

FINITE ELEMENT METHODS WITH PATCHES AND APPLICATIONS

THÈSE N° 3478 (2006)

PRÉSENTÉE LE 17 MARS 2006
À LA FACULTÉ SCIENCES DE BASE
CHAIRE D'ANALYSE ET DE SIMULATION NUMÉRIQUE
SECTION DE MATHÉMATIQUES

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Joël WAGNER

ingénieur physicien diplômé EPF
et de nationalité luxembourgeoise

acceptée sur proposition du jury:

Prof. Ch.-E. Pfister, président du jury
Prof. J. Rappaz, directeur de thèse
Prof. C. Bernardi, rapporteur
Prof. J. He, rapporteur
Prof. A. Quarteroni, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Lausanne, EPFL

2006

Abstract

Theoretical and numerical aspects of multi-scale problems are investigated.

On one hand, mathematical analysis is done on a new method for numerically solving problems with multi-scale behavior using multiple levels of not necessarily nested grids. A particularly flexible multiplicative Schwarz method is presented, requiring no conformity between the meshes at the different scales. The relaxed iterative method consists in calculating successive corrections to the solution in regions where the variations of a problem are too strong to be captured by a coarse initial mesh. In these sub-domains patches of finite elements are applied. *A priori* and *a posteriori* error estimates are given and an exact spectral analysis of the iteration operator describing the algorithm is presented. Computational issues are addressed and numerical methods to obtain optimal convergence are given. Crucial implementation matters are discussed with special regard to usage of memory and CPU-time.

On the other hand, the efficiency of the introduced correction method is demonstrated on Laplace model problems, either with changing Dirichlet-Neumann boundary conditions or in a polygonal domain with entrant corner. The regularity of the solutions is studied as well as the improvement of the convergence order in the mesh size using various sizes of patches. The correction algorithm is also applied to improve the accuracy in the simulation of the stress field in glacier modeling. A simple model to obtain the effective stress field in the ice mass of a glacier is presented and concluding results are obtained using patches in the regions where changes in the basal boundary conditions are involved.

Version abrégée

Nous nous intéressons à quelques aspects théoriques et numériques de problèmes multi-échelles.

Dans la première partie de cette thèse, nous effectuons une analyse mathématique d'une nouvelle méthode pour résoudre numériquement des problèmes avec données multi-échelles utilisant plusieurs niveaux de grilles non nécessairement emboîtées. Nous présentons une méthode multiplicative de type Schwarz, particulièrement souple dans le sens qu'elle ne requiert pas de conformité entre les maillages aux différents niveaux. La méthode itérative relaxée consiste à calculer des corrections successives de la solution par régions où les variations du problème sont trop importantes pour être résolues sur une grille grossière initiale. Dans ces sous-domaines nous appliquons des patchs d'éléments finis. Nous donnons des estimations d'erreur *a priori* et *a posteriori*, et présentons une analyse spectrale complète de l'opérateur d'itération décrivant l'algorithme. Nous considérons les problèmes de calcul et proposons des méthodes numériques pour obtenir la convergence optimale. Nous discutons les points cruciaux dans l'implantation avec une attention particulière quant à l'utilisation de la mémoire et du temps CPU.

Dans la seconde partie de cette thèse, nous démontrons l'efficacité de la méthode de correction sur des problèmes modèle de Laplace, avec changement de conditions au bord du type Dirichlet-Neumann ou dans un domaine polygonal avec coin entrant. Nous étudions la régularité des solutions ainsi que l'amélioration de l'ordre de convergence dans la taille des éléments de la grille en utilisant différentes grandeurs de patchs. Nous appliquons également l'algorithme de correction pour améliorer la précision de la simulation du champ des contraintes dans la modélisation de glaciers. Un modèle simple pour obtenir la contrainte effective dans la masse de glace d'un glacier est présenté et nous obtenons des résultats concluants en utilisant des patchs dans les régions où des changements dans les conditions de bord à la base sont impliqués.

Acknowledgments

I am very much indebted to Prof. Jacques Rappaz for having accepted to be my thesis director and me being part of his research group. He introduced me to numerical analysis since I have been a 1st year student in Physics at EPFL. Through his confidence in me he gave me the opportunity to work on a very interesting topic, to collaborate with the University of Houston and to participate in many conferences. I have learned a lot from his suggestions and remarks, as well in research or in teaching. Merci Maître.

I am most grateful to Prof. Jiwen He from the University of Houston for his support during my two stays in Houston and for being part of my jury. Merci pour sa sympathie.

It is a pleasure to thank Prof. Alexei Lozinski for his kind and frequent support. I also thank Dr Marco Picasso and my colleague Vittoria Rezzonico for sharing their point of view on this work.

Prof. Alfio Quarteroni and Prof. Christine Bernardi who honored me by being part of my jury and by reading this work are gratefully acknowledged. Thanks also to Prof. Charles-Edouard Pfister, president of the jury.

I thank Hon. Prof. Philippe Choquard for the shared research interests in Theoretical Physics and our ongoing activity during my thesis. Merci pour toutes les discussions.

At this point, I also want to thank all the present and former members of the team of Numerical Analysis and Simulation. Thanks for the good atmosphere inside our group.

Finally I would like to thank all my friends who contributed from far or near to making this period rich and interesting. Special thanks to Mr Toshio Tamano and my friends from Goju-Ryu Shorei-Kan Karate.

Merci à Nicolas Michel pour son amitié.

Zu gudder lescht e grouse Merci u meng Elteren a mäi Brudder fir hier Ënnerstützung a Verständnis während all menge Studiejoeren.

Contents

Introduction	1
1 Two-scale algorithm	5
1.1 Situation	6
1.2 A priori and a posteriori error estimates	9
1.3 The algorithm	15
1.4 Some properties of vector spaces	18
1.5 Spectral analysis of the iteration operator	22
1.6 Convergence of the algorithm	32
2 Parameter discussion and computational considerations	35
2.1 Estimates of the C.B.S. constant	36
2.2 Numerical evaluation of $\tilde{\gamma}$ and optimal relaxation	42
2.3 Implementation issues and memory and CPU-time usage	48
3 Analysis of the algorithm on two Poisson problems	57
3.1 Preliminary results	58
3.2 Problem with change in boundary conditions	60
3.3 Problem in a domain with entrant corner	71
4 Application to glacier modeling	79
4.1 Introduction	80
4.2 Governing field equations and numerical model	82
4.3 Patches and precision of the glacier stress field	94
Conclusion	101
Bibliography	103
Appendix. On a class of solutions of the continuity and Euler's equations for inviscid and compressible fluids	109

Introduction

Despite increasing computational power making simulations faster, and available memory enabling the treatment of larger problems, the need for efficient computational methods is always of crucial importance. Simulating more complex systems with always better precision is primary for many industries and scientific applications, naming only the widely known issues of the airplane industry or meteorologic forecast and air quality management.

