \end{aligned}$$

Remark that the solution is even :  $u(-r) = u(r)$ . We get  $\sigma(r) = -r$  and the yield stress criterion leads to  $u(x) = 0$  when  $|r| \leq Bi$ : the load is weaker than the yield stress and the flow is null. When  $Bi > 1$  the solution is  $u = 0$ . Otherwise, when  $|r| > Bi$ , we get  $|u'(r)|^n + Bi = |r|$  and finally, with the boundary conditions and the continuity at  $r = \pm Bi$  :

$$u(r) = \frac{(1 - Bi)^{1+\frac{1}{n}} - \max(0, |r| - Bi)^{1+\frac{1}{n}}}{1 + \frac{1}{n}}$$

When  $n = 1$ , the second derivative of the solution is discontinuous at  $r = Bi$  and its third derivative is not square integrable. For any  $n > 0$ , an inspection of the integrability of the square of the solution derivatives shows that  $u \in H^{1+1/n}([ - 1, 1[, r dr)$  at the best.



Example file 5.17: mosolov\_exact\_circle.icc

```

1 struct u {
2   Float operator() (const point& x) const {
3     return (pow(1-Bi,1+1/n) - pow(max(Float(0),norm(x)-Bi),1+1/n))/(1+1/n);
4   }
5   u (Float Bi1, Float n1) : Bi(Bi1), n(n1) {}
6   protected: Float Bi, n;
7 };
8 struct grad_u {
9   point operator() (const point& x) const {
10    Float r = norm(x);
11    return (r <= Bi) ? point(0,0) : -pow(r-Bi, 1/n)*(x/r);
12  }
13  grad_u (Float Bi1, Float n1) : Bi(Bi1), n(n1) {}
14  protected: Float Bi, n;
15 };
16 struct sigma {
17   point operator() (const point& x) const { return -x; }
18   sigma (Float=0, Float=0) {}
19 };

```

When computing on a circular pipe section, the exact solution is known and it is also possible to compute the error: this is implemented in the file ‘mosolov\_error.cc’. This file it is not listed here but is available in the **Rheolef** example directory. The error analysis is obtained by:

```

make mosolov_error
mkgeo_ball -order 2 -t 10 > circle-P2-10.geo
./mosolov_augmented_lagrangian circle-P2-10.geo P2 0.2 0.5
zcat circle-P2-10.field | ./mosolov_error

```

Note that we use an high order isoparametric approximation of the flow domain for tacking into account the the curved boundaries. Observe on Fig. 5.10 that both the error in  $H^1$  norm for the velocity  $u$  behaves as  $\mathcal{O}(h^s)$  with  $s = \min(k, 2)$ . This is optimal, as, from interpolation theory [15, p. 109], we have:

$$\|u - \pi_h(u)\|_{1,2,\Omega} \leq Ch^s |u|_{s+1,2,\Omega}$$

with  $s = \min(k, 1/n)$  when  $u \in H^{1+1/n}(\Omega)$ . Then, the convergence rate of the error in  $H^1$  norm versus the mesh size is bounded by  $1/n$  for any polynomial order  $k$ . The rate for the error in  $L^\infty$  norm for for the velocity  $u$  is  $\min(k + 1, 1/n)$ , for any  $k \geq 1$  and  $n > 0$ . For the stress  $\sigma$ , the error in  $L^2$  norm behaves as  $\mathcal{O}(h)$  for any polynomial order and, in  $L^\infty$  norm, the rate is weaker, of about  $3/4$ . Finally, for  $n = 1/2$  there is no advantage of using polynomial order more than  $k = 2$  with quasi-uniform meshes. Conversely, for  $n = 1$ , we obtains that there is no advantage of using polynomial order more than  $k = 1$  with quasi-uniform meshes. This limitation can be circumvented by combining mesh adaptation with high order polynomials [79, 96].

### 5.2.5 Error analysis for the yield surface

The limit contour separating the rigid zones are expressed as a level set of the stress norm:

$$\Gamma_h = \{\mathbf{x} \in \Omega ; |\sigma_h(\mathbf{x})| = Bi\}$$

This contour is called the *yield surface* and its exact value is known from the exact solution  $\sigma(\mathbf{x}) = \mathbf{x}$  in the circle:

$$\Gamma = \{\mathbf{x} \in \Omega ; |\mathbf{x}| = Bi\}$$

Recall the stress  $\sigma_h$  converges in  $L^\infty$  norm, thus, there is some hope that its point-wise values converge. As these point-wise values appears in the definition of the yield surface, this suggests

that  $\Gamma_h$  could converge to  $\Gamma$  with mesh refinement: our aim is to check this conjecture. Let us introduce the area between  $\Gamma_h$  and  $\Gamma$  as a  $L^1$  measure of the distance between them:

$$\text{dist}(\Gamma, \Gamma_h) = \int_{\Omega} \delta(|\sigma_h| - Bi, |\sigma| - Bi) \, dx$$

where

$$\delta(\phi, \psi) = \begin{cases} 0 & \text{when } \phi\psi \geq 0 \\ 1 & \text{otherwise} \end{cases}$$

Example file 5.18: mosolov\_error\_yield\_surface.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "mosolov_exact_circle.icc"
5 Float delta (Float f, Float g) { return (f*g >= 0) ? 0 : 1; }
6 int main(int argc, char**argv) {
7     environment rheolef (argc,argv);
8     Float tol = (argc > 1) ? atof(argv[1]) : 1e-15;
9     Float Bi;
10    field sigma_h;
11    din >> catchmark("Bi") >> Bi
12         >> catchmark("sigma") >> sigma_h;
13    space Th = sigma_h.get_space();
14    geo omega = Th.get_geo();
15    quadrature_option qopt;
16    qopt.set_family(quadrature_option::gauss);
17    qopt.set_order(4*(Th.degree()+1));
18    Float err_ys_l1 = integrate (omega,
19        compose(delta, norm(sigma_h)-Bi, norm(sigma()-Bi), qopt);
20    dout << "err_ys_l1 = " << err_ys_l1 << endl;
21    return err_ys_l1 < tol ? 0 : 1;
22 }

```

The computation of the error for the yield surface prediction writes:

```

make mosolov_error_yield_surface
zcat circle-P2-10.field.gz | ./mosolov_error_yield_surface

```

Fig. 5.11.right shows the result: the yield surface error converges as  $\mathcal{O}(h)$  for any  $k \geq 0$ .

Recall that the yield surface is the zero isosurface of the level set function  $\phi(\mathbf{x}) = |\sigma_h(\mathbf{x}) - Bi$ . Its visualization, shown on Fig. 5.11.left, is provided by the following commands:

```

make mosolov_yield_surface
zcat circle-P2-10.field.gz | ./mosolov_yield_surface | \
    field - -proj P1 -n-iso 10 -n-iso-negative 5

```

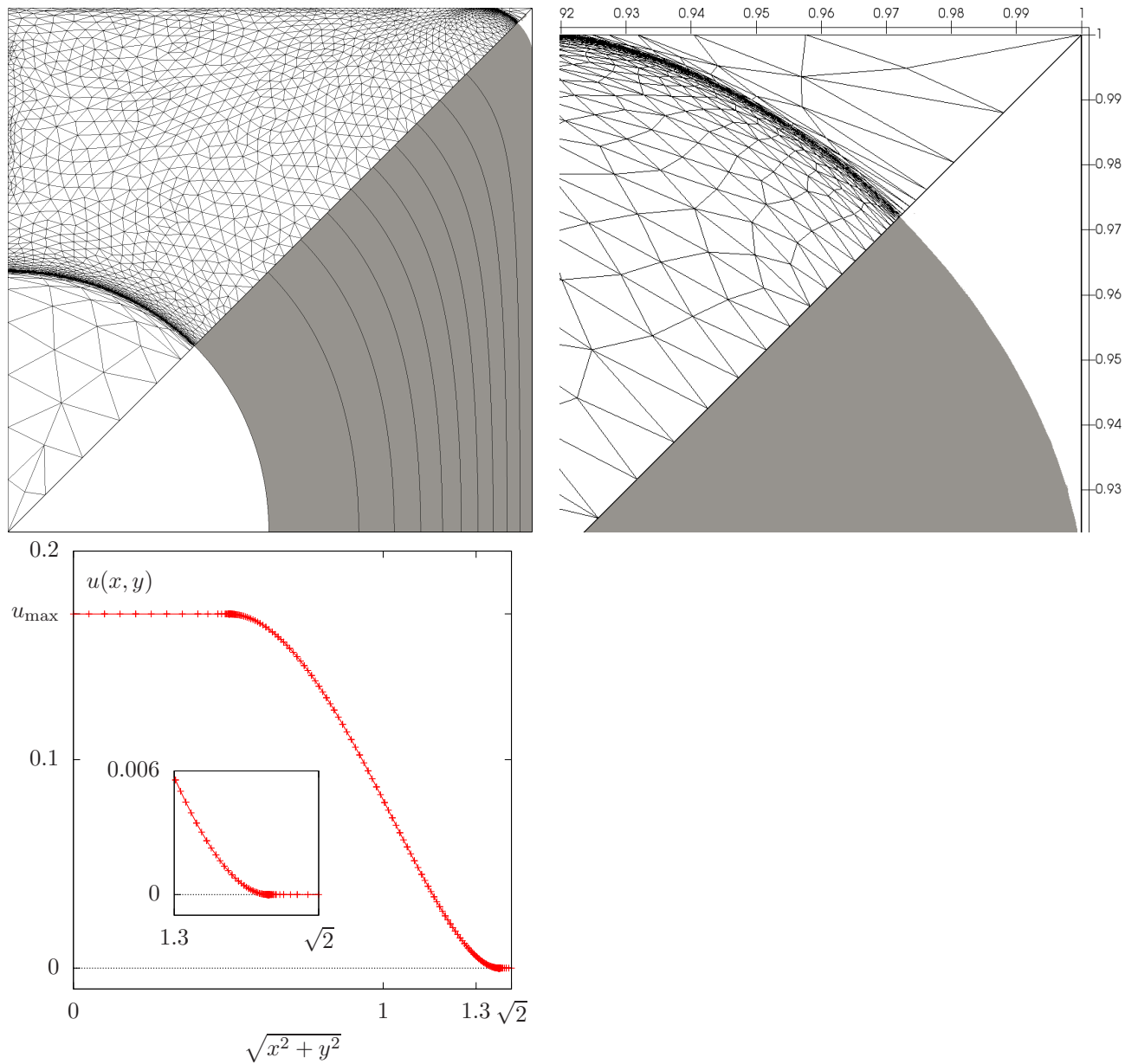


Figure 5.8: Auto-adaptive meshes for the Mosolov problem with  $Bi = 0.5$  and  $n = 1$  and the  $P_2$  element for the velocity: (top) Yielded regions in gray and nine isovalues of the velocity inside  $]0, u_{\max}[$  with  $u_{\max} = 0.169865455242435$ . The auto-adaptive mesh is mirrored. (top-right) Zoom in the outer corner of the square section. (bottom) Velocity cut along the first bisector.

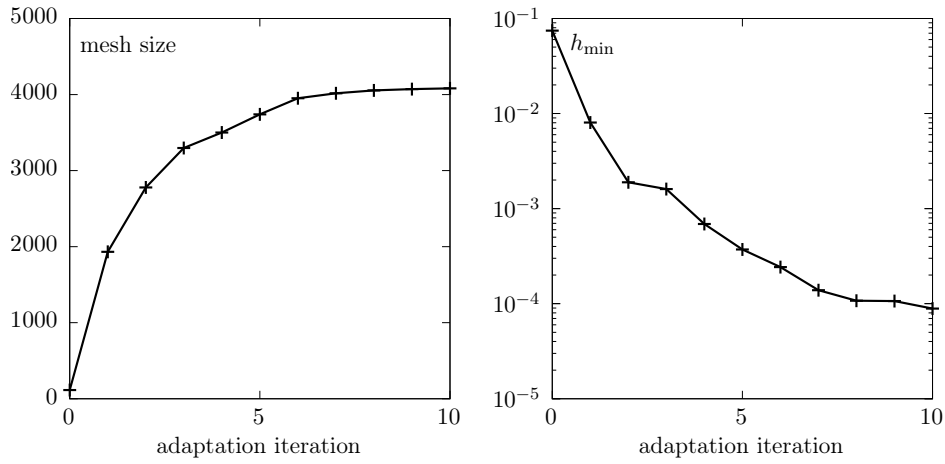


Figure 5.9: Auto-adaptive meshes for the Mosolov problem: evolution of the mesh size (left) and the minimal edge length during the adaptation loop.

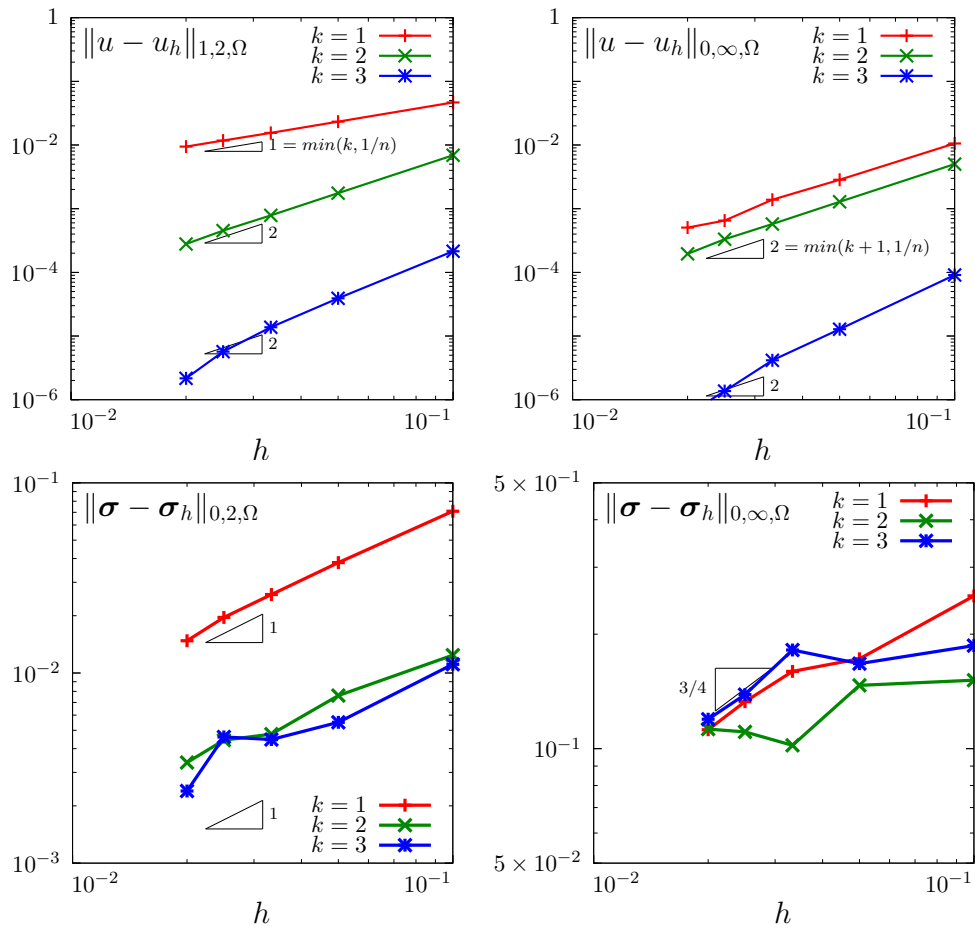


Figure 5.10: The augmented Lagrangian algorithm on the Mosolov problem with  $Bi = 0.2$  and  $n = 1/2$ : error versus mesh parameter  $h$  for various polynomial order  $k$ .

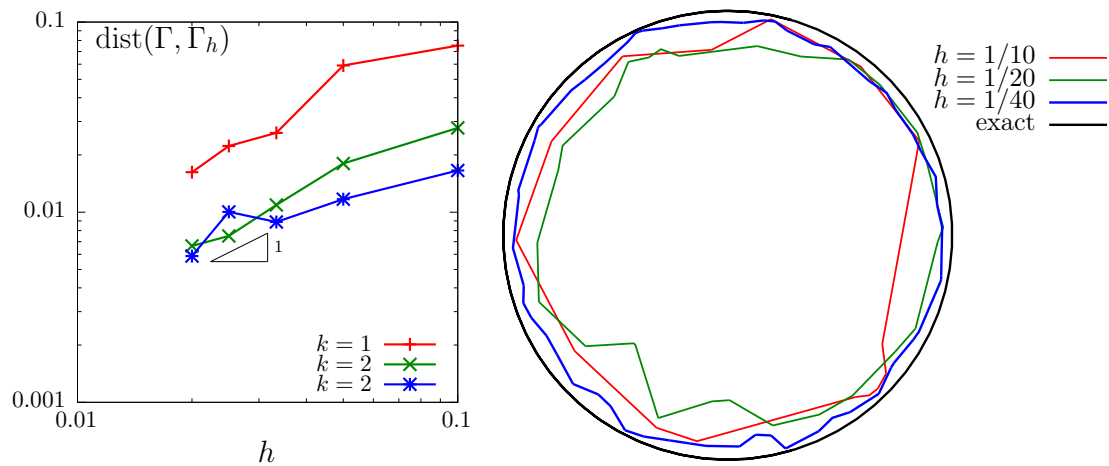


Figure 5.11: The augmented Lagrangian algorithm on the Mosolov problem with  $Bi = 0.2$  and  $n = 1/2$ : (left) convergence of the yield surface versus mesh parameter  $h$  for various polynomial order  $k$  ; (right) visualization of the yield surface ( $P_1$  approximation).

## 5.3 [New] Viscoelastic fluids

### 5.3.1 A tensor transport equation

The aim of this chapter is to introduce to the numerical resolution of equations involving tensor derivatives by using discontinuous approximations. See [90, chap. 4] for an introduction to tensor derivatives and [93] for discontinuous Galerkin methods.

The tensor derivative of a symmetric tensor  $\boldsymbol{\sigma}$  is defined by:

$$\frac{\mathcal{D}_a \boldsymbol{\sigma}}{\mathcal{D}t} = \frac{\partial \boldsymbol{\sigma}}{\partial t} + (\mathbf{u} \cdot \nabla) \boldsymbol{\sigma} + \boldsymbol{\sigma} \mathbf{g}_a(\mathbf{u}) + \mathbf{g}_a^T(\mathbf{u}) \boldsymbol{\sigma}, \quad (5.10)$$

where  $\mathbf{u}$  is a given velocity field,

$$\mathbf{g}_a(\mathbf{u}) = ((1-a) \nabla \mathbf{u} - (1+a) \nabla \mathbf{u}^T) / 2 \quad (5.11)$$

is a generalized velocity gradient and  $a \in [-1, 1]$  is the parameter of the tensor derivative. Problems involving tensor derivatives appear in viscoelasticity (polymer solution and polymer melt, see e.g. [94]), in fluid-particle suspension modelling (see e.g. [68]), in turbulence modelling ( $R_{ij} - \epsilon$  models) or in liquid crystals modelling. Let  $\Omega \subset \mathbb{R}^d$  be a bounded open domain.

The time-dependent tensor transport problem writes:

(P): find  $\boldsymbol{\sigma}$ , defined in  $]0, T[ \times \Omega$ , such that

$$\begin{aligned} \frac{\mathcal{D}_a \boldsymbol{\sigma}}{\mathcal{D}t} + \nu \boldsymbol{\sigma} &= \chi \quad \text{in } ]0, T[ \times \Omega \\ \boldsymbol{\sigma} &= \boldsymbol{\sigma}_\Gamma \quad \text{on } ]0, T[ \times \partial\Omega_- \\ \boldsymbol{\sigma}(0) &= \boldsymbol{\sigma}_0 \quad \text{in } \Omega \end{aligned}$$

where  $\boldsymbol{\sigma}$  is the tensor valued unknown and  $\nu > 0$  is a constant that represents the inverse of the Weissenberg number. Also  $T > 0$  is a given final time, the data  $\chi$ ,  $\boldsymbol{\sigma}_\Gamma$  and  $\boldsymbol{\sigma}_0$  are known and  $\partial\Omega_-$  denotes the upstream boundary (see **Rheolef** documentation [93], section 4.1.1, page 141).

The steady version of the tensor transport problem writes:

(S): find  $\boldsymbol{\sigma}$ , defined in  $\Omega$ , such that

$$\begin{aligned} (\mathbf{u} \cdot \nabla) \boldsymbol{\sigma} + \boldsymbol{\sigma} \mathbf{g}_a(\mathbf{u} + \mathbf{g}_a^T(\mathbf{u})) + \nu \boldsymbol{\sigma} &= f \quad \text{in } \Omega \\ \boldsymbol{\sigma} &= \boldsymbol{\sigma}_\Gamma \quad \text{on } \partial\Omega_- \end{aligned}$$

A sufficient condition this problem to be well posed is [85, 90]:

$$\mathbf{u} \in W^{1,\infty}(\Omega)^d \quad \text{and} \quad 2\nu - \|\operatorname{div} \mathbf{u}\|_{0,\infty,\Omega} - 2a \|D(\mathbf{u})\|_{0,\infty,\Omega} > 0$$

Note that this condition is always satisfied when  $\operatorname{div} \mathbf{u} = 0$  and  $a = 0$ . We introduce the space:

$$X = \{\boldsymbol{\tau} \in L^2(\Omega)_s^{d \times d}; (\mathbf{u} \cdot \nabla) \boldsymbol{\tau} \in L^2(\Omega)_s^{d \times d}\}$$

and, for all  $\boldsymbol{\sigma}, \boldsymbol{\tau} \in X$

$$\begin{aligned} a(\boldsymbol{\sigma}, \boldsymbol{\tau}) &= \int_{\Omega} ((\mathbf{u} \cdot \nabla) \boldsymbol{\sigma} + \boldsymbol{\sigma} \mathbf{g}_a(\mathbf{u}) + \mathbf{g}_a^T(\mathbf{u}) \boldsymbol{\sigma} + \nu \boldsymbol{\sigma}) : \boldsymbol{\tau} \, dx + \int_{\partial\Omega} \max(0, -\mathbf{u} \cdot \mathbf{n}) \boldsymbol{\sigma} : \boldsymbol{\tau} \, ds \\ l(\boldsymbol{\tau}) &= \int_{\Omega} \chi : \boldsymbol{\tau} \, dx + \int_{\partial\Omega} \max(0, -\mathbf{u} \cdot \mathbf{n}) \boldsymbol{\sigma}_\Gamma : \boldsymbol{\tau} \, ds \end{aligned}$$

Then, the variational formulation of the steady problem writes:

(FV): find  $\boldsymbol{\sigma} \in X$  such that

$$a(\boldsymbol{\sigma}, \boldsymbol{\tau}) = l(\boldsymbol{\tau}), \quad \forall \boldsymbol{\tau} \in X$$

Notice that the term  $\max(0, -\mathbf{u}\cdot\mathbf{n}) = (|\mathbf{u}\cdot\mathbf{n}| - \mathbf{u}\cdot\mathbf{n})/2$  is positive and vanishes everywhere except on  $\partial\Omega_-$ . Thus, the boundary condition  $\phi = \phi_\Gamma$  is weakly imposed on  $\partial\Omega_-$  via the integrals on the boundary. We aim at adapting the discontinuous Galerkin method to this problem. The *discontinuous* finite element space is defined by:

$$X_h = \{\boldsymbol{\tau}_h \in L^2(\Omega)_s^{d \times d}; \boldsymbol{\tau}_h|_K \in P_k, \forall K \in \mathcal{T}_h\}$$

where  $k \geq 0$  is the polynomial degree. Notice that  $X_h \not\subset X$  and that the  $\nabla \boldsymbol{\tau}_h$  term has no more sense for discontinuous functions  $\boldsymbol{\tau}_h \in X_h$ . We introduce the *broken gradient*  $\nabla_h$  as a convenient notation (see also rheolef documentation, volume 2).

$$(\nabla_h \boldsymbol{\tau}_h)|_K = \nabla(\boldsymbol{\tau}_h|_K), \forall K \in \mathcal{T}_h$$

Thus

$$\int_{\Omega} ((\mathbf{u}\cdot\nabla_h)\boldsymbol{\sigma}_h) : \boldsymbol{\tau}_h \, dx = \sum_{K \in \mathcal{T}_h} \int_K ((\mathbf{u}\cdot\nabla)\boldsymbol{\sigma}_h) : \boldsymbol{\tau}_h \, dx, \forall \boldsymbol{\sigma}_h, \boldsymbol{\tau}_h \in X_h$$

This leads to a discrete version  $a_h$  of the bilinear form  $a$ , defined for all  $\boldsymbol{\sigma}_h, \boldsymbol{\tau}_h \in X_h$  by:

$$\begin{aligned} a_h(\boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) &= t_h(\mathbf{u}; \boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) + \nu \int_{\Omega} \boldsymbol{\sigma}_h : \boldsymbol{\tau}_h \, dx \\ t_h(\mathbf{u}; \boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) &= \int_{\Omega} ((\mathbf{u}\cdot\nabla_h)\boldsymbol{\sigma}_h + \boldsymbol{\sigma} g_a(\mathbf{u}) + g_a^T(\mathbf{u})\boldsymbol{\sigma}) : \boldsymbol{\tau}_h \, dx + \int_{\partial\Omega} \max(0, -\mathbf{u}\cdot\mathbf{n}) \boldsymbol{\sigma}_h : \boldsymbol{\tau}_h \, ds \\ &\quad + \sum_{S \in \mathcal{S}_h^{(i)}} \int_S \left( -\mathbf{u}\cdot\mathbf{n} [[\boldsymbol{\sigma}_h]] : \{\{\boldsymbol{\tau}_h\}\} + \frac{\alpha}{2} |\mathbf{u}\cdot\mathbf{n}| [[\boldsymbol{\sigma}_h]] : [[\boldsymbol{\tau}_h]] \right) ds \end{aligned} \quad (5.12)$$

(FV)<sub>h</sub>: find  $\boldsymbol{\sigma}_h \in X_h$  such that

$$a_h(\boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) = l(\boldsymbol{\tau}_h), \forall \boldsymbol{\tau}_h \in X_h$$

The following code implement this problem in the **Rheolef** environment.

Example file 5.19: transport\_tensor\_dg.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "transport_tensor_exact.icc"
5 int main(int argc, char**argv) {
6     environment rheolef (argc, argv);
7     geo omega (argv[1]);
8     space Xh (omega, argv[2], "tensor");
9     Float alpha = (argc > 3) ? atof(argv[3]) : 1;
10    Float nu = (argc > 4) ? atof(argv[4]) : 3;
11    Float t0 = (argc > 5) ? atof(argv[5]) : acos(-1.)/8;
12    Float a = 0;
13    trial sigma (Xh); test tau (Xh);
14    tensor ma = 0.5*((1-a)*grad_u - (1+a)*trans(grad_u));
15    auto beta_a = sigma*ma + trans(ma)*sigma;
16    form ah = integrate (ddot(grad_h(sigma)*u + beta_a + nu*sigma, tau))
17              + integrate ("boundary",
18                          max(0, -dot(u, normal()))*ddot(sigma, tau))
19              + integrate ("internal_sides",
20                          - dot(u, normal())*ddot(jump(sigma), average(tau))
21                          + 0.5*alpha*abs(dot(u, normal()))
22                          *ddot(jump(sigma), jump(tau)));
23    field lh = integrate (ddot(chi(nu, t0), tau))
24                  + integrate ("boundary",
25                              max(0, -dot(u, normal()))*ddot(sigma_g(nu, t0), tau));
26    solver sah (ah.uu());
27    field sigma_h(Xh);
28    sigma_h.set_u() = sah.solve(lh.u());
29    dout << catchmark("nu") << nu << endl
30          << catchmark("t0") << t0 << endl
31          << catchmark("sigma") << sigma_h;
32 }

```

### Running the program

Let  $d = 2$  and  $\Omega = ]-1/2, 1/2[^2$ . We consider the rotating field  $\mathbf{u} = (-x_2, x_1)$ . A particular solution of the time-dependent problem with zero right-hand side is given by:

$$\sigma(x, t) = \frac{1}{2} \exp \left\{ -\frac{t}{\lambda} - \frac{(x_1 - x_{1,c}(t))^2 + (x_2 - x_{2,c}(t))^2}{r_0^2} \right\} \times \begin{pmatrix} 1 + \cos(2t) & \sin(2t) \\ \sin(2t) & 1 - \cos(2t) \end{pmatrix}$$

where  $x_{1,c}(t) = \bar{x}_{1,c} \cos(t) - \bar{x}_{2,c} \sin(t)$  and  $x_{2,c}(t) = \bar{x}_{1,c} \sin(t) + \bar{x}_{2,c} \cos(t)$  with  $r_0 > 0$  and  $(\bar{x}_{1,c}, \bar{x}_{2,c}) \in \mathbb{R}^2$ . The initial condition is chosen as  $\sigma_0(x) = \sigma(0, x)$ . This exact solution is implemented in the file ‘transport\_tensor\_exact.icc’. This file it is not listed here but is available in the **Rheolef** example directory. For the steady problem, the right-hand side could be chosen as  $\chi = -\frac{\partial \sigma}{\partial t}$  and then  $t = t_0$  is fixed. The numerical tests correspond to  $\nu = 3$ ,  $r_0 = 1/10$ ,  $(\bar{x}_{1,c}, \bar{x}_{2,c}) = (1/4, 0)$  and a fixed time  $t_0 = \pi/8$ .

```
make transport_tensor_dg
mkgeo_grid -t 80 -a -0.5 -b 0.5 -c -0.5 -d 0.5 > square2.geo
./transport_tensor_dg square2 P1d > square2.field
field square2.field -comp 00 -elevation
field square2.field -comp 01 -elevation
```

The computation could also be performed with any Pkd, with  $k \geq 0$ .

### Error analysis

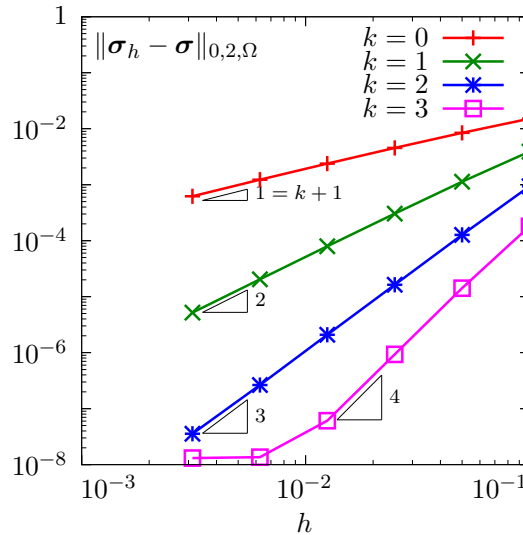


Figure 5.12: Transport tensor problem: convergence versus mesh size.

The file ‘transport\_tensor\_error\_dg.cc’ implement the computation of the error between the approximate solution  $\sigma_h$  and the exact one  $\sigma$ . This file it is not listed here but is available in the **Rheolef** example directory. The computation of the error is obtained by:

```
make transport_tensor_error_dg
./transport_tensor_error_dg < square2.field
```

The error is plotted on Fig. 5.12 for various mesh size  $h$  and polynomial order  $k$ : observe the optimality of the convergence properties. For  $k = 4$  and on the finest mesh, the error saturates



at about  $10^{-8}$ , due to finite machine precision effects. In the next chapter, transport tensor approximation are applied to viscoelastic fluid flow computations.

### 5.3.2 The Oldroyd model

We consider the following viscoelastic fluid flow problem (see e.g. [94, chap. 4]):

(P): find  $\tau$ ,  $\mathbf{u}$  and  $p$  defined in  $]0, T[ \times \Omega$  such that

$$We \frac{\mathcal{D}_a \tau}{\mathcal{D}t} + \tau - 2\alpha D(\mathbf{u}) = 0 \text{ in } ]0, T[ \times \Omega \quad (5.13a)$$

$$Re \left( \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) - \operatorname{div} (\tau + 2(1 - \alpha)D(\mathbf{u}) - p.I) = 0 \text{ in } ]0, T[ \times \Omega \quad (5.13b)$$

$$-\operatorname{div} \mathbf{u} = 0 \text{ in } ]0, T[ \times \Omega \quad (5.13c)$$

$$\tau = \tau_\Gamma \text{ on } ]0, T[ \times \partial\Omega_- \quad (5.13d)$$

$$\mathbf{u} = \mathbf{u}_\Gamma \text{ on } ]0, T[ \times \partial\Omega \quad (5.13e)$$

$$\tau(0) = \tau_0 \text{ and } \mathbf{u}(0) = \mathbf{u}_0 \text{ in } \Omega \quad (5.13f)$$

where  $\tau_0$ ,  $\mathbf{u}_0$ ,  $\tau_\Gamma$  and  $\mathbf{u}_\Gamma$  are given. The first equation corresponds to a generalized Oldroyd model [63]: when  $a = -1$  we obtain the Oldroyd-A model, when  $a = 1$ , the Oldroyd-B model, and when  $a \in ]-1, 1[$  a generalization of these two models. The dimensionless number  $We \geq 0$  is the Weissenberg number: this is the main parameter for this problem. The dimensionless Reynolds number  $Re \geq 0$  is often chosen small: as such fluids are usually slow, the  $\mathbf{u} \cdot \nabla \mathbf{u}$  inertia term is also neglected here for simplicity. The parameter  $\alpha \in [0, 1]$  represent a retardation. When  $\alpha = 1$  we obtain the Maxwell model, that is a reduced version of the Oldroyd one. The total Cauchy stress tensor is expressed by:

$$\boldsymbol{\sigma}_{\text{tot}} = -p.I + 2(1 - \alpha)D(\mathbf{u}) + \tau \quad (5.14)$$

### 5.3.3 The $\theta$ -scheme algorithm

The  $\theta$ -scheme is considered for the time discretization [85] (see also [94, chap. 4]): this leads to a semi-implicit splitting algorithm that defines a sequence  $(\tau^{(n)}, \mathbf{u}^{(n)}, p^{(n)})_{n \geq 0}$  as

- $n = 0$ : set  $(\tau^{(0)}, \mathbf{u}^{(0)}) = (\tau_0, \mathbf{u}_0)$  and  $p^0$  arbitrarily chosen.
- $n \geq 0$ : let  $(\tau^{(n)}, \mathbf{u}^{(n)})$  being known, then  $(\tau^{(n+1)}, \mathbf{u}^{(n+1)}, p^{(n+1)})$  is defined in three sub-steps.