Very often a great or better precision is required in certain regions of the computational domain in which the solution is defined. In order to fix the ideas, let us focus on a couple of model problems.

First, consider situations with multi-scale data, in the form of a sharp right-hand side or sharp coefficients in the differential operator in a small region of the domain. Point sources for example give rise to models needing careful examination of the space-scale. Getting an accurate simulation on large scales is linked to a simulation in subregions around the sources using finer grids.

We can also think of problems whose solution present singularities arising from changing Dirichlet-Neumann boundary conditions, alike for velocity of the ice of a glacier on its basal surface where adhesion or free sliding conditions are to be found and give rise to locally high stress fields. Computational domains with entrant corners inherit similar irregularities.

Finally, engineering problems with complex geometry are of concern: take, e.g., an aluminium production cell, where locally more precision is needed but the meshing of the cell is provided a priori and re-meshing has to be avoided.

Efficient approaches in the above mentioned situations include for instance adaptive mesh refinement techniques or domain decomposition methods. However, the first are sometimes inappropriate as the meshing of the problem can be given once for all a priori, or, e.g. in three dimensions, mesh generation can be time-consuming. The second cover a broad spectrum of algorithms but many of them are not always as flexible as needed (see [53] for an overview¹).

¹Note that many methods exist in the literature to treat suchlike problems by similar methods then the one that will be presented in this thesis. References appearing in this work, although citing the most

In this work we investigate a new method for numerically solving elliptic problems with multi-scale behavior using multiple levels of not necessarily nested grids. The algorithm is a domain decomposition method and is closest to a Chimera method [25, 60]. Being a multiplicative Schwarz method, it can be compared to the Fast Adaptive Composite grid method by McCormick [49], a hierarchical method (introduced first by Yserentant [69]), or the Successive Subspace Correction method by Xu [64]. However, by further and detailed comparison done in this work, we claim that the presented method requiring no conformity between the meshes is of much more flexible use. We have published this algorithm originally in [37], and discussed and illustrated it in [36] and [35]. We consider situations where the multi-scale behavior originates from one of the problems described above. Our relaxed iterative method consists in calculating successive corrections to the solution in critical regions where we apply finite element patches whose discretizations are not necessarily conforming.

This thesis consists of two parts. In a first part we introduce the method and investigate its properties from a theoretical point of view. The second part focuses on the application to model situations and the modeling of glaciers.

The first part of this thesis (Chapters 1 and 2) is devoted to theoretical studies related to the presented algorithm.

The objective of Chapter 1 is to introduce the method. The algorithm which we first described in [37] is a method for numerically solving elliptic problems with multi-scale data using two levels of not necessarily nested grids. We use a relaxed iterative method which consists in calculating successive corrections to the solution in patches of finite elements. First we introduce a two-scale setting and its notation. We next give *a priori* and *a posteriori* error estimates for this situation and introduce the two-level algorithm itself which consists in the following: first we solve the problem on a coarse mesh of the computational domain. Therein we consider a patch with corresponding fine mesh wherein we would like to obtain more accuracy. Thus we calculate successively corrections to the solution in the patch. The latter, consecutively computed in the patch (fine correction) and over the whole domain (coarse correction), are iteratively added to the solution using a relaxation parameter. Finally, we are left with the analysis of the method. After considering a general setting of vector spaces and deriving properties for the spectral analysis of the iteration operator describing the algorithm, we establish the spectral properties of the operator and conclude with the convergence of the proposed algorithm including the idea of how to optimize the relaxation with respect to the convergence speed. We consider an alternative introducing two relaxation parameters for the coarse and fine corrections

pertinent, give only a non-exhaustive list of examples.

respectively and show that optimality requires both parameters to be equal.

In Chapter 2 we discuss the practical usage of the algorithm. Obtained results show that the speed of convergence depends mainly on two parameters: the grids characterized by an abstract angle between their respective finite element spaces, and the relaxation. We give estimates for the first of the parameters in order to optimize the convergence properties of the method. We illustrate the influence of nested and non-nested grid constellations in one dimension. We give estimates in some particular two-dimensional cases and briefly consider a particular 2D or 3D situation where the patch is entirely included in one element of the coarse grid. Next we show how to evaluate numerically the best relaxation parameter and what is the influence of patches size on the convergence of the method. Finally, we care about implementation and computational issues, in particular concerning integration of the scalar products, and assess the convergence of the method in practice with respect to the usage of memory and CPU-time.

The second part of this thesis (Chapters 3 and 4) treats applications to model problems and contributes to part of the problem of glacier modeling.

In Chapter 3 we use the introduced correction method and consider regularity and convergence order issues for problems with singularities due to changing Dirichlet-Neumann boundary conditions or domains with entrant corners: we consider a Laplace problem with changing Dirichlet-Neumann boundary conditions and a Poisson-Dirichlet problem on a polygonal domain with entrant corner. In both cases, we study the regularity of the solutions. We analyze how the application of patches improves the quality of the solution efficiently with respect to the usage of memory. We also study the convergence orders in the mesh sizes obtained for the two models and various types of patches.

The objective of Chapter 4 is to introduce the correction algorithm into the modeling of glaciers. Following the work of Reist [56] we consider a 2D vertical cut in the direction of the motion of the glacier and present a model to simulate the velocity field of the ice mass and the effective stress field. Basal boundary conditions are a crucial issue of the study of glaciers and the model involves Dirichlet and Neumann boundary conditions. With the knowledge acquired through the model problems treated in Chapter 3 we apply patches in certain regions on the glacier domain. We present an improvement in the precision of the stress field on the model of the Gries glacier (Swiss Alps).

Note.