\* sub-step 1: compute explicitly:

$$\begin{aligned} \gamma &= \mathbf{u}^{(n)} \cdot \nabla \tau^{(n)} + \tau^{(n)} M_a(\mathbf{u}^{(n)}) + M_a^T(\mathbf{u}^{(n)}) \tau^{(n)} \\ \tilde{\mathbf{f}} &= \lambda \mathbf{u}^{(n)} + \operatorname{div} (c_1 \tau^{(n)} + c_2 \gamma) \end{aligned}$$

where

$$c_1 = \frac{We}{We + \theta \Delta t} \quad \text{and} \quad c_2 = -\frac{We \theta \Delta t}{We + \theta \Delta t}$$

Then determine  $(\mathbf{u}^{(n+\theta)}, p^{(n+\theta)})$  such that

$$\lambda \mathbf{u}^{(n+\theta)} - \operatorname{div} (2\eta D(\mathbf{u}^{(n+\theta)})) + \nabla p^{(n+\theta)} = \tilde{\mathbf{f}} \text{ in } \Omega \quad (5.15a)$$

$$-\operatorname{div} \mathbf{u}^{(n+\theta)} = 0 \text{ in } \Omega \quad (5.15b)$$

$$\mathbf{u} = \mathbf{u}_\Gamma((n + \theta)\Delta t) \text{ on } \partial\Omega \quad (5.15c)$$

and finally, compute explicitly

$$\tau^{(n+\theta)} = c_1 \tau^{(n)} + c_2 \gamma + 2c_3 D(\mathbf{u}^{(n+\theta)}) \quad (5.15d)$$

where

$$\lambda = \frac{Re}{\theta \Delta t}, \quad \eta = \frac{(1-\alpha)We + \theta \Delta t}{We + \theta \Delta t} \quad \text{and} \quad c_3 = \frac{\alpha \theta \Delta t}{We + \theta \Delta t}$$

\* sub-step 2:  $(\tau^{(n+\theta)}, \mathbf{u}^{(n+\theta)})$  being known, compute explicitly

$$\begin{aligned} \mathbf{u}^{(n+1-\theta)} &= \frac{1-\theta}{\theta} \mathbf{u}^{(n+\theta)} - \frac{1-2\theta}{\theta} \mathbf{u}^{(n)} \\ \xi &= c_4 \tau_{n+\theta} + 2c_5 D(\mathbf{u}^{(n+\theta)}) \end{aligned}$$

and then find  $\tau^{(n+1-\theta)}$  such that

$$\begin{aligned} \mathbf{u}^{(n+1-\theta)} \cdot \nabla \tau^{(n+1-\theta)} + \tau^{(n+1-\theta)} M_a(\mathbf{u}^{(n+1-\theta)}) + M_a^T(\mathbf{u}^{(n+1-\theta)}) \tau^{(n+1-\theta)} \\ + \nu \tau^{(n+1-\theta)} = \xi \quad \text{in } \Omega \end{aligned} \quad (5.16a)$$

$$\tau^{(n+1-\theta)} = \tau_\Gamma((n+1-\theta)\Delta t) \quad \text{on } \partial\Omega_- \quad (5.16b)$$

where

$$\nu = \frac{1}{(1-2\theta)\Delta t}, \quad c_4 = \frac{1}{(1-2\theta)\Delta t} - \frac{1}{We} \quad \text{and} \quad c_5 = \frac{\alpha}{We}$$

\* sub-step 3 is obtained by replacing  $n$  and  $n + \theta$  by  $n + 1 - \theta$  and  $n + 1$ , respectively.

Thus, sub-step 1 and 2 reduces to two similar generalized Stokes problems while sub-step 3 involves a stress transport problem. Here  $\Delta t > 0$  and  $\theta \in ]0, 1/2[$  are numerical parameters. A good choice is  $\theta = 1 - 1/\sqrt{2}$  [87]. This algorithm was first proposed in [84] and extended in [86] to Phan-Thien and Tanner viscoelastic models. See also [105] for another similar approach in the context of FENE viscoelastic models. In [100] are presented some benchmarks of this algorithm while [22] presents a numerical analysis of its convergence properties. The main advantage of this time-depend algorithm is its flexibility: while most time-dependent splitting algorithms for viscoelastic are limited to  $\alpha \leq 1/2$  (see e.g. [70]), here the full range  $\alpha \in ]0, 1]$  is available.

Let us introduce the finite element spaces:

$$\begin{aligned} T_h &= \{\tau_h \in (L^2(\Omega))_s^{d \times d}; \tau_{h|K} \in (P_1)_s^{d \times d}, \forall K \in \mathcal{T}_h\} \\ X_h &= \{\mathbf{v}_h \in (H^1(\Omega))^d; \tau_{h|K} \in (P_2)^d, \forall K \in \mathcal{T}_h\} \\ Q_h &= \{q_h \in L^2(\Omega); q_{h|K} \in P_1, \forall K \in \mathcal{T}_h\} \end{aligned}$$

Note the discontinuous approximation of pressure: it presents a major advantage, as  $\text{div}(X_h) \subset Q_h$ , it leads to an exact *divergence-free* approximation of the velocity: for any field  $\mathbf{v}_h \in X_h$  satisfying  $\int_\Omega q_h \text{div}(\mathbf{v}_h) dx = 0$  for all  $q_h \in Q_h$ , we have  $\text{div} \mathbf{v}_h = 0$  point-wise, everywhere in  $\Omega$ . The pair  $(X_h, Q_h)$  is known as the Scott-Vogelius lowest-order finite element approximation [99]. This is a major advantage when dealing with a transport equation. The only drawback is that the pair  $(X_h, Q_h)$  does not satisfy the inf-sup condition for an arbitrary mesh. There exists a solution to this however: Arnold and Qin [7] proposed a macro element technique applied to the mesh [7] that allows satisfying the inf-sup condition: for any triangular finite element mesh, it is sufficient to split each triangle in three elements from its barycenter (see also [91]). Notice that the macro element technique extends to quadrilateral meshes [7] and to the three-dimensional case [113]. By this way, the approximate velocity field satisfies exactly the incompressibility constraint: this is

an essential property for the operator splitting algorithm to behave correctly, combining stress transport equation with a divergence-free velocity approximation.

The tensor transport term is discretized as in the previous chapter, by using the  $t_h$  trilinear form introduced in (5.12) page 214. The bilinear forms  $b$ ,  $c$ ,  $d$  are defined by:

$$\begin{aligned} b(\boldsymbol{\sigma}_h, \mathbf{v}_h) &= \int_{\Omega} \boldsymbol{\sigma}_h : D(\mathbf{v}_h) \, dx \\ c(\mathbf{u}_h, \mathbf{v}_h) &= \int_{\Omega} D(\mathbf{u}_h) : D(\mathbf{v}_h) \, dx \\ d(\mathbf{u}_h, \mathbf{q}_h) &= \int_{\Omega} \operatorname{div}(\mathbf{u}_h) q_h \, dx \end{aligned}$$

Let  $T$ ,  $B$ ,  $C$ ,  $D$  and  $M$  be the discrete operators (i.e. the matrix) associated to the forms  $t_h$ ,  $b$ ,  $c$ ,  $d$  and the  $L^2$  scalar product in  $T_h$ . Assume that a stationary state is reached for the discretized algorithm. Then, (5.15a)-(5.15d) writes

$$\begin{aligned} WeT\tau_h + M\tau_h - 2\alpha B^T u_h &= 0 \\ B(c_1\tau_h + M^{-1}T\tau_h) + 2\eta C u_h + D^T p_h &= 0 \\ Du_h &= 0 \end{aligned}$$

Notice that (5.16a)-(5.16b) reduces also to the first equation of the previous system. Expanding the coefficients, combining the two previous equations, and using  $C = BM^{-1}B^T$ , we obtain the system characterizing the stationary solution of the discrete version of the algorithm:

$$\begin{aligned} WeT\tau_h + M\tau_h - 2\alpha B^T u_h &= 0 \\ B\tau_h + 2(1 - \alpha)C u_h + D^T p_h &= 0 \\ Du_h &= 0 \end{aligned}$$

Notice that this is exactly the discretized version of the stationary equation<sup>1</sup>. In order to check that the solution reaches a stationary state, the residual terms of this stationary equations are computed at each iteration, together with the relative error between two iterates.

---

<sup>1</sup>The original  $\theta$ -scheme presented in [85] has an additional relaxation parameter  $\omega$ . The present version correspond to  $\omega = 1$ . When  $\omega \neq 1$ , the stationary solution still depends slightly upon  $\Delta t$ , as the stationary system do not simplifies completely.

Example file 5.20: oldroyd\_theta\_scheme.h

```

1  template<class Problem>
2  struct oldroyd_theta_scheme {
3      oldroyd_theta_scheme();
4      void initial (const geo& omega, field& tau_h, field& uh, field& ph,
5                  string restart);
6      bool solve (field& tau_h, field& uh, field& ph);
7  protected:
8      void step      (const field& tau_h0, const field& uh0, const field& ph0,
9                    field& tau_h, field& uh, field& ph) const;
10     void sub_step1 (const field& tau_h0, const field& uh0, const field& ph0,
11                   field& tau_h, field& uh, field& ph) const;
12     void sub_step2 (const field& uh0, const field& tau_h1, const field& uh1,
13                   field& tau_h, field& uh) const;
14     Float residue (field& tau_h, field& uh, field& ph) const;
15     void reset (const geo& omega);
16     void update_transport_stress (const field& uh) const;
17 public:
18     Float We, alpha, a, Re, delta_t, tol;
19     size_t max_iter;
20 protected:
21     space Th, Xh, Qh;
22     form b, c, d, mt, inv_mt, mu, mp;
23     mutable form th;
24     mutable field thb;
25     Float theta, lambda, eta, nu, c1, c2, c3, c4, c5;
26     solver_abtb stokes;
27 };
28 #include "oldroyd_theta_scheme1.icc"
29 #include "oldroyd_theta_scheme2.icc"
30 #include "oldroyd_theta_scheme3.icc"

```

Example file 5.21: oldroyd\_theta\_scheme1.icc

```

1  template<class P>
2  oldroyd_theta_scheme<P>::oldroyd_theta_scheme()
3  : We(0), alpha(8./9), a(1), Re(1), delta_t(0.025), tol(1e-6), max_iter(500),
4    Th(), Xh(), Qh(), b(), c(), d(), mt(), inv_mt(), mu(), mp(), th() {}
5  template<class P>
6  void oldroyd_theta_scheme<P>::reset(const geo& omega) {
7      Th = space (omega, "P1d", "tensor");
8      Xh = P::velocity_space (omega, "P2");
9      Qh = space (omega, "P1d");
10     theta = 1-1/sqrt(2.);
11     lambda = Re/(theta*delta_t);
12     eta = ((1 - alpha)*We + theta*delta_t)/(We + theta*delta_t);
13     nu = 1/((1-2*theta)*delta_t);
14     c1 = We/(We + theta*delta_t);
15     c2 = - We*theta*delta_t/(We + theta*delta_t);
16     c3 = alpha*theta*delta_t/(We + theta*delta_t);
17     c4 = 1/((1-2*theta)*delta_t) - 1/We;
18     c5 = alpha/We;
19     trial u (Xh), tau(Th), p (Qh);
20     test v (Xh), xi (Th), q (Qh);
21     mt = integrate (ddot(tau,xi));
22     mu = integrate (dot(u,v));
23     mp = integrate (p*q);
24     integrate_option fopt;
25     fopt.invert = true;
26     inv_mt = integrate (ddot(tau,xi), fopt);
27     b = integrate (-ddot(tau,D(v)));
28     c = integrate (lambda*dot(u,v) + 2*eta*ddot(D(u),D(v)));
29     d = integrate (-div(u)*q);
30     stokes = solver_abtb (c.uu(), d.uu(), mp.uu());
31 }

```

Example file 5.22: oldroyd\_theta\_scheme2.icc

```

1  template <class P>
2  bool oldroyd_theta_scheme<P>::solve(field& tau_h, field& uh, field& ph) {
3      reset (uh.get_geo());
4      field tau_h0 = tau_h, uh0 = uh, ph0 = ph;
5      derr << "# n t rel_err residue lambda_min" << endl;
6      Float r = residue (tau_h, uh, ph);
7      Float rel_err = 0;
8      derr << "0 0 0 " << r << endl;
9      for (size_t n = 1; n <= max_iter; ++n) {
10         step (tau_h0, uh0, ph0, tau_h, uh, ph);
11         Float rel_err_prec = rel_err, r_prec = r;
12         r = residue (tau_h, uh, ph);
13         rel_err = field(tau_h-tau_h0).max_abs() + field(uh-uh0).max_abs();
14         derr << n << " " << n*delta_t << " " << rel_err << " " << r << endl;
15         if (rel_err < tol) return true;
16         if (rel_err_prec != 0 && ((rel_err > 10*rel_err_prec && r > 10*r_prec) |
17             (rel_err > 1e5 && r > 1e5))) return false;
18         tau_h0 = tau_h; uh0 = uh; ph0 = ph;
19     }
20     return (rel_err < sqrt(tol));
21 }
22 template <class P>
23 void oldroyd_theta_scheme<P>::initial (
24     const geo& omega, field& tau_h, field& uh, field& ph, string restart) {
25     reset (omega);
26     ph = field(Qh,0);
27     if (restart == "") {
28         uh = P::velocity_field (Xh);
29         trial u (Xh); test v (Xh), xi(Th);
30         form c0 = integrate (2*ddot(D(u),D(v)));
31         solver_abtb s0 (c0.uu(), d.uu(), mp.uu());
32         s0.solve (-(c0.ub()*uh.b()), -(d.ub()*uh.b()),
33             uh.set_u(), ph.set_u());
34         field Duh = inv_mt*integrate(ddot(D(uh),xi));
35         tau_h = 2*alpha*Duh;
36     } else {
37         tau_h = field(Th);
38         uh = field(Xh);
39         idiststream in (restart, "field");
40         in >> catchmark("tau") >> tau_h
41             >> catchmark("u") >> uh
42             >> catchmark("p") >> ph;
43     }
44 }
45 template <class P>
46 void oldroyd_theta_scheme<P>::step (
47     const field& tau_h0, const field& uh0, const field& ph0,
48     field& tau_h, field& uh, field& ph) const {
49     field tau_h1 = tau_h0, uh1 = uh0, ph1 = ph0;
50     sub_step1 (tau_h0, uh0, ph0, tau_h1, uh1, ph1);
51     field tau_h2 = tau_h1, uh2 = uh1;
52     sub_step2 (uh0, tau_h1, uh1, tau_h2, uh2);
53     sub_step1 (tau_h2, uh2, ph1, tau_h, uh, ph);
54 }
55 template <class P>
56 Float
57 oldroyd_theta_scheme<P>::residue(field& tau_h, field& uh, field& ph) const{
58     update_transport_stress (uh);
59     test xi (Th);
60     field Duh = inv_mt*integrate(ddot(D(uh),xi));
61     field gh = 2*Duh;
62     field rt = We*(th*tau_h-thb) + integrate (ddot(tau_h - alpha*gh, xi));
63     field ru = b*(tau_h + (1-alpha)*gh) - d.trans_mult(ph);
64     ru.set_b() = 0;
65     field rp = d*uh;
66     return rt.u().max_abs() + ru.u().max_abs() + rp.u().max_abs();
67 }

```

Example file 5.23: oldroyd\_theta\_scheme3.icc

```

1  template <class P>
2  void oldroyd_theta_scheme<P>::sub_step1 (
3      const field& tau_h0, const field& uh0, const field& ph0,
4      field& tau_h, field& uh, field& ph) const
5  {
6      update_transport_stress (uh0);
7      field gamma_h = inv_mt*(th*tau_h0 - thb);
8      test v (Xh), xi (Th);
9      field lh = lambda*integrate (dot(uh0,v))
10         + b*(c1*tau_h0 + c2*gamma_h);
11     ph = ph0;
12     uh.set_u() = uh0.u();
13     stokes.solve (lh.u() - c.ub()*uh.b(), -d.ub()*uh.b(),
14                 uh.set_u(), ph.set_u());
15     field Duh = inv_mt*integrate(ddot(D(uh),xi));
16     tau_h = c1*tau_h0 + c2*gamma_h + 2*c3*Duh;
17 }
18 template <class P>
19 void oldroyd_theta_scheme<P>::sub_step2 (
20     const field& uh0,
21     const field& tau_h1, const field& uh1,
22     field& tau_h, field& uh) const
23 {
24     uh = (1-theta)/theta*uh1 - (1-2*theta)/theta*uh0;
25     test xi (Th);
26     if (We == 0) {
27         field Duh = inv_mt*integrate(ddot(D(uh),xi));
28         tau_h = 2*alpha*Duh;
29         return;
30     }
31     update_transport_stress (uh);
32     form th_nu = th + nu*mt;
33     typename P::tau_upstream tau_up (Th.get_geo(), We, alpha);
34     field lh = integrate (ddot(c4*tau_h1 + 2*c5*D(uh1),xi))
35         + integrate ("boundary",
36                 max(0, -dot(uh,normal()))*ddot(tau_up,xi));
37     solver sth (th_nu.uu());
38     tau_h.set_u() = sth.solve(lh.u());
39 }
40 template <class P>
41 void
42 oldroyd_theta_scheme<P>::update_transport_stress (const field& uh) const {
43     typename P::tau_upstream tau_up (Th.get_geo(), We, alpha);
44     trial tau (Th); test xi (Th);
45     auto ma = 0.5*((1-a)*grad(uh) - (1+a)*trans(grad(uh)));
46     auto beta_a = tau*ma + trans(ma)*tau;
47     th = integrate (ddot(grad_h(tau)*uh + beta_a,xi))
48         + integrate ("boundary", max(0, -dot(uh,normal()))*ddot(tau,xi))
49         + integrate ("internal_sides",
50             - dot(uh,normal())*ddot(jump(tau),average(xi))
51             + 0.5*abs(dot(uh,normal()))*ddot(jump(tau),jump(xi)));
52     thb = integrate ("boundary", max(0, -dot(uh,normal()))*ddot(tau_up,xi));
53 }

```

### 5.3.4 Flow in an abrupt ontraction

Fig. 5.13 represents the contraction flow geometry. Let us denote by  $\Gamma_u$ ,  $\Gamma_d$ ,  $\Gamma_w$  and  $\Gamma_s$  the upstream, downstream, wall and symmetry axis boundary domains, respectively. This geometry has already been studied in the **Rheolef** documentation, volume 1 [92], section 2.2.2, in the context of a Newtonian fluid. Here, the fluid is more complex and additional boundary conditions on the upstream domain are required for the extra-stress tensor  $\boldsymbol{\tau}$ .

For the geometry, we assume that the lengths  $L_u$  and  $L_d$  are sufficiently large for the Poiseuille flows to be fully developed at upstream and downstream. Assuming also  $a = 1$ , i.e. the Oldroyd-B model, the boundary conditions at upstream are explicitly known as the solution of the fully

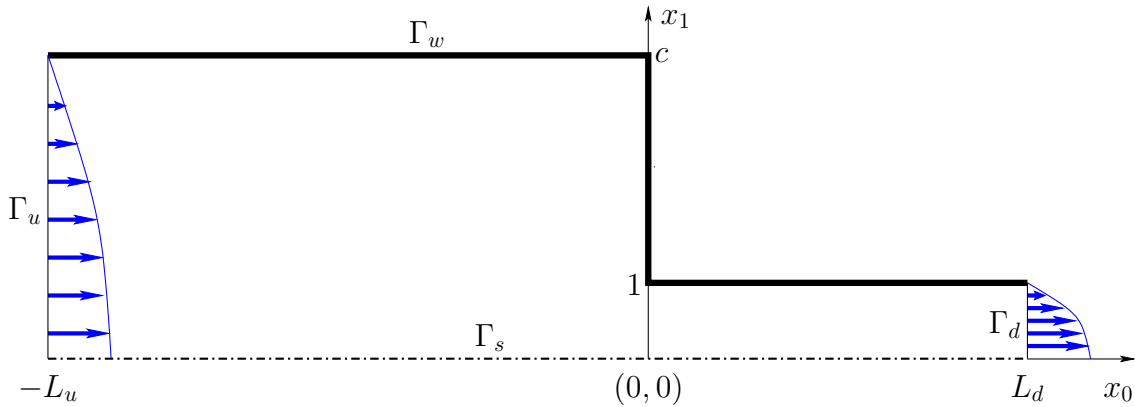


Figure 5.13: The Oldroyd problem in the abrupt contraction: schematic view of the flow domain.

Poiseuille flow for a plane pipe with half width  $c$ :

$$\begin{aligned}
 u_0(x_1) &= \bar{u} \left( 1 - \left( \frac{x_1}{c} \right)^2 \right) \\
 \dot{\gamma}(x_1) &= u'_0(x_1) = -\frac{2\bar{u}x_1}{c^2} \\
 \tau_{00}(x_1) &= 2\alpha We \dot{\gamma}^2(x_1) \\
 \tau_{01}(x_1) &= \tau_{10}(x_1) = \alpha \dot{\gamma}(x_1) \\
 \tau_{11}(x_1) &= 0
 \end{aligned}$$

where  $\bar{u}$  denotes the maximal velocity of the Poiseuille flow. Without loss of generality, thanks to a dimensional analysis, it can be adjusted with the contraction ratio  $c$  for obtaining a flow rate equal to one:

$$\bar{u} = \begin{cases} 3/(2c) & \text{for a planar geometry} \\ 4/c^2 & \text{for an axisymmetric one} \end{cases}$$

Example file 5.24: oldroyd\_contraction.icc

```

1 #include "contraction.icc"
2 struct oldroyd_contraction: contraction {
3     struct tau_upstream: base {
4         tau_upstream (geo omega, Float We1, Float alpha1)
5             : base(omega), We(We1), alpha(alpha1) {}
6         tensor operator() (const point& x) const {
7             tensor tau;
8             Float dot_gamma = - 2*base::umax*x[1]/sqr(base::c);
9             tau(0,0) = 2*alpha*We*sqr(dot_gamma);
10            tau(0,1) = tau(1,0) = alpha*dot_gamma;
11            tau(1,1) = 0;
12            return tau;
13        }
14        Float We, alpha;
15    };
16 };

```

The class `contraction`, already used for Newtonian fluids, is here reused and extended with the boundary condition function `tau_upstream`, as a derived class `oldroyd_contraction`. We are now able to write the main program for solving a viscoelastic fluid flow problem in a contraction.

Example file 5.25: oldroyd\_contraction.cc

```

1 #include "rheolef.h"
2 using namespace rheolef;
3 using namespace std;
4 #include "oldroyd_theta_scheme.h"
5 #include "oldroyd_contraction.icc"
6 int main(int argc, char**argv) {
7     environment rheolef (argc, argv);
8     cin >> noverbose;
9     oldroyd_theta_scheme<oldroyd_contraction> pb;
10    geo omega (argv[1]);
11    Float We_incr      = (argc > 2) ? atof(argv[2]) : 0.1;
12    Float We_max       = (argc > 3) ? atof(argv[3]) : 0.1;
13    Float delta_t0     = (argc > 4) ? atof(argv[4]) : 0.005;
14    string restart     = (argc > 5) ? argv[5] : "";
15    pb.tol              = 1e-3;
16    pb.max_iter        = 50000;
17    pb.alpha           = 8./9;
18    pb.a               = 1;
19    pb.delta_t         = delta_t0;
20    Float delta_t_min  = 1e-5;
21    Float We_incr_min  = 1e-5;
22    dout << catchmark("alpha") << pb.alpha << endl
23          << catchmark("a") << pb.a << endl;
24    branch even ("We", "tau", "u", "p");
25    field tau_h, uh, ph;
26    pb.initial (omega, tau_h, uh, ph, restart);
27    dout << even (pb.We, tau_h, uh, ph);
28    bool ok = true;
29    do {
30        if (ok) pb.We += We_incr;
31        derr << "# We = " << pb.We << " delta_t = " << pb.delta_t << endl;
32        field tau_h0 = tau_h, uh0 = uh, ph0 = ph;
33        ok = pb.solve (tau_h, uh, ph);
34        if (ok) {
35            dout << even (pb.We, tau_h, uh, ph);
36        } else {
37            pb.delta_t /= 2;
38            tau_h = tau_h0; uh = uh0; ph = ph0;
39            if (pb.delta_t < delta_t_min) {
40                derr << "# solve failed: decreases We_incr and retry..." << endl;
41                pb.delta_t = delta_t0;
42                We_incr /= 2;
43                pb.We -= We_incr;
44                if (We_incr < We_incr_min) break;
45            } else {
46                derr << "# solve failed: decreases delta_t and retry..." << endl;
47            }
48        }
49        derr << endl << endl;
50    } while (true);
51 }

```

The splitting element technique for the Scott-Vogelius element is implemented as an option by the command `mkgeo_contraction`. This file it is not listed here but is available in the **Rheolef** example directory. The mesh generation for an axisymmetric contraction writes:

```

mkgeo_contraction 3 -c 4 -zr -Lu 20 -Ld 20 -split
geo contraction.geo -paraview

```

This command generates a mesh for the axisymmetric 4:1 abrupt contraction with upstream and downstream length  $L_u = L_d = 20$ . Such high lengths are required for the Poiseuille flow to be fully developed at upstream and downstream for large values of  $We$ . The 3 first argument of `mkgeo_contraction` is a number that characterizes the mesh density: when increasing, the average edge length decreases. Then, the program is started:

```

make oldroyd_contraction

```



```
./oldroyd_contraction contraction.geo 0.1 10 0.01 > contraction.branch
```

The program computes stationary solutions by increasing  $We$  with the time-dependent algorithm. It performs a continuation algorithm, using solution at a lowest  $We$  as initial condition. The recurrence starts from an the Newtonian solution associated to  $We = 0$ . The others model parameters for this classical benchmark are fixed here as  $\alpha = 8/9$  and  $a = 1$  (Oldroyd-B model). The computation can take a while as there are two loops: one outer, on  $We$ , and the other inner on time, and there are two generalized Stokes subproblem and one tensorial transport one to solve at each time iteration. The inner time loop stops when the relative error is small enough. Thus, a parallel run, when available, could be a major advantage: this can be obtained by adding e.g. `mpirun -np 8` at the beginning of the command line. Recall that the time scheme is conditionally stable: the time step should be small enough for the algorithm to converge to a stationary solution. When the solver fails, it restarts with a smaller time step. Note that the present solver can be dramatically improved: by using a Newton method, as shown in [91], it is possible to directly reach the stationary solution, but such more sophisticated implementation is out of the scope of the present documentation.

The visualization of the stream function writes:

```
branch contraction.branch -toc
branch contraction.branch -extract 3 > contraction-We-0.3.field
make streamf_contraction
field -mark u contraction-We-0.3.field -field | ./streamf_contraction > psi.field
field psi.field -n-iso 15 -n-iso-negative 10
field psi.field -n-iso 15 -n-iso-negative 10 -bw
```

The file ‘`streamf_contraction.cc`’ has already been introduced in the **Rheolef** documentation, volume 1 [92], in the context of a Newtonian fluid. The result is shown on Fig. 5.14. The vortex growths with  $We$ : this is the major effet observed on this problem. Observe the two color maps representation: positive values of the stream functions are associated to the vortex and negatives values to the main flow region. The last command is a black-and-white variant view.

The vortex activity is obtained by:

```
field psi.field -max
```

Recall that the minimal value of the stream function is  $-1$ , thanks to the dimensionless procedure used here.

Cuts along the axis of symmetry are obtained by:

```
field contraction-We-0.3.field -domain axis -mark u -comp 0 -elevation -gnuplot
field contraction-We-0.3.field -domain axis -mark tau -comp 00 -elevation -gnuplot
field contraction-We-0.3.field -domain axis -mark tau -comp 11 -elevation -gnuplot
field contraction-We-0.3.field -domain axis -mark tau -comp 22 -elevation -gnuplot
```

These cuts are plotted on Fig. 5.16. Observe the overshoot of the velocity along the axis when  $We > 0$  while this effect is not perceptible when  $We = 0$ . Also, the normal extra stress  $\tau_{zz}$  growth dramatically in the entry region.

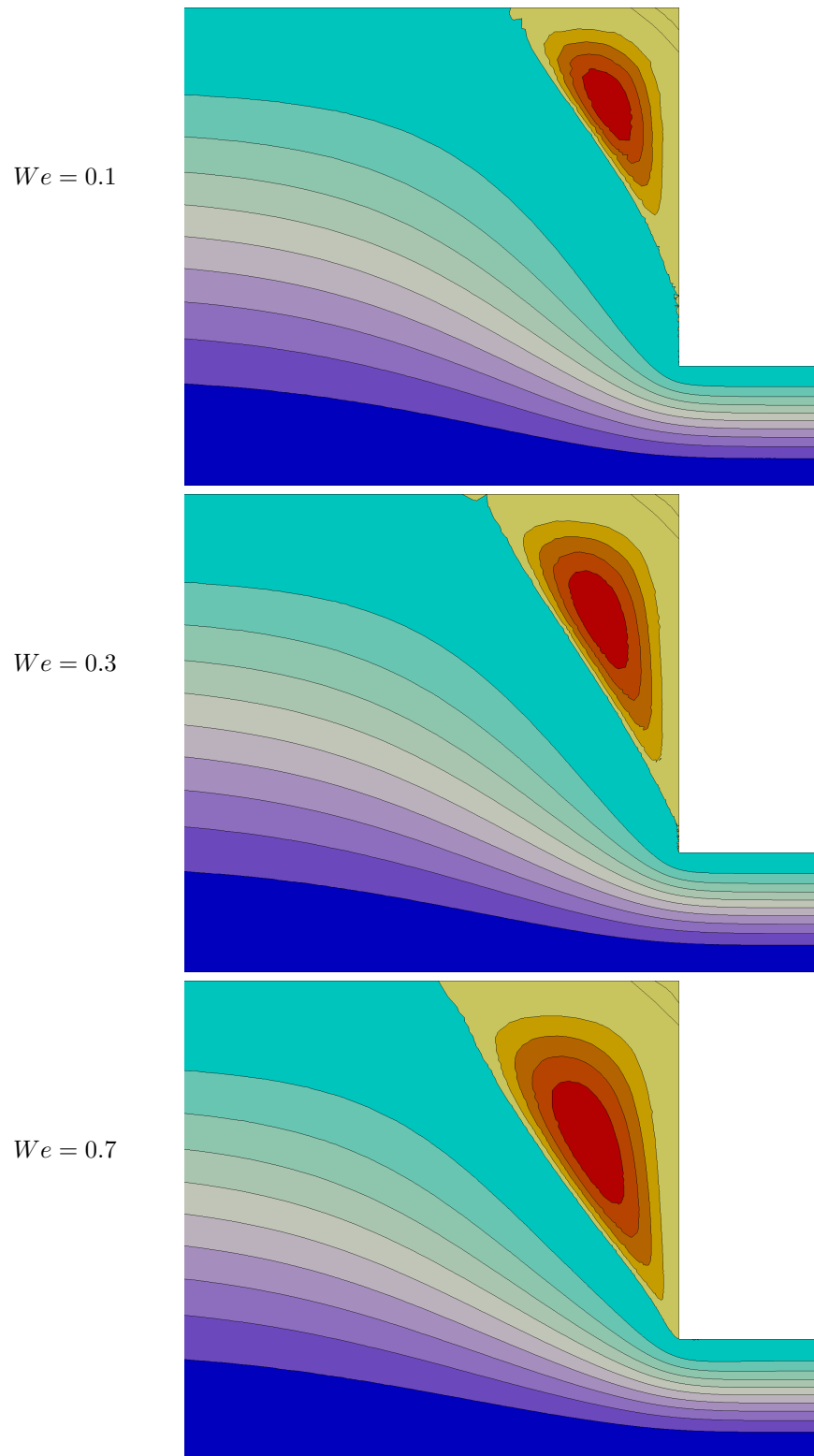


Figure 5.14: The Oldroyd problem in the axisymmetric contraction: stream function for  $We = 0.1, 0.3$  and  $0.7$ , from top to bottom.

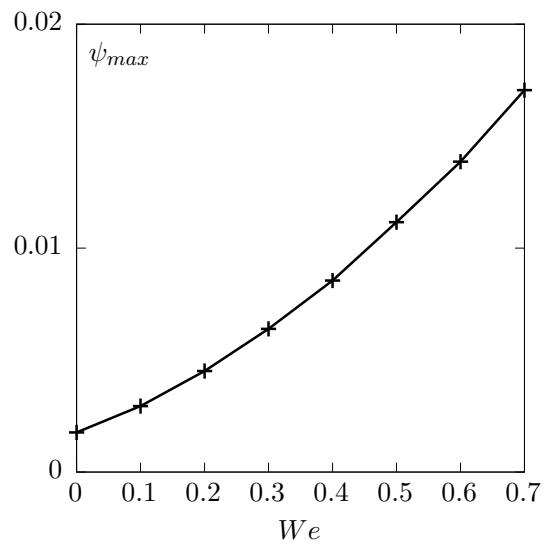


Figure 5.15: The Oldroyd-B problem in the axisymmetric contraction: vortex activity vs  $We$  ( $\alpha = 8/9$ ).