This thesis is supplemented with an appendix out of context. A short review of research done in Theoretical Physics on a class of solutions of the continuity and Euler's equations for inviscid and compressible fluids is presented. The work aims to describe the dynamics of conservative, very large and dense systems experiencing strongly correlated motion of their constituents which interact via long range potentials, and this, by means of canonically conjugated collective variables.

Chapter 1

Two-scale algorithm

The objective of this chapter is to introduce a new method first described in [37]. It is an algorithm for numerically solving elliptic problems with multi-scale data using two levels of not necessarily nested grids. We use a relaxed iterative method which consists in calculating successive corrections to the solution in patches of finite elements. Some parts of this chapter concerning the analysis of the spectral properties of the iteration operator are extracted from [36], a paper in collaboration with Glowinski, He, Lozinski and Rappaz. We conclude with the convergence of the algorithm.

The outline of this chapter is the following:

1.1	Situation	6
1.2	A priori and a posteriori error estimates	9
1.3	The algorithm	15
1.4	Some properties of vector spaces	18
1.5	Spectral analysis of the iteration operator	22
1.6	Convergence of the algorithm	32

In Section 1.1 we introduce a two-scale setting and its notation. In the next section we give *a priori* and *a posteriori* error estimates for the introduced situation. In Section 1.3 we introduce the two-level algorithm. Hence we are left with the analysis of the latter. We consider in Section 1.4 a general setting of vector spaces. We derive properties used afterwards in the spectral analysis of the iteration operator describing the algorithm. In Section 1.5 we establish the spectral properties of the operator and, finally, in Section 1.6 we prove the convergence of the proposed algorithm.

1.1 Situation

In numerical approximation of elliptic problems by finite element method, a great precision of solutions is often required in certain regions of the domain in which the solution is defined. Our objective is to present a method to solve numerically elliptic problems with multi-scale data using two levels of not necessarily nested grids.

Consider a multi-scale problem with sharp data in small sub-domains. We solve the problem on a coarse mesh of the computational domain Ω . Therein we consider a patch Λ (or multiple patches, see [36, §5]) with corresponding fine mesh wherein we would like to obtain more accuracy. Thus we calculate successively corrections to the solution in the patch. The coarse and fine discretizations are not necessarily conforming, as illustrated in two dimensions in Figure 1.1.

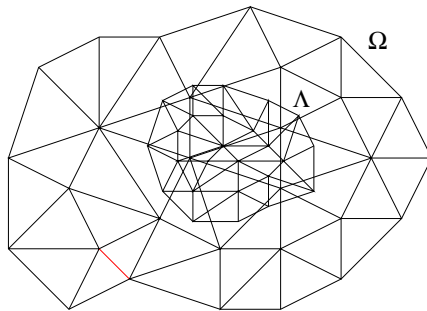


Figure 1.1: Two-scale situation: Computational domain Ω and patch Λ .

Before presenting the algorithm, we introduce a two-scale setting and its notation.

Let $\Omega \subset \mathbb{R}^d$, $d = 2$ or 3 , be an open polygonal or polyhedral domain and consider a bilinear, symmetric, continuous and coercive form

$$a : H_0^1(\Omega) \times H_0^1(\Omega) \rightarrow \mathbb{R}. \quad (1.1)$$

Here and in the sequel we use standard notation for the Sobolev spaces: $H^1(\Omega)$ denotes the usual Sobolev space of functions with first derivatives in $L^2(\Omega)$. The subscript 0 indicates the subspace of functions with trace zero on the boundary $\partial\Omega$.

The usual $H^1(\Omega)$ -norm in $H_0^1(\Omega)$ is equivalent to the a -norm defined by $\|v\| = a(v, v)^{\frac{1}{2}}$, $\forall v \in H_0^1(\Omega)$. If $f \in H^{-1}(\Omega)$, due to Riesz' representation Theorem there exists a unique $u \in H_0^1(\Omega)$ such that

$$a(u, \varphi) = \langle f | \varphi \rangle, \quad \forall \varphi \in H_0^1(\Omega), \quad (1.2)$$

where $\langle \cdot | \cdot \rangle$ denotes the duality $H^{-1}(\Omega) - H_0^1(\Omega)$. Let us point out that (1.2) is the weak formulation of a problem of type

$$\begin{cases} \mathcal{L}(u) = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \quad (1.3)$$

where $\mathcal{L}(\cdot)$ is a second order, linear, symmetric, strongly elliptic operator. In the following we consider problems of the form (1.3) with homogeneous Dirichlet boundary conditions. If the Dirichlet boundary conditions of the initial problem are not homogeneous, we extend the discussion of the current situation. In Section 3.2 we also consider a situation with non-homogeneous Dirichlet and homogeneous Neumann boundary conditions.

For instance, an operator $\mathcal{L}(\cdot)$ for problem (1.3) can be given by

$$\mathcal{L}(u)(x) = - \sum_{i,j=1}^d \frac{\partial}{\partial x_i} \left(a_{ij}(x) \frac{\partial u}{\partial x_j}(x) \right), \quad (1.4)$$

where $a_{ij} \in L^\infty(\Omega)$, $a_{ij}(x) = a_{ji}(x)$, $1 \leq i, j \leq d$, and $\sum_{i,j=1}^d a_{ij}(x) \xi_i \xi_j \geq \alpha |\xi|^2, \forall \xi \in \mathbb{R}^d, \forall x \in \Omega$, where α is a positive constant. In this case the form $a(\cdot, \cdot)$ is defined by

$$a(\psi, \varphi) = \sum_{i,j=1}^d \int_{\Omega} a_{ij} \frac{\partial \psi}{\partial x_j} \frac{\partial \varphi}{\partial x_i} dx, \quad \forall \psi, \varphi \in H_0^1(\Omega). \quad (1.5)$$

Our objective is to find an approximation of the solution $u \in H_0^1(\Omega)$ of problem (1.2). A Galerkin approximation consists in

- building a finite dimensional subspace $V_{Hh} \subset H_0^1(\Omega)$, and
- solving the problem: Find $u_{Hh} \in V_{Hh}$ satisfying

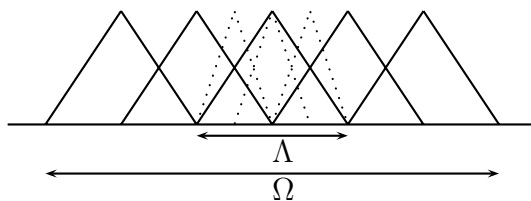
$$a(u_{Hh}, \varphi) = \langle f | \varphi \rangle, \quad \forall \varphi \in V_{Hh}. \quad (1.6)$$

In the following the construction of the space V_{Hh} is presented. Firstly, we introduce a regular family of triangulations \mathcal{T}_H of $\bar{\Omega}$ (see Ciarlet [28, Sect. 7] or [29, Sect. 17]), a union of triangles K of diameter less than or equal to H .