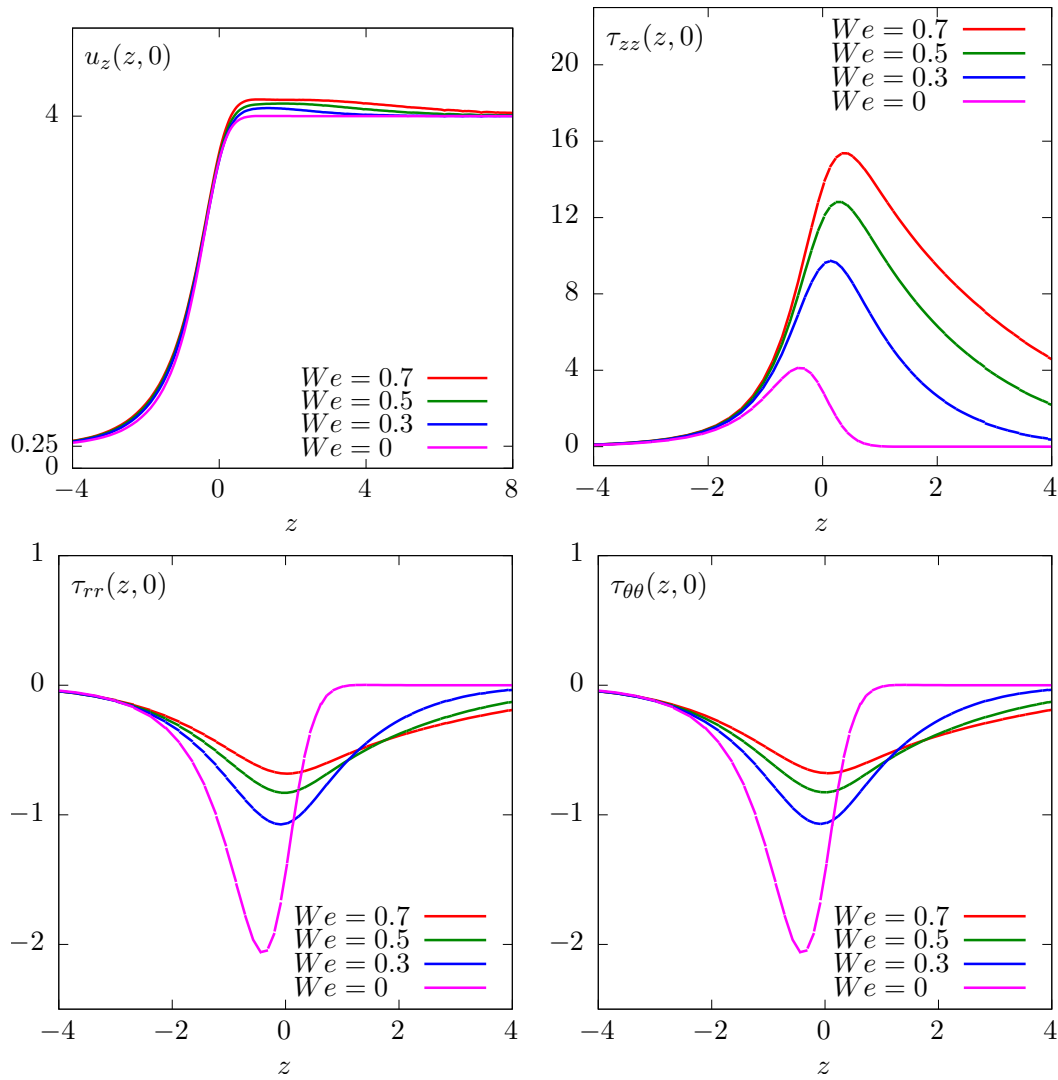


Figure 5.16: The Oldroyd problem in the axisymmetric contraction: velocity and stress components along the axis.



# Appendix A

## Technical appendices

### A.1 How to write a variational formulation ?

The major keypoint for using **Rheolef** is to put the problem in variational form. Then this variational form can be efficiently translated into C++ language. This appendix is dedicated to readers who are not fluent with variational formulations and some related functional analysis tools.

#### A.1.1 The Green formula

Let us come back to the model problem presented in section 1.1.1, page 12, equations (1.1)-(1.2) and details how this problem is transformed into (1.3).

Let  $H_0^1(\Omega)$  the space of functions whose gradient square has a finite sum over  $\Omega$  and that vanishes on  $\partial\Omega$ :

$$H_0^1(\Omega) = \{v \in L^2(\Omega); \nabla v \in L^2(\Omega)^d \text{ and } v = 0 \text{ on } \partial\Omega\}$$

We start by multiplying (1.1) by an arbitrarily test-function  $v \in H_0^1(\Omega)$  and then integrate over  $\Omega$  :

$$-\int_{\Omega} \Delta u v \, dx = \int_{\Omega} f v \, dx, \quad \forall v \in H_0^1(\Omega)$$

The next step is to invoke an integration by part, the so-called Green formula:

$$\int_{\Omega} \Delta u v \, dx + \int_{\Omega} \nabla u \cdot \nabla v \, dx = \int_{\partial\Omega} \frac{\partial u}{\partial n} v \, ds, \quad \forall u, v \in H^1(\Omega)$$

Since our test-function  $v$  vanishes on the boundary, the integral over  $\partial\Omega$  is zero and the problem becomes:

$$\int_{\Omega} \nabla u \cdot \nabla v \, dx = \int_{\Omega} f v \, dx, \quad \forall v \in H_0^1(\Omega)$$

This is exactly the variational formulation (1.3), page 12.

#### A.1.2 The vectorial Green formula

In this section, we come back to the linear elasticity problem presented in section 2.1.1, page 41, equations (2.1)-(2.2) and details how this problem is transformed into (2.3).

Let  $\Gamma_d$  (resp.  $\Gamma_n$ ) denotes the parts of the boundary  $\partial\Omega$  related to the homogeneous Dirichlet boundary condition  $\mathbf{u} = 0$  (resp. the homogeneous Neumann boundary condition  $\sigma(\mathbf{u}) \mathbf{n} = 0$ ). We suppose that  $\partial\Omega = \overline{\Gamma_d} \cap \overline{\Gamma_n}$ . Let us introduce the following functional space:

$$\mathbf{V} = \{\mathbf{v} \in H^1(\Omega)^d; \mathbf{v} = 0 \text{ on } \Gamma_d\}$$

Then, multiplying the first equation of (2.2) by an arbitrarily test-function  $\mathbf{v} \in \mathbf{V}$  and then integrate over  $\Omega$  :

$$-\int_{\Omega} \mathbf{div}(\sigma(\mathbf{u})) \cdot \mathbf{v} \, dx = \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, dx, \quad \forall \mathbf{v} \in \mathbf{V}$$

The next step is to invoke an integration by part:

$$\int_{\Omega} \mathbf{div} \tau \cdot \mathbf{v} \, dx + \int_{\Omega} \tau : D(\mathbf{v}) \, dx = \int_{\partial\Omega} \tau : (\mathbf{v} \otimes \mathbf{n}) \, ds, \quad \forall \tau \in L^2(\Omega)^{d \times d}, \quad \forall \mathbf{v} \in \mathbf{V}$$

Recall that  $\mathbf{div} \tau$  denotes  $\left( \sum_{j=0}^{d-1} \partial_j \tau_{i,j} \right)_{0 \leq i < d}$ , i.e. the vector whose component are the divergence of each row of  $\tau$ . Also,  $\sigma : \tau$  denote the double contracted product  $\sum_{i,j=0}^{d-1} \sigma_{i,j} \tau_{i,j}$  for any tensors  $\sigma$  and  $\tau$ , and that  $\mathbf{u} \otimes \mathbf{v}$  dotes the  $\tau_{i,j} = u_i v_j$  tensor, vectors  $\mathbf{u}$  and  $\mathbf{v}$ . Remark that  $\tau : (\mathbf{u} \otimes \mathbf{v}) = (\tau \mathbf{v}) \cdot \mathbf{u} = \sum_{i,j=0}^{d-1} \tau_{i,j} u_i v_j$ . Choosing  $\tau = \sigma(\mathbf{u})$  in the previous equation leads to:

$$\int_{\Omega} \sigma(\mathbf{u}) : D(\mathbf{v}) \, dx = \int_{\partial\Omega} (\sigma(\mathbf{u}) \mathbf{n}) \cdot \mathbf{v} \, ds + \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, dx, \quad \forall \mathbf{v} \in \mathbf{V}$$

Since our test-function  $v$  vanishes on  $\Gamma_d$  and the solution satisfies the homogeneous Neumann boundary condition  $\sigma(\mathbf{u}) \mathbf{n} = 0$  on  $\Gamma_n$ , the integral over  $\partial\Omega$  is zero and the problem becomes:

$$\int_{\Omega} \sigma(\mathbf{u}) : D(\mathbf{v}) \, dx = \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, dx, \quad \forall \mathbf{v} \in \mathbf{V}$$

From the definition of  $\sigma(\mathbf{u})$  in (2.1) page 41 we have:

$$\begin{aligned} \sigma(\mathbf{u}) : D(\mathbf{v}) &= \lambda \operatorname{div}(\mathbf{u}) (I : D(\mathbf{v})) + 2\mu D(\mathbf{u}) : D(\mathbf{v}) \\ &= \lambda \operatorname{div}(\mathbf{u}) \operatorname{div}(\mathbf{v}) + 2\mu D(\mathbf{u}) : D(\mathbf{v}) \end{aligned}$$

and the previous relation becomes:

$$\int_{\Omega} \lambda \operatorname{div}(\mathbf{u}) \operatorname{div}(\mathbf{v}) \, dx + \int_{\Omega} 2\mu D(\mathbf{u}) : D(\mathbf{v}) \, dx = \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, dx, \quad \forall \mathbf{v} \in \mathbf{V}$$

This is exactly the variational formulation (2.3), page 41.

### A.1.3 The Green formula on a surface

Let  $\Gamma$  a closed and orientable surface of  $\mathbb{R}^d$ ,  $d = 2, 3$  and  $\mathbf{n}$  its unit normal. From [54], appendix C we have the following integration by part:

$$\int_{\Gamma} \operatorname{div}_s \mathbf{v} \xi \, ds + \int_{\Gamma} \mathbf{v} \cdot \nabla_s \xi \, ds = \int_{\Gamma} \mathbf{v} \cdot \mathbf{n} \xi \operatorname{div} \mathbf{n} \, ds$$

for all  $\xi \in H^1(\Gamma)$  and  $\mathbf{v} \in H^1(\Gamma)^d$ . Notice that  $\operatorname{div} \mathbf{n}$  represent the surface curvature. Next, we choose  $\mathbf{v} = \nabla_s \varphi$ , for any  $\varphi \in H^2(\Gamma)$ . Remarking that  $\mathbf{v} \cdot \mathbf{n} = 0$  and that  $\operatorname{div}_s \mathbf{v} = \Delta_s \varphi$ . Then:

$$\int_{\Gamma} \Delta_s \xi \, ds + \int_{\Gamma} \nabla_s \varphi \cdot \nabla_s \xi \, ds = 0$$

This formula is the starting point for all variational formulations of problems defined on a surface (see chapter 3.1).

## A.2 How to prepare a mesh ?

Since there is many good mesh generators, **Rheolef** does not provide a built-in mesh generator. There are several ways to prepare a mesh for **Rheolef**.

We present here several procedures: by using the **bamg** bidimensional anisotropic mesh generator, written by Frédéric Hecht [46], and the **gmsh** mesh generator, suitable when  $d = 1, 2$  and  $3$ , and written by Christophe Geuzaine and Jean-François Remacle [39].

### A.2.1 Bidimensional mesh with bamg

We first create a ‘`square.bamgcad`’ file:

```
MeshVersionFormatted
0
Dimension
2
Vertices
4
0 0 1
1 0 2
1 1 3
0 1 4
Edges
4
1 2 101
2 3 102
3 4 103
4 1 104
hVertices
0.1 0.1 0.1 0.1
```

This is an uniform mesh with element size  $h = 0.1$ . We refer to the `bamg` documentation [46] for the complete file format description. Next, enter the mesh generator commands:

```
bamg -g square.bamgcad -o square.bamg
```

Then, create the file ‘`square.dmn`’ that associate names to the four boundary domains of the mesh. Here, there is four boundary domains:

```
EdgeDomainNames
4
bottom
right
top
left
```

and enter the translation command:

```
bamg2geo square.bamg square.dmn > square.geo
```

This command creates a ‘`square.geo`’ file. Look at the mesh via the command:

```
geo square
```

This presents the mesh in a graphical form, usually with `gnuplot`. You can switch to the `paraview` or `mayavi` renders:

```
geo square -paraview
geo square -mayavi
```

A finer mesh could be generated by:

```
bamg -coef 0.5 -g square.bamgcad -o square-0.5.bamg
```



### A.2.2 Unidimensional mesh with gmsh

The simplest unidimensional mesh is a line:

```
h_local = 0.1;
Point(1) = {0, 0, 0, h_local};
Point(2) = {1, 0, 0, h_local};
Line(3) = {1,2};
Physical Point("left") = {1};
Physical Point("right") = {2};
Physical Point("boundary") = {1,2};
Physical Line("interior") = {3};
```

The mesh generation command writes:

```
gmsh -1 line.mshcad -format msh -o line.msh
```

Then, the conversion to '.geo' format and the visualization:

```
msh2geo line.msh > line.geo
geo line
```

### A.2.3 Bidimensional mesh with gmsh

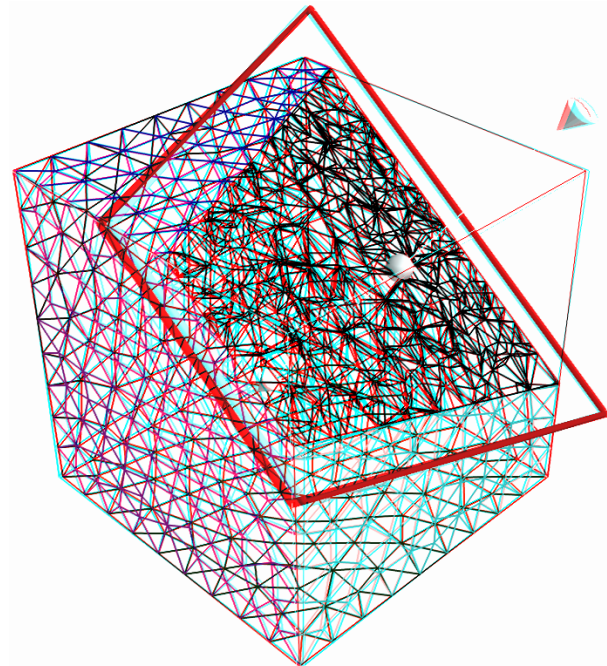
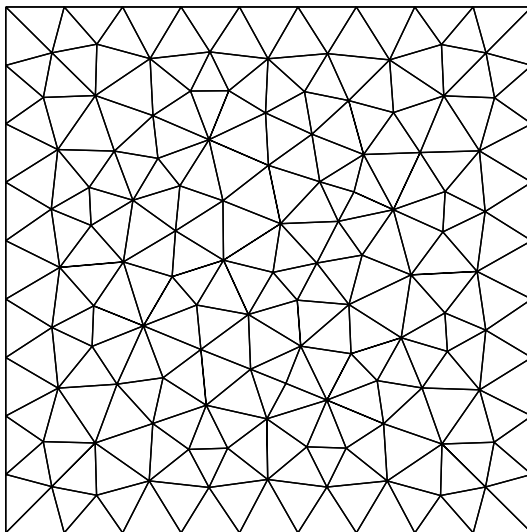


Figure A.1: Visualization of the gmsh meshes 'square.geo' and 'cube.geo'.

We first create a 'square.mshcad' file:

```
n = 10.0;
hloc = 1.0/n;
Point(1) = {0, 0, 0, hloc};
Point(2) = {1, 0, 0, hloc};
```

```

Point(3) = {1, 1, 0, hloc};
Point(4) = {0, 1, 0, hloc};
Line(1) = {1,2};
Line(2) = {2,3};
Line(3) = {3,4};
Line(4) = {4,1};
Line Loop(5) = {1,2,3,4};
Plane Surface(6) = {5} ;
Physical Point("left_bottom") = {1};
Physical Point("right_bottom") = {2};
Physical Point("right_top") = {3};
Physical Point("left_top") = {4};
Physical Line("boundary") = {1,2,3,4};
Physical Line("bottom") = {1};
Physical Line("right") = {2};
Physical Line("top") = {3};
Physical Line("left") = {4};
Physical Surface("interior") = {6};

```

This is a uniform mesh with element size  $h = 0.1$ . We refer to the `gmsh` documentation [39] for the complete file format description. Next, enter the mesh generator commands:

```
gmsh -2 square.mshcad -format msh -o square.msh
```

Then, enter the translation command:

```
msh2geo square.msh > square.geo
```

This command creates a ‘`square.geo`’ file. Look at the mesh via the command:

```
geo square
```

Remark that the domain names, defined in the `.mshcad` file, are included in the `gmsh .msh` input file and are propagated in the `.geo` by the format conversion.

### A.2.4 Tridimensional mesh with gmsh

First, create a ‘`cube.mshcad`’ file:

```

Mesh.Algorithm = 7; // bamg
Mesh.Algorithm3D = 7; // mmg3d
a = 0; c = 0; f = 0;
b = 1; d = 1; g = 1;
n = 10;
hloc = 1.0/n;
Point(1) = {a, c, f, hloc};
Point(2) = {b, c, f, hloc};
Point(3) = {b, d, f, hloc};
Point(4) = {a, d, f, hloc};
Point(5) = {a, c, g, hloc};
Point(6) = {b, c, g, hloc};
Point(7) = {b, d, g, hloc};
Point(8) = {a, d, g, hloc};
Line(1) = {1,2};
Line(2) = {2,3};

```

```

Line(3) = {3,4};
Line(4) = {4,1};
Line(5) = {5,6};
Line(6) = {6,7};
Line(7) = {7,8};
Line(8) = {8,5};
Line(9) = {1,5};
Line(10) = {2,6};
Line(11) = {3,7};
Line(12) = {4,8};
Line Loop(21) = {-1,-4,-3,-2};
Plane Surface(31) = {21} ;
Line Loop(22) = {5,6,7,8};
Plane Surface(32) = {22} ;
Line Loop(23) = {1,10,-5,-9};
Plane Surface(33) = {23} ;
Line Loop(24) = {12,-7,-11,3};
Plane Surface(34) = {24} ;
Line Loop(25) = {2,11,-6,-10};
Plane Surface(35) = {25} ;
Line Loop(26) = {9,-8,-12,4};
Plane Surface(36) = {26} ;
Surface Loop(41) = {31,32,33,34,35,36};
Volume(51) = {41};
Physical Surface("bottom") = {31};
Physical Surface("top") = {32};
Physical Surface("left") = {33};
Physical Surface("front") = {35};
Physical Surface("right") = {34};
Physical Surface("back") = {36};
Physical Volume("internal") = {51};

```

Next, enter the mesh generator commands:

```
gmsht -3 cube.mshcad -format msh -o cube.msh
```

Then, enter the translation command:

```
msh2geo cube.msh > cube.geo
```

This command creates a 'cube.geo' file. Look at the mesh via the command:

```
geo cube
geo cube.geo -cut
```

The second command allows to see inside the mesh.

### A.3 Migrating to Rheolef version 6.0

Due to its new distributed memory and computation support, **Rheolef** version 6.0 presents some backward incompatibilities with previous versions: codes using previous versions of the library should be slightly modified. This appendix presents some indications for migrating existing code.

### A.3.1 What is new in Rheolef 6.0 ?

The major main features are:

- support **distributed architectures**: the code looks sequential, is easy to read and write but can be run massively parallel and distributed, based on the MPI library.
- **high order polynomial** approximation:  $P_k$  basis are introduced in this version, for  $k \geq 0$ . This feature will be improved in the future developments.
- **mesh adaptation** and the **characteristic method** are now available for **three-dimensional** problems.

In order to evaluate in these directions, internal data structures inside the library are completely rewritten in a different way, and thus this version is a completely new library.

Conversely, the library and unix command interfaces was as less as possible modified.

Nevertheless, the user will find some few backward incompatibilities: 5.93 based codes will not directly compile with the 6.0 library version. Let us review how to move a code from 5.93 to 6.0 version.

### A.3.2 What should I have to change in my 5.x code ?

#### 1. Namespace

The namespace `rheolef` was already introduced in last 5.93 version. Recall that a code usually starts with:

```
#include "rheolef.h"
using namespace rheolef;
```

#### 2. Environment

The MPI library requires initialisation and the two command line arguments. This initialisation is performed via the `boost::mpi` class `environment`: The code entry point writes:

```
int main (int argc, char** argv) {
environment rheolef (argc,argv);
...
}
```

#### 3. Fields and forms data accessors

The accesses to unknown and blocked data was of a field `uh` was direct, as `uh.u` and `uh.b`. This access is no more possible in a distributed environment, as non-local value requests may be optimized and thus, read and write access may be controled thought accessors. These accessors are named `uh.u()` and `uh.b()` for read access, and `uh.set_u()` and `uh.set_b()` for write access. Similarly, a form `a` has accessors as `a.uu()`.

A typical 5.93 code writes:

```
ssk<Float> sa = ldl(a.uu);
uh.u = sa.solve (lh.u - a.ub*uh.b);
```

and the corresponding 6.0 code is:

```
solver sa (a.uu());
uh.set_u() = sa.solve (lh.u() - a.ub()*uh.b());
```

This major change in the library interface induces the most important work when porting to the 6.0 version.

Notice also that the old `ssk<Float>` class has been superseded by the `solver` class, that manages both direct and iterative solvers in a more effective way. For three-dimensional problems, the iterative solver is the default while direct solvers are used otherwise. In the same spirit, a `solver_abtb` has been introduced, for Stokes-like mixed problem. These features facilitate the dimension-independent coding style provided by the **Rheolef** library.

#### 4. Distributed input and output streams

Input and output *sequential* standard streams `cin`, `cout` and `cerr` may now be replaced by *distributed Rheolef* streams `din`, `dout` and `derr` as:

```
din >> omega;
dout << uh;
```

These new streams are available together with the `idiststream` and `odiststream` classes of the **Rheolef** library.

#### 5. File formats ‘.geo’ and ‘.field’ have changed

The ‘.geo’ and ‘.field’ file formats have changed. The ‘.mfield’ is now obsolete: it has been merged into the ‘.field’ format that supports now multi-component fields. Also, the corresponding `mfield` unix command is obsolete, as these features are integrated in the `field` unix command.

At this early stage of the 6.0 version, it is not yet possible to read the old ‘.geo’ format, but this backward compatibility will be assured soon.

#### 6. Space on a domain

A space defined on a domain "boundary" of a mesh `omega` was defined in the 5.93 version as:

```
space Wh (omega["boundary"], omega, "P1");
```

It writes now:

```
space Wh (omega["boundary"], "P1");
```

as the repetition of `omega` is no more required.

#### 7. Nonlinear expressions involving fields

Non-linear operations, such as `sqrt(uh)` or `1/uh` was directly supported in **Rheolef-5.x**.

```
space Xh (omega, "P1");
field uh (Xh, 2.);
field vh = 1/uh;
```

Notice that non-linear operations as `1/uh` do not return in general piecewise polynomials while `uh*uh` is piecewise quadratic. In **Rheolef-5.x**, the returned value was implicitly the Lagrange interpolant of the nonlinear expression in space `Xh`. For more clarity, **Rheolef-6.x** requires an explicit call to the `interpolate` function and the code should write:

```
field vh = interpolate (Xh, 1/uh);
```

Notice that when the expression is linear, there is no need to call `interpolate`.

### A.3.3 New features in Rheolef 6.4

The **Rheolef-6.x** code is in active developments. While backward compatibility is maintained since 6.0, some styles and idioms evolve in order to increase the expressivity and the flexibility of the interface library. Here is the summary of these evolutions.

#### 1. Nonlinear expressions

Nonlinear expressions have been extended since **Rheolef-6.4** to expression mixing `field` and functions or functors. For instance, when `u_exact` is a functor, an  $L^2$  error could be computed using a nonlinear expression submitted to the `integrate` function:

```
Float err_l2 = sqrt (integrate (omega, sqr (uh - u_exact()), qopt));
```

Until Rheolef version 6.6, the class `field_functor` was used to mark such functors. From version 6.7, the `field_functor` class is obsolete and any function or functor that is callable with a `point` as argument is valid.

#### 2. Right-hand-side specification

For specifying a right-hand-side involving `f`, previous code style, from **Rheolef-6.0** to 6.3 was using:

```
field lh = riesz (Xh, f());
```

**Rheolef-6.4** introduces:

```
test v (Xh);
field lh = integrate (f()*v);
```

This feature opens new possibilities of right-hand-side specifications, e.g. expressions involving some derivatives of the test-function `v`. The `riesz` function is no more needed: it is maintained for backward compatibility purpose.

#### 3. Form specification

For specifying a bilinear form, previous code style, from **Rheolef-6.0** to 6.3 was using a specification based on a name:

```
form a (Xh, Xh, "grad_grad");
```

**Rheolef-6.4** introduces:

```
trial u (Xh); test v (Xh);
form a = integrate (dot(grad(u), grad(v)));
```

This feature opens new possibilities for form specifications and more flexibility. The `form` specification based on a name is no more needed: it is maintained for backward compatibility purpose.



# Appendix B

## GNU Free Documentation License

Version 1.1, March 2000

Copyright © 2000 Free Software Foundation, Inc.

59 Temple Place, Suite 330, Boston, MA 02111-1307 USA

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

### Preamble

The purpose of this License is to make a manual, textbook, or other written document “free” in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of “copyleft”, which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

### Applicability and Definitions

This License applies to any manual or other work that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. The “Document”, below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as “you”.

A “Modified Version” of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A “Secondary Section” is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document’s overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (For example, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.



The “Invariant Sections” are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License.

The “Cover Texts” are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License.

A “Transparent” copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, whose contents can be viewed and edited directly and straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup has been designed to thwart or discourage subsequent modification by readers is not Transparent. A copy that is not “Transparent” is called “Opaque”.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format,  $\LaTeX$  input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML designed for human modification. Opaque formats include PostScript, PDF, proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML produced by some word processors for output purposes only.

The “Title Page” means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, “Title Page” means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text.

## Verbatim Copying

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

## Copying in Quantity

If you publish printed copies of the Document numbering more than 100, and the Document’s license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a publicly-accessible computer-network location containing a complete Transparent copy of the Document, free of added material, which the general network-using public has access to download anonymously at no charge using public-standard network protocols. If you

use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

## Modifications

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has less than five).
- State on the Title page the name of the publisher of the Modified Version, as the publisher.
- Preserve all the copyright notices of the Document.
- Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- Include an unaltered copy of this License.
- Preserve the section entitled "History", and its title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- In any section entitled "Acknowledgements" or "Dedications", preserve the section's title, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

- Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- Delete any section entitled “Endorsements”. Such a section may not be included in the Modified Version.
- Do not retitle any existing section as “Endorsements” or to conflict in title with any Invariant Section.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version’s license notice. These titles must be distinct from any other section titles. You may add a section entitled “Endorsements”, provided it contains nothing but endorsements of your Modified Version by various parties – for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

## Combining Documents

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections entitled “History” in the various original documents, forming one section entitled “History”; likewise combine any sections entitled “Acknowledgements”, and any sections entitled “Dedications”. You must delete all sections entitled “Endorsements.”

## Collections of Documents

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

## Aggregation With Independent Works

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, does not as a whole count as a Modified Version of the Document, provided no compilation copyright is claimed for the compilation. Such a compilation is called an “aggregate”, and this License does not apply to the other self-contained works thus compiled with the Document, on account of their being thus compiled, if they are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one quarter of the entire aggregate, the Document’s Cover Texts may be placed on covers that surround only the Document within the aggregate. Otherwise they must appear on covers around the whole aggregate.

## Translation

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License provided that you also include the original English version of this License. In case of a disagreement between the translation and the original English version of this License, the original English version will prevail.

## Termination

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

## Future Revisions of This License

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

\*

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright © YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.1 or any later version published by the Free Software Foundation; with the

Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST. A copy of the license is included in the section entitled “GNU Free Documentation License”.

If you have no Invariant Sections, write “with no Invariant Sections” instead of saying which ones are invariant. If you have no Front-Cover Texts, write “no Front-Cover Texts” instead of “Front-Cover Texts being LIST”; likewise for Back-Cover Texts.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

# Bibliography

- [1] E. M. Abdalass. *Résolution performante du problème de Stokes par mini-éléments, maillages auto-adaptatifs et méthodes multigrilles – applications*. PhD thesis, Thèse de l'école centrale de Lyon, 1987. [62](#)
- [2] L. Abouorm. *Méthodes mathématiques pour les écoulements sur des surfaces*. Master's thesis, M2R Université J. Fourier, Grenoble, 2010. [98](#), [99](#)
- [3] P. Alart. Méthode de Newton généralisée en mécanique du contact. *Journal de Mathématiques Pures et Appliquées*, 76(1):83–108, 1997. [192](#)
- [4] P. R. Amestoy, I. S. Duff, J.-Y. L'Excellent, and J. Koster. A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM J. Matrix Anal. Appl.*, 23(1):15–41, 2001. [19](#)
- [5] P. R. Amestoy, A. Guermouche, J.-Y. L'Excellent, and S. Pralet. Hybrid scheduling for the parallel solution of linear systems. *Parallel Comput.*, 32(2):136–156, 2006. [19](#)
- [6] D. N. Arnold, F. Brezzi, and M. Fortin. A stable finite element for the Stokes equations. *Calcolo*, 21:337–344, 1984. [60](#)
- [7] D. N. Arnold and J. Qin. Quadratic velocity/linear pressure Stokes elements. *Adv. Comput. Meth. Partial Diff. Eqn.*, 7:28–34, 1992. [217](#)
- [8] U. M. Ascher, S. J. Ruuth, and R. J. Spiteri. Implicit-explicit Runge-Kutta methods for time-dependent partial differential equations. *Appl. Numer. Math.*, 25(2):151–167, 1997. [161](#)
- [9] C. Ashcraft and J. W. H. Liu. Robust ordering of sparse matrices using multisection. *SIAM J. Matrix Anal. Appl.*, 19(3):816–832, 1998. [19](#)
- [10] F. Auteri, N. Parolini, and L. Quartapelle. Numerical investigation on the stability of singular driven cavity flow. *J. Comput. Phys.*, 183(1):1–25, 2002. [83](#)
- [11] G. K. Batchelor. *An introduction to fluid dynamics*. Cambridge university press, UK, sixth edition, 1967. [67](#)
- [12] R. Bird, R. C. Armstrong, and O. Hassager. *Dynamics of polymeric liquids. Volume 1. Fluid mechanics*. Wiley, New-York, second edition, 1987. [207](#)
- [13] H. Borouchaki, P. L. George, F. Hecht, P. Laug, B. Mohammadi, and E. Saltel. Mailleur bidimensionnel de Delaunay gouverné par une carte de métriques. Partie II: applications. Technical Report RR-2760, INRIA, 1995. [47](#)
- [14] K. Boukir, Y. Maday, B. Metivet, and E. Razafindrakoto. A high-order characteristic/finite element method for the incompressible Navier-Stokes equations. *Int. J. Numer. Meth. Fluids*, 25:1421–1454, 1997. [79](#)
- [15] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*. Springer, second edition, 2002. [208](#)

- [16] H. Brezis. *Analyse fonctionnelle. Théorie et application*. Masson, Paris, 1983. [105](#)
- [17] F. Brezzi and J. Pitkäranta. On the stabilization of finite element approximation of the Stokes equations. In *Efficient solutions of elliptic systems, Kiel, Notes on numerical fluid mechanics*, volume 10, pages 11–19, 1984. [62](#)
- [18] M. P. Calvo, J. de Frutos, and J. Novo. Linearly implicit Runge-Kutta methods for advection-reaction-diffusion equations. *Appl. Numer. Math.*, 37(4):535–549, 2001. [161](#)
- [19] G. F. Carey and B. Jiannng. Least-squares finite elements for first-order hyperbolic systems. *Int. J. Numer. Meth. Eng.*, 26(1):81–93, 1988. [148](#)
- [20] P. Castillo. Performance of discontinuous Galerkin methods for elliptic PDEs. *SIAM J. Sci. Comput.*, 24(2):524–547, 2002. [154](#)
- [21] M. J. Castro-Diaz, F. Hecht, B. Mohammadi, and O. Pironneau. Anisotropic unstructured mesh adaption for flow simulations. *Int. J. Numer. Meth. Fluids*, 25(4):475–491, 1997. [47](#)
- [22] J. C. Crispell, V. J. Ervin, and E. W. Jenkins. A fractional step  $\theta$ -method approximation of time-dependent viscoelastic fluid flow. *J. Comput. Appl. Math.*, 232(2):159–175, 2009. [217](#)
- [23] B. Cockburn. *An introduction to the discontinuous Galerkin method for convection-dominated problems*, chapter 2, pages 151–268. Springer, 1998. [146](#)
- [24] B. Cockburn, B. Dong, J. Guzmán, and J. Qian. Optimal convergence of the original DG method on special meshes for variable transport velocity. *SIAM J. Numer. Anal.*, 48(1):133–146, 2010. [143](#)
- [25] B. Cockburn, S. Hou, and C.-W. Shu. The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. IV. the multidimensional case. *Math. Comput.*, 54(190):545–581, 1990. [148](#)
- [26] B. Cockburn, G. Kanschat, and D. Schötzau. A locally conservative LDG method for the incompressible Navier-Stokes equations. *Math. Comput.*, 74(251):1067–1095, 2005. [175](#)
- [27] B. Cockburn, G. Kanschat, D. Schötzau, and C. Schwab. Local discontinuous Galerkin methods for the Stokes system. *SIAM J. Numer. Anal.*, 40(1):319–343, 2002. [167](#)
- [28] M. Crouzeix and J. Rappaz. *On numerical approximation in bifurcation theory*. Masson, Paris, 1990. [126](#)
- [29] K. Deckelnick, G. Dziuk, C.M. Elliott, and C.-J. Heine. An h-narrow band finite element method for elliptic equations on implicit surfaces. *IMA Journal of Numerical Analysis*, to appear:0, 2009. [89](#)
- [30] D. A. di Pietro and A. Ern. Discrete functional analysis tools for discontinuous Galerkin methods with application to the incompressible Navier-Stokes equations. *Math. Comp.*, 79:1303–1330, 2010. [167](#), [171](#), [175](#)
- [31] D. A. di Pietro and A. Ern. *Mathematical aspects of discontinuous Galerkin methods*. Springer, 2012. [141](#), [142](#), [144](#), [145](#), [154](#), [156](#), [157](#), [161](#), [168](#), [171](#), [175](#), [180](#)
- [32] M. Dicko. Méthodes mathématiques pour les écoulements sur des surfaces. Master’s thesis, M2P Université J. Fourier, Grenoble, 2011. [98](#)
- [33] J. Donea and A. Huerta. *Finite element methods for flow problems*. Wiley, New-York, 2003. [86](#)