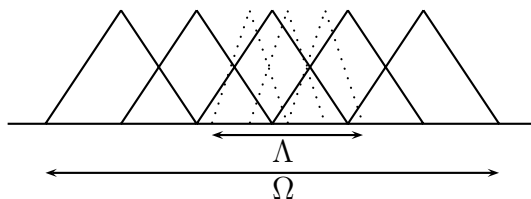
Consider now in two dimensions a multi-scale situation with a solution that is very sharp, i.e. varies rapidly, in a small polygonal sub-domain Λ of Ω , but smooth, i.e. varies slowly, in $\Omega \setminus \Lambda$. This means that the solution can be well approximated on a coarse mesh in $\Omega \setminus \Lambda$ but needs a fine mesh in Λ . We would like to stress that $\bar{\Lambda}$ is not necessarily the union of several triangles K of \mathcal{T}_H . Besides Λ can be determined in practice by an a priori knowledge of the solution behavior or an a posteriori error estimator (Proposition 1.2), for example. Let \mathcal{T}_h be a regular family of triangulations of $\bar{\Lambda}$ with triangles K such that $h = \max_{K \in \mathcal{T}_h} \text{diam}(K)$.

In a first step we approximate u by a finite element method of order r on the triangulation \mathcal{T}_H of $\bar{\Omega}$ by using

$$V_H = \{ \psi \in H_0^1(\Omega) : \psi|_K \in \mathbb{P}_r(K), \forall K \in \mathcal{T}_H \}, \quad (1.7)$$



(a) Nested elements.



(b) Non-nested elements.

Figure 1.2: Linear finite elements in 1D on Ω (plain lines) and Λ (dotted lines).

where $\mathbb{P}_r(K)$ is the space of polynomials of degree $\leq r$ on triangle K . In a second step we rectify the solution on the finite element space

$$V_h = \{\psi \in H_0^1(\Omega) : \psi|_K \in \mathbb{P}_s(K), \forall K \in \mathcal{T}_h \text{ and } \psi = 0 \text{ in } \bar{\Omega} \setminus \Lambda\}, \quad (1.8)$$

where \mathcal{T}_h is the triangulation of $\bar{\Lambda}$.

Consequently we set $V_{Hh} = V_H + V_h$. We observe that in practice, it is generally not possible to determine a finite element basis of V_{Hh} . Hence it is impossible to compute directly u_{Hh} . The goal of our method is to evaluate efficiently u_{Hh} without having a basis of V_{Hh} , but only a basis of V_H and a basis of V_h .

Let us mention that a priori $V_H \cap V_h$ does not necessarily reduce to the element zero as shown in Figure 1.2(a) where a one-dimensional situation is illustrated by the “hat” functions in Ω and in Λ . In the case when \mathcal{T}_H and \mathcal{T}_h are not nested, as illustrated by Figure 1.2(b) where we have translated the patch, it is not possible to easily exhibit a finite element-basis of V_{Hh} from the bases of V_H and V_h . Note also that moving from the situation depicted in Figure 1.2(a) to the one in Figure 1.2(b), the dimension of V_{Hh} increases by 1. For a more detailed discussion on these issues, we refer the reader to Section 2.1.

All these difficulties impose an iterative method for solving problem (1.6). We introduce the algorithm in Section 1.3 and analyze it through Sections 1.4–1.6.

1.2 A priori and a posteriori error estimates

Before showing how to compute u_{Hh} (Section 1.3), we investigate the convergence orders through an *a priori* estimate (Proposition 1.1), and the local quality of the solution by introducing an *a posteriori* error estimator (Proposition 1.2).

We use the norm $\|\cdot\|$ based on the scalar product a introduced in (1.1) and equivalent in $H_0^1(\Omega)$ to the usual $H^1(\Omega)$ -norm. Recall that $H = \max_{K \in \mathcal{T}_H} \text{diam}(K)$ and $h = \max_{K \in \mathcal{T}_h} \text{diam}(K)$.

Proposition 1.1 (A priori error estimate, see [36]). *Let $q = \max(r, s) + 1$ and suppose that the solution u of (1.2) is in $H^q(\Omega)$. Then the approximation u_{Hh} to u satisfies the a priori error estimate*

$$\|u - u_{Hh}\| \leq C \left(H^r \|u\|_{H^q(\Omega \setminus \bar{\Lambda})} + h^s \|u\|_{H^q(\Lambda)} \right), \quad (1.9)$$

where C is a constant independent of H , h and u .

For reader's convenience, we establish the proof below:

Proof. The boundary $\partial\Lambda$ being locally Lipschitz, due to the Stein Extension Theorem (see Adams and Fournier [5], Thm. 5.24), there exists a bounded extension operator $E : H^q(\Omega \setminus \bar{\Lambda}) \rightarrow H^q(\Omega)$, i.e. $Ev|_{\Omega \setminus \bar{\Lambda}} = v|_{\Omega \setminus \bar{\Lambda}}$, $\forall v \in H^q(\Omega \setminus \bar{\Lambda})$. Let u be the solution of (1.2). We define \tilde{u} the extension of $u|_{\Omega \setminus \bar{\Lambda}}$ to Ω such that $\tilde{u} = Eu$ if $\|Eu\|_{H^q(\Lambda)} \leq \|u\|_{H^q(\Lambda)}$ and $\tilde{u} = u$ otherwise. We have that $\tilde{u} = u$ in $\Omega \setminus \bar{\Lambda}$,

$$\|\tilde{u}\|_{H^q(\Omega)} \leq C \|u\|_{H^q(\Omega \setminus \bar{\Lambda})}, \quad (1.10)$$

where here, like in the sequel, C denotes a generic constant, and

$$\|\tilde{u}\|_{H^q(\Lambda)} \leq \|u\|_{H^q(\Lambda)}. \quad (1.11)$$

Note that $u - \tilde{u} \in H_0^q(\Lambda)$. Let r_H and r_h be the standard interpolants to the spaces V_H and V_h respectively. We introduce $\tilde{u}_H = r_H \tilde{u}$ and $\tilde{u}_h = r_h(u - \tilde{u})$. Define $\tilde{u}_{Hh} = \tilde{u}_H + \tilde{u}_h$ and $v_{Hh} = u_{Hh} - \tilde{u}_{Hh}$. By the definitions of u and u_{Hh} we have $a(u, v_{Hh}) = a(u_{Hh}, v_{Hh})$. This and the previous definitions lead to the equality $a(v_{Hh}, v_{Hh}) = a(u - \tilde{u}_{Hh}, v_{Hh})$, from which we derive using the Cauchy-Schwarz inequality that $\|v_{Hh}\|^2 \leq \|u - \tilde{u}_{Hh}\| \|v_{Hh}\|$. Thus

$$\|u_{Hh} - \tilde{u}_{Hh}\| \leq \|u - \tilde{u}_{Hh}\|. \quad (1.12)$$

With $u - u_{Hh} = (u - \tilde{u}_{Hh}) + (\tilde{u}_{Hh} - u_{Hh})$ and (1.12), we have

$$\|u - u_{Hh}\| \leq \|u - \tilde{u}_{Hh}\| + \|u_{Hh} - \tilde{u}_{Hh}\| \leq 2\|u - \tilde{u}_{Hh}\|. \quad (1.13)$$

Writing $u - \tilde{u}_{Hh} = (\tilde{u} - \tilde{u}_H) + [(u - \tilde{u}) - \tilde{u}_h]$, we get by standard interpolation results

$$\|u - \tilde{u}_{Hh}\| \leq \|\tilde{u} - \tilde{u}_H\| + \|(u - \tilde{u}) - \tilde{u}_h\| \quad (1.14)$$

$$\leq C \left(H^r \|\tilde{u}\|_{H^q(\Omega)} + h^s \|u - \tilde{u}\|_{H^q(\Lambda)} \right), \quad (1.15)$$

and furthermore, with $\|u - \tilde{u}\|_{H^q(\Lambda)} \leq \|u\|_{H^q(\Lambda)} + \|\tilde{u}\|_{H^q(\Lambda)}$ and using the relations (1.10) and (1.11), we obtain

$$\|u - \tilde{u}_{Hh}\| \leq C \left(H^r \|u\|_{H^q(\Omega \setminus \bar{\Lambda})} + h^s \|u\|_{H^q(\Lambda)} \right). \quad (1.16)$$

Hence, combining the results (1.13) and (1.16) completes the proof. \square

Let us outline an illustration of the *a priori* error estimate (1.9). For doing this, we consider a two-dimensional Poisson problem with sharp right-hand side and suppose that we have computed an approximation $u_{Hh} \in V_{Hh}$ to the solution u .

More precisely, consider the problem

$$\begin{cases} -\Delta u = f & \text{in } \Omega = (-1; 1)^2, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \quad (1.17)$$

Set $f = f_0 + f_1$, where $f_0 = -\Delta u_0$ and $f_1 = -\Delta u_1$, such that the exact solution to the problem is given by $u = u_0 + u_1$. Take $u_0(x, y) = \cos(\frac{\pi}{2}x) \cos(\frac{\pi}{2}y)$ and $u_1(x, y) = \eta \chi(R) \exp \epsilon_f^{-2} \exp(-1/|\epsilon_f^2 - R^2|)$, where $R(x, y) = \sqrt{x^2 + y^2}$ and $\chi(R) = 1$ if $R \leq \epsilon_f$, $\chi(R) = 0$ if $R > \epsilon_f$; η and ϵ_f are parameters. We choose $\eta = 10$ and $\epsilon_f = 0.4$.

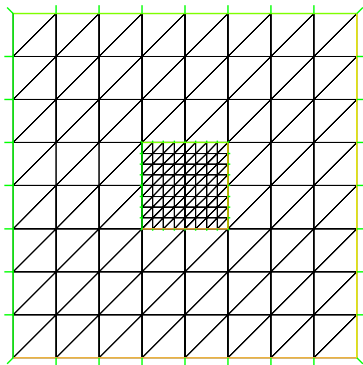
Away from the origin $(0, 0)$ the solution is smooth and varies slowly. In a region close to $(0, 0)$ the solution has a peak, we want to apply a patch for more precision.

For the triangulation \mathcal{T}_H of $\bar{\Omega}$, we use a coarse structured, respectively unstructured, grid and linear finite elements for V_H ($r = 1$). We consider the patch $\Lambda = (-0.25; 0.25)^2$ with a fine structured triangulation and set $s = 1$. We set $H = 2/N$ and $h = 0.5/M$, N, M being the number of intervals on one side of the squares Ω and Λ respectively. An illustration of the grid constellations with $N = 8$ and $H/h = 4$ is given in Figures 1.3(a) and 1.4(a) resp. for structured and unstructured \mathcal{T}_H , nested and non-nested \mathcal{T}_h .

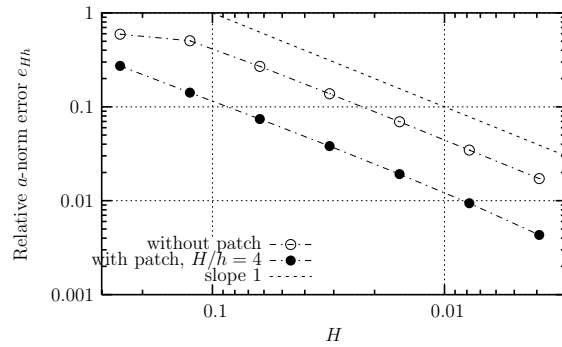
We compute the approximation u_H on the coarse grid (without patch) and evaluate u_{Hh} using the method defined in Section 1.3. We follow the implementation details specified in Sections 2.2 and 2.3. For the evaluation of the errors, we consider the associated a -norm to the problem (1.17), induced by the scalar-product

$$a(\psi, \varphi) = \int_{\Omega} \nabla \psi \cdot \nabla \varphi \, dx, \quad \forall \psi, \varphi \in V_{Hh}. \quad (1.18)$$