- [34] Y. Epshteyn and B. Rivière. Estimation of penalty parameters for symmetric interior penalty Galerkin methods. *J. Comput. Appl. Math.*, 206(2):843–872, 2007. [154](#)
- [35] E. Erturk, T. C. Corke, and C. Gökçol. Numerical solutions of 2-D steady incompressible driven cavity flow at high Reynolds numbers. *Int. J. Numer. Meth. Fluids*, 48:747–774, 2005. [83](#)
- [36] G. Fourestey and S. Piperno. A second-order time-accurate ALE Lagrange-Galerkin method applied to wind engineering and control of bridge profiles. *Comput. Methods Appl. Mech. Engrg.*, 193:4117–4137, 2004. [79](#)
- [37] T. Gelhard, G. Lube, M. A. Olshanskii, and J. H. Starcke. Stabilized finite element schemes with LBB-stable elements for incompressible flows. *J. Comput. Appl. Math.*, 177:243–267, 2005. [83](#)
- [38] A. George. Nested dissection of a regular finite element mesh. *SIAM J. Numer. Anal.*, 10:345–363, 1973. [19](#)
- [39] C. Geuzaine and J.-F. Remacle. Gmsh: a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities. *Int. J. Numer. Meths Engrg.*, 79(11):1309–1331, 2009. [230](#), [233](#)
- [40] U. Ghia, K. N. Ghia, and C. T. Shin. High  $Re$  solutions for incompressible flow using the Navier-Stokes equations and a multigrid method. *J. Comput. Phys.*, 48:387–411, 1982. [83](#), [85](#)
- [41] V. Girault and P. A. Raviart. *Finite element methods for the Navier-Stokes equations. Theory and algorithms*. Springer, 1986. [31](#), [55](#), [56](#), [57](#)
- [42] S. Gottlieb and C.-W. Shu. Total variation diminishing Runge-Kutta schemes. *Math. Comput.*, 67(221):73–85, 1998. [145](#), [146](#)
- [43] S. Gottlieb, Chi-W. Shu, and E. Tadmor. Strong stability-preserving high-order time discretization methods. *SIAM review*, 43(1):89–112, 2001. [145](#), [146](#)
- [44] M. M. Gupta and J. C. Kalita. A new paradigm for solving Navier-Stokes equations: streamfunction-velocity formulation. *J. Comput. Phys.*, 207:52–68, 2005. [83](#)
- [45] A. Harten, B. Engquist, S. Osher, and S. R. Chakravarthy. Uniformly high order accurate essentially non-oscillatory schemes, III. *J. Comput. Phys.*, 71(2):231–303, 1987. [148](#)
- [46] F. Hecht. *BAMG: bidimensional anisotropic mesh generator*, 2006. <http://www.ann.jussieu.fr/~hecht/ftp/bamg>. [47](#), [230](#), [231](#)
- [47] J. S. Hesthaven and T. Warburton. *Nodal discontinuous Galerkin methods. Algorithms, analysis and applications*. Springer, 2008. [141](#)
- [48] A. J. Hoffman, M. S. Martin, and D. J. Rose. Complexity bounds for regular finite difference and finite element grids. *SIAM J. Numer. Anal.*, 10(2):364–369, 1973. [19](#)
- [49] P. Hood and C. Taylor. A numerical solution of the Navier-Stokes equations using the finite element technique. *Comp. and Fluids*, 1:73–100, 1973. [52](#), [59](#), [80](#)
- [50] J. S. Howell. Computation of viscoelastic fluid flows using continuation methods. *J. Comput. Appl. Math.*, 225(1):187–201, 2009. [138](#)
- [51] Jr. J. E. Dennis and R. B. Schnablel. *Numerical methods for unconstraint optimization and nonlinear equations*. Prentice Hall, Englewood Cliff, N. J., 1983. [120](#)



- [52] C. Johnson and J. Pitkäranta. An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation. *Math. Comp.*, 46(173):1–26, 1986. 143
- [53] A. Klawonn. An optimal preconditioner for a class of saddle point problems with a penalty term. *SIAM J. Sci. Comput.*, 19(2):540–552, 1998. 53, 60
- [54] A. Laadhari, C. Misbah, and P. Saramito. On the equilibrium equation for a generalized biological membrane energy by using a shape optimization approach. *Phys. D*, 239:1568–1572, 2010. 230
- [55] R. J. Labeur and G. N. Wells. A Galerkin interface stabilisation method for the advection-diffusion and incompressible Navier-Stokes equations. *Comput. Meth. Appl. Mech. Engrg.*, 196(49–52):4985–5000, 2007. 86
- [56] S. Melchior, V. Legat, P. Van Dooren, and A. J. Wathen. Analysis of preconditioned iterative solvers for incompressible flow problems. *Int. J. Numer. Meth. Fluids*, 68(3):269–286, 2012. 83
- [57] P. D. Mineev and C. R. Ethier. A characteristic/finite element algorithm for the 3-D Navier-Stokes equations using unstructured grids. *Comput. Meth. in Appl. Mech. and Engrg.*, 178(1-2):39–50, 1998. 83
- [58] P. P. Mosolov and V. P. Miasnikov. Variational methods in the theory of the fluidity of a viscous-plastic medium. *J. Appl. Math. Mech.*, 29(3):545–577, 1965. 200
- [59] P. P. Mosolov and V. P. Miasnikov. On stagnant flow regions of a viscous-plastic medium in pipes. *J. Appl. Math. Mech.*, 30(4):841–853, 1966. 200
- [60] P. P. Mosolov and V. P. Miasnikov. On qualitative singularities of the flow of a viscoplastic medium in pipes. *J. Appl. Math. Mech.*, 31(3):609–613, 1967. 200
- [61] D. R. Musser and A. Saini. *C++ STL tutorial and reference guide*. Addison Wesley, Reading, 1996. 23
- [62] D. R. Musser and A. Saini. *STL tutorial and reference guide*. Addison-Wesley, 1996. 75
- [63] J. G. Oldroyd. On the formulation of rheological equations of states. *Proc. R. Soc. Lond. A*, 200:523–541, 1950. 216
- [64] M. A. Olshanskii and A. Reusken. A finite element method for surface PDEs: matrix properties. *Numer. Math.*, 114:491–520, 2010. 99
- [65] M. A. Olshanskii, A. Reusken, and J. Grande. A finite element method for elliptic equations on surfaces. *SIAM J. Num. Anal.*, 47(5):3339–3358, 2009. 93, 98, 99
- [66] M. A. Olshanskii, A. Reusken, and X. Xu. On surface meshes induced by level set functions. *Computing and visualization in science*, 15(2):53–60, 2012. 96
- [67] M. L. Ould Salihi. *Couplage de méthodes numériques en simulation directe d’écoulements incompressibles*. PhD thesis, Université J. Fourier, Grenoble, 1998. 83
- [68] O. Ozenda, P. Saramito, and G. Chambon. A new rate-independent tensorial model for suspensions of non-colloidal rigid particles in newtonian fluids. *HAL preprint*, pages hal-01528817, 2017. <https://hal.archives-ouvertes.fr/hal-01528817v3/document>. 213
- [69] C. C. Paige and M. A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal.*, 12(4):617–629, 1975. 31
- [70] T.-W. Pan, J. Hao, and R. Glowinski. On the simulation of a time-dependent cavity flow of an Oldroyd-B fluid. *Int. J. Numer. Meth. Fluids*, 60(7):791–808, 2009. 217

- [71] J.-C. Paumier. *Bifurcation et méthodes numériques. Applications aux problèmes elliptiques semi-linéaires*. Masson, Paris, 1997. 126, 129, 132, 135
- [72] F. Pellegrini. *PT-Scotch and libscotch 5.1 user's guide*. Université de Bordeaux and INRIA, France, 2010. 19
- [73] T. E. Peterson. A note on the convergence of the discontinuous Galerkin method for a scalar hyperbolic equation. *SIAM J. Numer. Anal.*, 28(1):133–140, 1991. 143
- [74] O. Pironneau. *Méthode des éléments finis pour les fluides*. Masson, Paris, 1988. 74
- [75] O. Pironneau and M. Tabata. Stability and convergence of a galerkin-characteristics finite element scheme of lumped mass type. *Int. J. Numer. Meth. Fluids*, 64:1240–1253, 2010. 75
- [76] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C. The art of scientific computing*. Cambridge University Press, UK, second edition, 1997. Version 2.08. 120
- [77] P. A. Raviart and J. M. Thomas. *Introduction à l'analyse numérique des équations aux dérivées partielles*. Masson, Paris, 1983. 108
- [78] G. R. Richter. An optimal-order error estimate for the discontinuous galerkin method. *Math. Comput.*, 50(181):75–88, 1988. 143
- [79] N. Roquet, R. Michel, and P. Saramito. Errors estimate for a viscoplastic fluid by using Pk finite elements and adaptive meshes. *C. R. Acad. Sci. Paris, ser. I*, 331(7):563–568, 2000. 47, 207, 208
- [80] N. Roquet and P. Saramito. An adaptive finite element method for Bingham fluid flows around a cylinder. *Comput. Meth. Appl. Mech. Engrg.*, 192(31-32):3317–3341, 2003. 205
- [81] N. Roquet and P. Saramito. Stick-slip transition capturing by using an adaptive finite element method. *M2AN*, 38(2):249–260, 2004. 187, 192
- [82] N. Roquet and P. Saramito. An adaptive finite element method for viscoplastic flows in a square pipe with stick-slip at the wall. *J. Non-Newt. Fluid Mech.*, 155:101–115, 2008. 200, 205
- [83] H. Rui and M. Tabata. A second order characteristic finite element scheme for convection diffusion problems. *Numer. Math. (to appear)*, 2001. 74
- [84] P. Saramito. *Simulation numérique d'écoulements de fluides viscoélastiques par éléments finis incompressibles et une méthode de directions alternées; applications*. PhD thesis, Institut National Polytechnique de Grenoble, 1990. 67, 217
- [85] P. Saramito. Numerical simulation of viscoelastic fluid flows using incompressible finite element method and a  $\theta$ -method. *Math. Model. Numer. Anal.*, 28(1):1–35, 1994. 213, 216, 218
- [86] P. Saramito. Efficient simulation of nonlinear viscoelastic fluid flows. *J. Non Newt. Fluid Mech.*, 60:199–223, 1995. 217
- [87] P. Saramito. Operator splitting in viscoelasticity. *Élasticité, Viscoélasticité et Contrôle Optimal, Lyon, décembre 1995, ESAIM: Proceedings*, 2:275–281, 1997. 217
- [88] P. Saramito. **Rheolef** home page. <https://www-ljk.imag.fr/membres/Pierre.Saramito/rheolef>, 2012. 23
- [89] P. Saramito. *Language C++ et calcul scientifique*. College Publications, London, 2013. 19

- [90] P. Saramito. *Méthodes numériques en fluides complexes : théorie et algorithmes*. CNRS-CCSD, 2013. <http://cel.archives-ouvertes.fr/cel-00673816>. 213
- [91] P. Saramito. On a modified non-singular log-conformation formulation for Johnson-Segalman viscoelastic fluids. *J. Non-Newt. Fluids Mech.*, 211:16–30, 2014. 217, 224
- [92] P. Saramito. *Efficient C++ finite element computing with Rheolef*. CNRS and LJK, 2015. <http://cel.archives-ouvertes.fr/cel-00573970>. 182, 191, 197, 221, 224
- [93] P. Saramito. *Efficient C++ finite element computing with Rheolef: volume 2: discontinuous Galerkin methods*. CNRS and LJK, 2015. <http://cel.archives-ouvertes.fr/cel-00863021>. 83, 213
- [94] P. Saramito. *Complex fluids: modelling and algorithms*. Springer, 2016. 105, 187, 200, 213, 216
- [95] P. Saramito. A damped Newton algorithm for computing viscoplastic fluid flows. *J. Non-Newt. Fluid Mech.*, 238:6–15, 2016. 205
- [96] P. Saramito and N. Roquet. An adaptive finite element method for viscoplastic fluid flows in pipes. *Comput. Meth. Appl. Mech. Engrg.*, 190(40-41):5391–5412, 2001. 18, 200, 201, 205, 208
- [97] P. Saramito and A. Wachs. Progress in numerical simulation of yield stress fluid flows. *J. Rheol.*, 56(3):211–230, 2017. 200
- [98] Pierre Saramito. Are curved and high order gmsh meshes really high order ?, 2012. <http://www.geuz.org/pipermail/gmsh/2012/006967.html>. 94
- [99] L. R. Scott and M. Vogelius. Norm estimates for a maximal right inverse of the divergence operator in spaces of piecewise polynomials. *M2AN*, 19(1):111–143, 1985. 217
- [100] N. D. Scurtu. *Stability analysis and numerical simulation of non-Newtonian fluids of Oldroyd kind*. PhD thesis, U. Nürnberg, Deutschland, 2005. 217
- [101] K. Shahbazi. An explicit expression for the penalty parameter of the interior penalty method. *J. Comput. Phys.*, 205(2):401–407, 2005. 154
- [102] J. Shen. Hopf bifurcation of the unsteady regularized driven cavity flow. *J. Comp. Phys.*, 95:228–245, 1991. <http://www.math.purdue.edu/~shen/pub/Cavity.pdf>. 83
- [103] C.-W. Shu. TVB boundary treatment for numerical solutions of conservation laws. *Math. Comput.*, 49(179):123–134, 1987. 147
- [104] C.-W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *J. Comput. Phys.*, 77(2):439–471, 1988. 145, 146
- [105] P. Singh and L. G. Leal. Finite-element simulation of the start-up problem for a viscoelastic fluid in an eccentric rotating cylinder geometry using a third-order upwind scheme. *Theor. Comput. Fluid Dyn.*, 5(2-3):107–137, 1993. 217
- [106] B. Stroustrup. C++ programming styles and libraries. *InformIt.com*, 0:0, 2002. 3
- [107] G. I. Taylor. On the decay of vortices in a viscous fluid. *Philos. Mag.*, 46:671–674, 1923. 166
- [108] M.-G. Vallet. Génération de maillages anisotropes adaptés, application à la capture de couches limites. Technical Report RR-1360, INRIA, 1990. 47
- [109] H. Wang, C.-W. Shu, and Q. Zhang. Stability analysis and error estimates of local discontinuous Galerkin methods with implicit–explicit time-marching for nonlinear convection–diffusion problems. *Appl. Math. Comput.*, 2015. 161

- 
- [110] H. Wang, C.-W. Shu, and Q. Zhang. Stability and error estimates of local discontinuous Galerkin methods with implicit-explicit time-marching for advection-diffusion problems. *SIAM J. Numer. Anal.*, 53(1):206–227, 2015. 161, 163
- [111] Wikipedia. The Stokes stream function. *Wikipedia*, 2012. [http://en.wikipedia.org/wiki/Stokes\\_stream\\_function](http://en.wikipedia.org/wiki/Stokes_stream_function). 67
- [112] N. Wirth. *Algorithm + data structure = programs*. Prentice Hall, NJ, USA, 1985. 3
- [113] S. Zhang. A new family of stable mixed finite elements for the 3d Stokes equations. *Math. Comput.*, 74(250):543–554, 2005. 217
- [114] X. Zhong and C.-W. Shu. A simple weighted essentially nonoscillatory limiter for Runge-Kutta discontinuous galerkin methods. *J. Comput. Phys.*, 232(1):397–415, 2013. 151