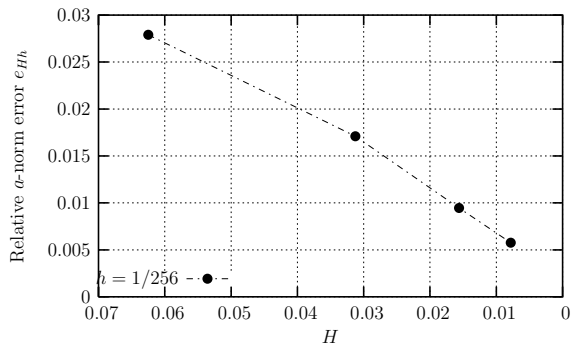
We evaluate numerically the integral terms appearing in the iterative method following the discussion in Section 2.3, and more precisely using the formulas (2.31) and (2.32).



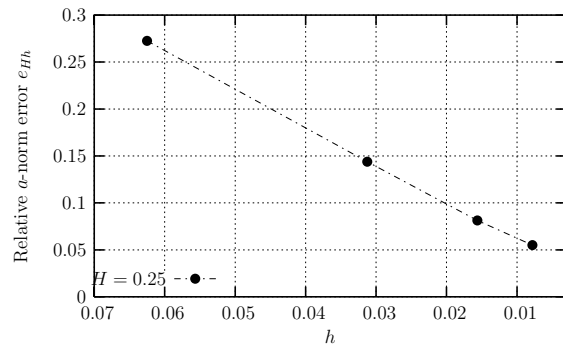
(a) Structured coarse \mathcal{T}_H and nested fine \mathcal{T}_h .



(b) Convergence of u_{Hh} to u in the mesh size H with $H/h = 4$ for the triangulations in (a).

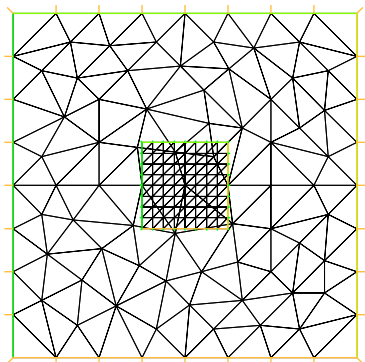


(c) Convergence of u_{Hh} to u in the mesh size H with h fixed for the triangulations in (a).

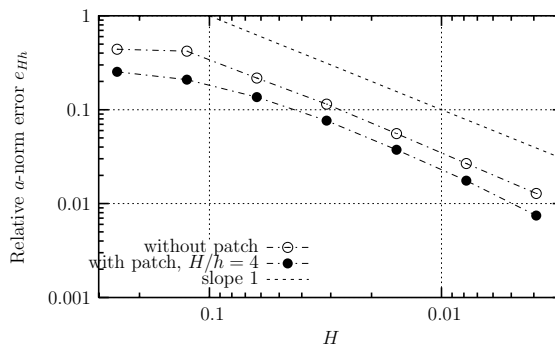


(d) Convergence of u_{Hh} to u in the mesh size h with H fixed for the triangulations in (a).

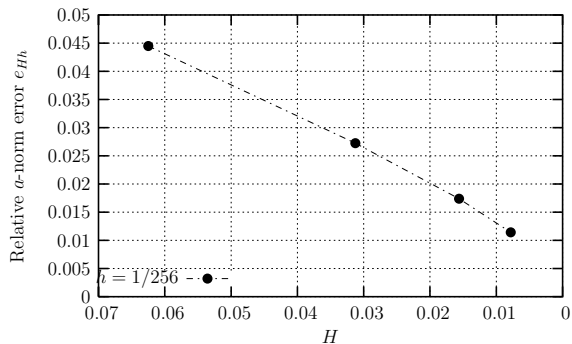
Figure 1.3: Illustration of the *a priori* error estimate for linear finite elements on nested grids.



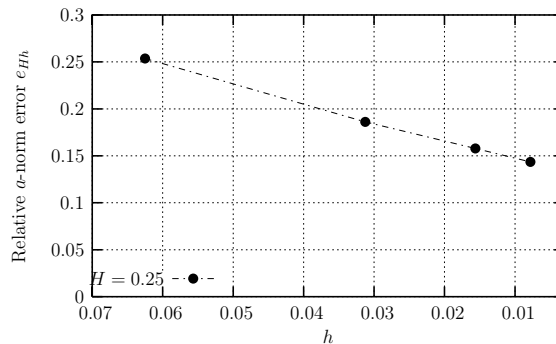
(a) Unstructured coarse \mathcal{T}_H and structured fine \mathcal{T}_h .



(b) Convergence order in the mesh size of u_{Hh} to u for the triangulations in (a).



(c) Convergence of u_{Hh} to u in the mesh size H with h fixed for the triangulations in (a).



(d) Convergence of u_{Hh} to u in the mesh size h with H fixed for the triangulations in (a).

Figure 1.4: Illustration of the *a priori* error estimate for linear finite elements on non-nested grids.

We are now able to report the relative error of u_H and u_{Hh} to u for increasing N , $N = 2^3, \dots, 2^9$. First we take $H/h = 4$ fixed, i.e. $M = N$. The results in Figures 1.3(b) and 1.4(b) exemplify the *a priori* error estimate of Proposition 1.1 in the nested case resp. the case of an unstructured triangulation \mathcal{T}_H . In both cases, we observe optimal convergence (order one) in the mesh size H , i.e. $O(H)$ -accuracy for the a -norm which is equivalent to the H^1 -norm. Furthermore, we observe that the error when using a patch is smaller in comparison to the case without a patch. Finally, through Figures 1.3 and 1.4, (c) and (d), we confirm Proposition 1.1. In the Figures 1.3(c) and 1.4(c) we show the linear convergence in H of u_{Hh} to u when keeping h fixed. In the Figures 1.3(d) and 1.4(d) we illustrate the *a priori* convergence result with H fixed.

In practical problems, due to the nature of the data in certain regions, it happens that

the solution of a boundary-value problem is less regular. As discussed earlier, we would like to increase the accuracy of the finite element approximation by applying patches with refined grids in those sub-domains where it is needed. That is why we carry out finite element calculations either on a provisional coarse grid, or on a given set of a grid with patches placed by some a priori knowledge. Then we can compute an *a posteriori* estimate for the error. The purpose is to indicate what part of which grid induces large errors. Using this information, we apply a patch, and repeat the finite element computation.

To simplify our discussion, we restrict ourselves in the sequel of this section to the case of the Laplace operator $\mathcal{L} = -\Delta$ in two dimensions ($d = 2$), i.e. the Poisson equation

$$-\Delta u = f \quad \text{in } \Omega \subset \mathbb{R}^2, \quad (1.19)$$

with homogeneous Dirichlet boundary conditions. We consider its finite element approximation in the setting introduced earlier. We solve the problem to find $u_{Hh} \in V_{Hh} \subset H_0^1(\Omega)$ satisfying

$$a(u_{Hh}, \varphi) = \langle f | \varphi \rangle, \quad \forall \varphi \in V_{Hh}, \quad (1.20)$$

where $a(\psi, \varphi) = \int_{\Omega} \nabla \psi \cdot \nabla \varphi \, dx$. We suppose that the boundary $\partial\Omega$ is conforming with edges of triangles of \mathcal{T}_h . Our objective is to give an estimate based on the triangulations \mathcal{T}_H of Ω and \mathcal{T}_h of Λ .

Following [22, §8] and [54], our estimate uses the approach of residual estimators first introduced in [13]. When we insert the finite element solution u_{Hh} of (1.20) in the differential equation in its classical form (1.19), we get a residual. Moreover the approximation differs from the solution in that its gradient has jumps on the edges of the elements of the triangulation. The error is bounded on an element K in terms of the size of the area-based residual and the edge-based jumps on all inter-element boundaries, i.e. edges of the triangles which lie in the interior of Ω resp. Λ .

For any triangle K of \mathcal{T}_H or \mathcal{T}_h with boundary ∂K , diameter h_K and edges of length $h_{l,K}$, $l = 1, 2, 3$, we define the local quantity

$$\eta(K, u) = h_K \|\Delta u + f\|_{L^2(K)} + \sum_{l=1}^3 \sqrt{h_{l,K}} \left\| \left[\frac{\partial u}{\partial n} \right] \right\|_{L^2(\partial K)}, \quad (1.21)$$

where u is a polynomial and $\left[\frac{\partial u}{\partial n} \right]$ denotes the jump of the normal derivative of u on ∂K when we have fixed a normal direction n on each internal side of the triangulation. With this notation, we set $-\frac{1}{2} \left[\frac{\partial u}{\partial n} \right] = \frac{\partial u}{\partial n}$ on the sides of triangles which are on the boundaries $\partial\Omega$ of Ω and $\partial\Lambda$ of Λ . We introduce $\hat{\mathcal{T}}_H$, the restriction of \mathcal{T}_H to $\Omega \setminus \bar{\Lambda}$.

If r_H and r_h are the standard interpolants to the spaces V_H and V_h respectively, we introduce a decomposition of $u_{Hh} \in V_{Hh}$: we define $\tilde{u}_H = r_H u_{Hh} \in V_H$ and $\tilde{u}_h = u_{Hh} - \tilde{u}_H \in V_h$ such that $u_{Hh} = \tilde{u}_H + \tilde{u}_h$.

Proposition 1.2 (A posteriori error estimate). *Let $\mathcal{L} = -\Delta$ and consider the boundary $\partial\Omega$ conforming with edges of triangles of \mathcal{T}_h . Then the approximation u_{Hh} to u satisfies the a posteriori error estimate*

$$\|u - u_{Hh}\| \leq C \left(\sqrt{\sum_{K \in \hat{\mathcal{T}}_H} \eta(K, \tilde{u}_H)^2} + \sqrt{\sum_{K \in \mathcal{T}_h} \eta(K, u_{Hh})^2} \right), \quad (1.22)$$

where C is a constant independent of H and h , and η is the local estimator defined by (1.21).

Proof. We prove (1.22) in evaluating the right-hand side of

$$\|u - u_{Hh}\| \leq C \sup_{\substack{v \in H_0^1(\Omega) \\ \|v\|=1}} \int_{\Omega} \nabla(u - u_{Hh}) \cdot \nabla v \, dx. \quad (1.23)$$

Consider the expression $I(v) = \int_{\Omega} \nabla(u - u_{Hh}) \cdot \nabla v \, dx = \int_{\Omega} (fv - \nabla u_{Hh} \cdot \nabla v) \, dx$ for $v \in H_0^1(\Omega)$, $\|v\| = 1$. With $\varphi = v_{Hh} \in V_{Hh}$ in (1.20), the decomposition $\Omega = (\Omega \setminus \bar{\Lambda}) \cup \Lambda$ and $u_{Hh} = \tilde{u}_H + \tilde{u}_h$ yield $\forall v_{Hh} \in V_{Hh}$

$$\begin{aligned} I(v) &= \int_{\Omega \setminus \bar{\Lambda}} (f(v - v_{Hh}) - \nabla \tilde{u}_H \cdot \nabla(v - v_{Hh})) \, dx \\ &\quad + \int_{\Lambda} (f(v - v_{Hh}) - \nabla u_{Hh} \cdot \nabla(v - v_{Hh})) \, dx \\ &= \sum_{K \in \hat{\mathcal{T}}_H} \left(\int_K (f + \Delta \tilde{u}_H)(v - v_{Hh}) \, dx - \int_{\partial K} \frac{\partial \tilde{u}_H}{\partial n} (v - v_{Hh}) \, ds \right) \\ &\quad + \sum_{K \in \mathcal{T}_h} \left(\int_K (f + \Delta u_{Hh})(v - v_{Hh}) \, dx - \int_{\partial K} \frac{\partial u_{Hh}}{\partial n} (v - v_{Hh}) \, ds \right). \end{aligned} \quad (1.24)$$