# List of example files

Makefile, 12  
burgers.icc, 146  
burgers\_dg.cc, 148  
burgers\_diffusion\_dg.cc, 159  
burgers\_diffusion\_exact.icc, 158  
burgers\_diffusion\_operators.icc, 160  
burgers\_flux\_godunov.icc, 146  
cavity.icc, 50  
cavity\_dg.icc, 177  
combustion.h, 124  
combustion1.icc, 124  
combustion2.icc, 125  
combustion\_continuation.cc, 130  
combustion\_error.cc, 128  
combustion\_exact.icc, 128  
combustion\_keller.cc, 136  
combustion\_keller\_post.cc, 137  
combustion\_newton.cc, 126  
contraction.icc, 61  
convect.cc, 72  
convect\_error.cc, 74  
cosinusprod.icc, 23  
cosinusprod\_error.cc, 23  
cosinusprod\_laplace.icc, 21  
cosinusrad.icc, 25  
cosinusrad\_laplace.icc, 24  
d\_projection\_dx.icc, 192  
dirichlet-nh.cc, 20  
dirichlet.cc, 10  
dirichlet.icc, 107  
dirichlet\_dg.cc, 153  
elasticity\_criterion.icc, 47  
elasticity\_solve.icc, 47  
elasticity\_taylor\_dg.cc, 163  
embankment.cc, 40  
embankment.icc, 41  
embankment\_adapt.cc, 46  
eta.icc, 106  
harten.icc, 147  
harten\_show.cc, 147  
heat.cc, 69  
helmholtz\_band.cc, 98  
helmholtz\_band\_iterative.cc, 96  
helmholtz\_s.cc, 86  
incompressible-elasticity.cc, 57  
inertia.icc, 169  
inertia\_cks.icc, 173  
inertia\_upw.icc, 178  
lambda2alpha.icc, 128  
lambda\_c.cc, 127  
lambda\_c.icc, 127  
laplace\_band.cc, 99  
laplace\_s.cc, 90  
level\_set\_sphere.cc, 93  
mosolov\_augmented\_lagrangian.cc, 201  
mosolov\_augmented\_lagrangian.h, 200  
mosolov\_augmented\_lagrangian1.icc, 199  
mosolov\_augmented\_lagrangian2.icc, 200  
mosolov\_error\_yield\_surface.cc, 207  
mosolov\_exact\_circle.icc, 205  
mosolov\_yield\_surface.cc, 203  
navier\_stokes\_cavity.cc, 79  
navier\_stokes\_criterion.icc, 79  
navier\_stokes\_dg.h, 175  
navier\_stokes\_dg1.icc, 175  
navier\_stokes\_dg2.icc, 177  
navier\_stokes\_solve.icc, 78  
navier\_stokes\_taylor\_dg.cc, 170  
navier\_stokes\_taylor\_newton\_dg.cc, 175  
navier\_stokes\_upw\_dg.h, 179  
navier\_stokes\_upw\_dg.icc, 179  
neumann-laplace.cc, 30  
neumann-nh.cc, 26  
neumann\_dg.cc, 155  
nu.icc, 115  
oldroyd\_contraction.cc, 220  
oldroyd\_contraction.icc, 220  
oldroyd\_theta\_scheme.h, 216  
oldroyd\_theta\_scheme1.icc, 217  
oldroyd\_theta\_scheme2.icc, 217  
oldroyd\_theta\_scheme3.icc, 218  
p\_laplacian.h, 113  
p\_laplacian1.icc, 114  
p\_laplacian2.icc, 115  
p\_laplacian\_circle.icc, 122  
p\_laplacian\_damped\_newton.cc, 119

p\_laplacian\_error.cc, 122  
p\_laplacian\_fixed\_point.cc, 104  
p\_laplacian\_newton.cc, 113  
phi.icc, 187  
poisson\_robin.icc, 188  
proj\_band.cc, 97  
projection.icc, 186  
robin.cc, 28  
rotating-hill.h, 73  
sgn.icc, 179  
sinusprod\_helmholtz.icc, 27  
sphere.icc, 87  
stokes\_cavity.cc, 50  
stokes\_contraction\_bubble.cc, 60  
stokes\_dirichlet\_dg.icc, 166  
stokes\_taylor\_dg.cc, 166  
streamf\_cavity.cc, 55  
streamf\_contraction.cc, 64  
stress.cc, 43  
taylor.icc, 164  
torus.icc, 90  
transmission.cc, 33  
transport\_dg.cc, 141  
transport\_tensor\_dg.cc, 212  
vector\_projection.icc, 199  
vortex\_position.cc, 81  
vorticity.cc, 53  
yield\_slip.h, 193  
yield\_slip1.icc, 193  
yield\_slip2.icc, 194  
yield\_slip\_augmented\_lagrangian.cc,  
188  
yield\_slip\_augmented\_lagrangian.icc,  
188  
yield\_slip\_damped\_newton.cc, 192  
cavity.icc, 80  
contraction.icc, 220  
contraction.mshcad, 62, 66  
convect.cc, 79  
cosinusprod\_error\_dg.cc, 155  
cosinusrad\_error.cc, 25  
cube.mshcad, 231  
dirichlet\_nh\_ball.cc, 25  
elasticity\_taylor\_error\_dg.cc, 164  
harten0.icc, 147  
helmholtz\_s\_error.cc, 89  
line.mshcad, 49, 230  
mkview\_mosolov, 204  
mosolov\_error.cc, 206  
mosolov\_residue.cc, 202  
navier\_stokes\_cavity\_newton\_dg.cc, 177  
navier\_stokes\_cavity\_newton\_upw\_dg.cc,  
179  
navier\_stokes\_taylor\_error\_dg.cc, 171  
neumann-nh.cc, 86  
runge\_kutta\_semiimplicit.icc, 159  
runge\_kutta\_ssp.icc, 144  
square.bamgcad, 48, 80, 229  
square.dmn, 48, 80, 229  
square.mshcad, 230  
stokes\_contraction.cc, 64  
stokes\_taylor\_error\_dg.cc, 167  
streamf\_cavity.cc, 80  
streamf\_contraction.cc, 66, 222  
stress.cc, 68  
taylor.icc, 167, 171  
taylor\_exact.icc, 164  
torus.mshcad, 91  
transport\_tensor\_error\_dg.cc, 213  
transport\_tensor\_exact.icc, 213  
yield\_slip\_circle.icc, 196  
yield\_slip\_error.cc, 196  
yield\_slip\_residue.cc, 190

# List of commands

bamg2geo, 229  
bamg, 48, 80, 229  
    -splitpbedge, 108  
branch, 73  
    -gnuplot, 71, 161  
    -paraview, 71, 75  
    -toc, 131  
    -umax, 161  
    -umin, 161  
convect, 73  
field, 13, 234  
    -, 15  
    -bw, 13, 64, 67, 80, 97  
    -comp, 43, 45, 67, 68, 80  
    -cut, 67, 68, 80, 107, 190  
    -domain, 190  
    -elevation, 13, 45, 97, 107, 190  
    -fill, 43  
    -gnuplot, 13, 67, 68, 80, 107,  
        190  
    -gray, 13  
    -mark, 52  
    -max, 65, 81, 131  
    -min, 81  
    -n-iso, 64, 204  
    -n-iso-negative, 64, 67, 80, 204,  
        222  
    -noclean, 56  
    -noexecute, 56  
    -nofill, 13, 42, 48  
    -normal, 67, 68, 80, 107, 190  
    -origin, 67, 68, 80, 107, 190  
    -proj, 44, 204  
    -scale, 58, 80  
    -stereo, 13, 16, 43, 45, 87, 97  
    -velocity, 52, 80  
    -volume, 16, 22  
geo, 13  
    -cut, 15  
    -fill, 15  
    -full, 15  
    -gnuplot, 92  
    -mayavi, 229  
    -paraview, 221, 229  
    -shrink, 15  
    -stereo, 15, 87  
    -subdivide, 88  
gmsh, 49, 63, 66, 87, 91, 108, 230  
gnuplot, 13, 15, 34, 43, 71, 74, 229  
gzip, 48  
library  
    boost, 11, 31, 233  
    CGAL, computational geometry, 72  
    MPI, message passing interface,  
        233  
    MPI, message passing interface, 11  
    mumps, linear system direct solver, 17  
    scotch, mesh partition library, 17  
    STL, standard template library, 73  
make, 12  
man, 14  
mayavi, 45, 229  
mkgeo\_ball, 87  
    -e, 87  
    -q, 25  
    -s, 87  
    -t, 25, 87  
mkgeo\_contraction, 221  
    -split, 221  
mkgeo\_grid, 13, 73, 108  
    -H, 16  
    -T, 15, 76  
    -a, 74  
    -b, 74  
    -c, 75  
    -d, 75  
    -e, 15  
    -f, 76  
    -g, 76  
    -q, 16  
    -region, 34  
    -t, 13  
    -zr, 65  
mkgeo\_sector, 203  
mkgeo\_ugrid, 23, 108  
mpirun, 18, 31, 34, 80, 203, 221, 222  
msh2geo, 63, 230, 231  
    -zr, 66

---

paraview, [13](#), [15](#), [16](#), [44](#), [45](#), [56](#), [71](#),  
[75](#), [204](#), [229](#)  
rheolef-config, [9](#)  
    -check, [9](#)  
    -docdir, [9](#)  
    -exampledir, [9](#)  
sed, [25](#)  
time, [80](#)  
visualization  
    mesh, [13](#)  
    deformed, [42](#)  
vlc, [71](#)  
zcat, [80](#)



# Index

- approximation, 10
  - P0, 45, 140
  - P1b-P1, 58
  - P1d, 45
  - P1, 11, 41, 45, 57, 58, 76, 167
  - P2-P1, Taylor-Hood, 50, 58, 59, 64
  - P2-P1d, Scott-Vogelius, 215
  - P2, 11, 14, 41, 45, 57, 76, 167
  - Pk, 11, 14, 233
  - bubble, 58
  - discontinuous, 33, 44, 45, 53, 139, 211
  - high-order, 11, 14, 34, 233
  - isoparametric, 88
  - mixed, 50
  - geometry
    - curved, 24
    - high-order, 24, 206
    - isoparametric, 24, 206
- argc, argv, command line arguments, 11, 233
- benchmark
  - driven cavity flow, 49, 76, 165, 167
  - Dziuk-Elliott-Heine on a sphere, 87
  - embankment, 39, 163
  - flow in an abrupt contraction, 59, 66
  - Olshanskii-Reusken-Grande on a torus, 91
  - pipe flow, 103
  - rotating hill, 72
- boundary condition
  - Dirichlet, 10, 20, 33, 39, 49, 56, 76, 103, 112, 152, 155, 165, 167, 185
  - mixed, 56, 59, 66
  - Neumann, 25, 29, 33, 39, 56
  - Poiseuille flow, 62
  - Robin, 27, 185
  - weakly imposed, 139, 152, 212, 220
- broken Sobolev space  $H^1(\mathcal{T}_h)$ , 152, 158
- class
  - Float, 21
  - adapt\_option, 48
  - band, 96
  - branch, 69, 75
  - characteristic, 72
  - communicator, 31
  - csr<T>, 31
  - environment, 233
  - eye, 97
  - field\_functor, 235
  - field, 11, 233
  - form, 11
  - geo, 11
  - idiststream, 234
  - integrate\_option, 60, 66
  - level\_set\_option, 93
  - odiststream, 47, 234
  - point, 21
  - quadrature\_option, 72, 106, 123
  - solver\_abtb, 79, 234
  - solver\_option, 16
  - solver, 12, 16, 51, 234
  - space, 11, 234
  - vec<T>, 31
- compilation, 12
- convergence
  - error
    - versus mesh, 22, 89, 121, 132, 141, 154
    - versus polynomial degree, 22, 89, 121
  - residue
    - rate, 108, 109
    - super-linear, 115
- coordinate system
  - axisymmetric, 62, 64, 65
  - Cartesian, 21, 34
  - spherical, 87
  - torus, 91
- directory of example files, 9, 62, 64, 196
- distributed computation, 11, 18, 31, 203, 221, 233
- element shape, 34
- error analysis, 22, 75, 89, 121, 132
- file format
  - ‘.bamgcad’ bamg geometry, 48, 229
  - ‘.bamg’ bamg mesh, 48, 229

- ‘.branch’ family of fields, 70
- ‘.dmn’ domain names, 229
- ‘.field’ field, 13, 234
- ‘.field’ multi-component field, 42, 234
- ‘.geo’ mesh, 13, 63, 65, 229–231, 234
- ‘.gz’ gzip compressed file, 48
- ‘.mshcad’ gmsh geometry, 49, 63, 66, 91, 230
- ‘.msh’ gmsh mesh, 49, 63, 66, 230
- ‘.ogv’ ogg vorbis/theora file (video), 71
- ‘.vtk’ vtk file, 56, 71
- form
  - $(\eta \nabla u) \cdot \nabla v$ , 114
  - $2D(\mathbf{u}) : D(\mathbf{v})$ , 50, 56
  - $2D(\mathbf{u}) : D(\mathbf{v}) + \lambda \operatorname{div} \mathbf{u} \operatorname{div} \mathbf{v}$ , 41
  - $2D(\mathbf{u}) : D(\mathbf{v}) + \mathbf{u} \cdot \mathbf{v}$ , 77
  - $\eta \nabla u \cdot \nabla v$ , 34, 105
  - $\llbracket u \rrbracket \{\{\nabla_h v \cdot \mathbf{n}\}\}$ , 152, 155, 159
  - $\llbracket u \rrbracket \{\{v\}\}$ , 140
  - $\llbracket u \rrbracket \llbracket v \rrbracket$ , 140, 152, 155, 159
  - $\nabla_s u \cdot \nabla_s v + uv$ , 86
  - $\nabla u \cdot \nabla v$ , 10
  - $\nabla u \cdot \nabla v + uv$ , 27
  - $uv$ , 23
  - $\operatorname{bcurl}(\mathbf{u}) \cdot \boldsymbol{\xi}$ , 64
  - $\operatorname{curl}(\mathbf{u}) \cdot \boldsymbol{\xi}$ , 53
  - $\operatorname{div}(\mathbf{u}) q$ , 50, 56
  - energy, 10, 33, 103
  - product, 60
  - weighted, 34, 105
    - quadrature formula, 106, 114
    - tensorial weight, 114
- formal adjoint, 192
- Fréchet derivative, 112
- function
  - adapt, 45, 48
  - catchmark, 31, 41, 52, 69
  - compose, 72, 80, 105, 114
  - damped\_newton, 119
  - diag, 97
  - dis\_wall\_time, 16
  - grad, 105, 114
  - integrate, 11, 27, 60, 64, 72, 86, 105, 114, 123, 235
    - on a band, 97
    - on the boundary, 27
  - interpolate, 22, 47, 75, 234
  - ldlt, 31
  - level\_set, 93
  - newton, 113
  - norm2, 47, 79, 105
  - riesz, 235
  - sqr, 47
  - class-function object, 21, 73, 115
  - functor, 22, 123, 235
  - geometry
    - axisymmetric, 62, 64
    - circle, 25, 87
    - contraction, 59, 66, 219
    - cube, 15, 231
    - curved, 88
    - line, 15, 230
    - pipe, 185, 198
    - sphere, 87
    - square, 13, 230
    - surface, 85
      - curvature, 228
    - torus, 91
  - Green formula, 86, 227
  - internal sides of a mesh, 140
  - Lagrange
    - interpolation, 20, 22, 26, 30
    - multiplier, 29, 51
    - node, 12
  - Lamé coefficients, 39
  - Makefile, 12
  - matrix
    - bloc-diagonal
      - inverse, 60
    - block structure, 12
    - concatenation, 31
    - diagonal, 97
    - factorization
      - Choleski, 12
    - identity, 97
    - indefinite, 31
    - singular, 31, 97
    - sparse, 31
  - mesh, 11, 228
    - adaptation, 203, 206, 233
      - anisotropic, 45, 80
    - connected components, 98
    - generation, 63, 66, 228
  - method
    - augmented Lagrangian, 185, 198
    - BDF2 scheme, 77
    - characteristic, 71, 77, 233
    - conjugate gradient algorithm, 12, 29, 51, 57
    - continuation, 130
    - Euler explicit scheme, 143
    - Euler implicit scheme, 68, 71
    - fixed-point, 103, 116
      - relaxation, 104, 110

- level set, 85, 93
  - banded, 96
- minres algorithm, 29, 97
- Newton, 112
  - damped, 118
- newton, 124, 127, 130, 187, 190, 203
- Runge-Kutta scheme, 143, 159
- upwind scheme, 178
  
- namespace
  - rheolef, 11, 233
  - std, 11
- norm
  - in  $W^{-1,p}$ , 108
    - discrete version, 108
  - in  $W^{1,p}$ , 103
  - in  $W_0^{1,p}$ , 103
  
- operator
  - average, accross sides, 140, 152
  - jump, accross sides, 140, 152
  - adjoint, 118
  - curl, 53
  - divergence, 39
  - gradient, 39
    - symmetric part, 39
  - Helmholtz, 25
  - Helmholtz-Beltrami, 85
  - Laplace, 10
  - Laplace-Beltrami, 85
  
- parallel computation, 11, 18, 31, 221
- penalty parameter, 152
- polar coordinate system, 65
- preconditioner, 51
  - Choleski incomplete factorization, 12
  - for nearly incompressible elasticity, 57
  - for Stokes problem, 51
  
- problem
  - Bingham, 198
  - Helmholtz, 25
  - Herschel-Bulkley, 198
  - Mosolov, 198
  - Navier-Stokes, 76, 167
  - Poisson, 10, 20, 27, 29, 69, 103, 105, 152, 155, 185
  - Stokes, 49, 58, 76, 165, 167
    - combustion, 124
  - convection-diffusion, 71
  - elasticity, 39, 163
  - heat, 68
  - linear tangent, 112
  - nonlinear, 76, 167
  - p-Laplacian, 103
  - stabilized Stokes, 60
  - transmission, 32
  - yield slip, 185
  - Burgers equation, 146
  - elasticity
    - incompressible, 56
  - hyperbolic nonlinear equation, 142
  - oldroyd, 214
  - Poisson
    - non-constant tensorial coefficients, 112
  - transport equation
    - steady, 139
    - tensor, 211
    - unsteady, 71, 75
  
- projection, 45, 54
  
- quadrature formula, 106
- quadrature formulae
  - Gauss, 73
  - Gauss-Lobatto, 73
  
- region, 32, 34
- residual term, 108, 112
  
- singular solution, 54
- space
  - $W^{-1,p}$ , 103
  - $W^{-1,p}$ , dual of  $W_0^{1,p}$ , 108
  - $W^{1,p}$ , 103
  - $W_0^{1,p}$ , 103
  - dual, 108
  - duality bracket  $\langle \cdot, \cdot \rangle$ , 108
  - weighted (axisymmetric), 65
  
- speedup, 18
- stabilization, 56
- stream function, 54, 64, 80
  - axisymmetric, 65
  
- tensor
  - Cauchy stress, 39, 68
  - field, 44
  - rate of deformation, 67
  - visualization as ellipsoid, 44
  
- unknow and blocked components, 12
- upstream boundary, 139
- upwinding, 140
  
- variable
  - derr, 234
  - din, 234
  - dout, 234
  
- visualization
  - animation, 71

- 
- elevation view, [13](#), [107](#)
  - stereoscopic anaglyph, [13](#), [43](#)
  - vortex, [64](#), [67](#)
  - vorticity, [53](#)