$$(1.25)$$

Hence

$$\begin{aligned} I(v) &\leq \sum_{K \in \hat{\mathcal{T}}_H} \left(\left| \int_K (f + \Delta \tilde{u}_H)(v - v_{Hh}) \, dx \right| + \frac{1}{2} \left| \int_{\partial K} \left[\frac{\partial \tilde{u}_H}{\partial n} \right] (v - v_{Hh}) \, ds \right| \right) \\ &\quad + \sum_{K \in \mathcal{T}_h} \left(\left| \int_K (f + \Delta u_{Hh})(v - v_{Hh}) \, dx \right| + \frac{1}{2} \left| \int_{\partial K} \left[\frac{\partial u_{Hh}}{\partial n} \right] (v - v_{Hh}) \, ds \right| \right) \end{aligned} \quad (1.26)$$

and for $v \in H_0^1(\Omega)$, $\|v\| = 1$,

$$\begin{aligned}
 I(v) \leq & \sup_{\substack{v \in H_0^1(\Omega) \\ \|v\|=1}} \inf_{v_{Hh} \in V_H} \left(\sum_{K \in \hat{\mathcal{T}}_H} \|\Delta \tilde{u}_H + f\|_{L^2(K)} \|v - v_{Hh}\|_{L^2(K)} \right. \\
 & + \frac{1}{2} \left\| \left[\frac{\partial \tilde{u}_H}{\partial n} \right] \right\|_{L^2(\partial K)} \|v - v_{Hh}\|_{L^2(\partial K)} \\
 & + \sum_{K \in \mathcal{T}_h} \|\Delta u_{Hh} + f\|_{L^2(K)} \|v - v_{Hh}\|_{L^2(K)} \\
 & \left. + \frac{1}{2} \left\| \left[\frac{\partial u_{Hh}}{\partial n} \right] \right\|_{L^2(\partial K)} \|v - v_{Hh}\|_{L^2(\partial K)} \right). \quad (1.27)
 \end{aligned}$$

At this point, we recall Clément's results on approximation [30]. Choose v_{Hh} to be Clément's interpolant of v . For given $v \in H_0^1(\Omega)$ we have the properties

$$\|v - v_{Hh}\|_{L^2(K)} \leq Ch_K \|v\|_{H^1(\Lambda_K)}, \quad \forall K \in \hat{\mathcal{T}}_H \text{ or } \mathcal{T}_h, \quad (1.28)$$

and

$$\|v - v_{Hh}\|_{L^2(l)} \leq C \sqrt{h_{k,l}} \|v\|_{H^1(\Lambda_K)}, \quad \forall K \in \hat{\mathcal{T}}_H \text{ or } \mathcal{T}_h, \forall l, \text{ edge of } K, l \in \hat{\Omega} \text{ or } \hat{\Lambda}, \quad (1.29)$$

where $\Lambda_K = \{T \in \mathcal{T} : T \cap K \neq \emptyset\}$ for $K \in \mathcal{T}$, \mathcal{T} being a triangulation.

The result (1.22) now follows from (1.27) with properties (1.28), (1.29), applying the Cauchy-Schwarz inequality and introducing the local quantity η defined by (1.21). \square

1.3 The algorithm

The idea is to solve the problem (1.6) on a domain Ω and consider therein a patch Λ wherein we would like to obtain more accuracy. Thus we calculate successively corrections to the solution in the patch.

We start from an initial approximation evaluated on a coarse triangulation over all the domain Ω . One step of our algorithm consist in two parts: first we calculate a correction on a fine triangulation in the patch Λ which we add to the initial solution to obtain an intermediate solution. Next, using this last update we evaluate a correction over the whole domain Ω on the coarse triangulation leading to an overall update. We also introduce a relaxation parameter ω . Hence we add ω times the correction at each step to update the solution.

In the following we define the algorithm and introduce its iteration operator (1.37), the key for the convergence analysis. Then we discuss the algorithm by comparing it to

existing methods.

Recall the notation from Section 1.1. An approximation of u by the finite element method of order r consists in using a regular triangulation \mathcal{T}_H of $\bar{\Omega}$ and the space V_H , see (1.7), and calculating $u_H \in V_H$ satisfying

$$a(u_H, \varphi) = \langle f | \varphi \rangle, \quad \forall \varphi \in V_H. \quad (1.30)$$

We consider a patch $\Lambda \subset \Omega$ wherein we would like to obtain a better precision on the solution u than the one given by u_H . We use \mathcal{T}_h a regular triangulation of $\bar{\Lambda}$ and consider V_h given by (1.8). Setting $V_{Hh} = V_H + V_h$ we search as approximation for u the function $u_{Hh} \in V_{Hh}$ satisfying (1.6), i.e.

$$a(u_{Hh}, \varphi) = \langle f | \varphi \rangle, \quad \forall \varphi \in V_{Hh}. \quad (1.31)$$

As we have mentioned at the end of Section 1.1, a priori $V_H \cap V_h$ does not necessarily reduce to the element zero and it is impossible, practically speaking, to exhibit a finite element basis of the space V_{Hh} and consequently to compute directly u_{Hh} . It is the reason for which we suggest the following algorithm for computing u_{Hh} .

Algorithm 1.3.

- Initialization: Set $u^0 = u_H \in V_H$ satisfying (1.30) and choose $\omega \in (0; 2)$.
- For $n = 1, 2, 3, \dots$

(i) find $w_h \in V_h$ such that

$$a(w_h, \varphi) = \langle f | \varphi \rangle - a(u^{n-1}, \varphi), \quad \forall \varphi \in V_h; \quad (1.32)$$

define

$$u^{n-\frac{1}{2}} = u^{n-1} + \omega w_h; \quad (1.33)$$

(ii) find $w_H \in V_H$ such that

$$a(w_H, \varphi) = \langle f | \varphi \rangle - a(u^{n-\frac{1}{2}}, \varphi), \quad \forall \varphi \in V_H; \quad (1.34)$$

define

$$u^n = u^{n-\frac{1}{2}} + \omega w_H. \quad (1.35)$$

When implementing the algorithm, the coarse and the fine parts of u^n and $u^{n-\frac{1}{2}}$ are stored separately. This and issues related to the integration are discussed in Section 2.3.

For analyzing Algorithm 1.3 we need to introduce some notation.

If $P_h : V_{Hh} \rightarrow V_h$ and $P_H : V_{Hh} \rightarrow V_H$ are orthogonal projectors from V_{Hh} onto V_h and V_H respectively with regard to the scalar product $a(\cdot, \cdot)$, we have

$$u_{Hh} - u^n = B(u_{Hh} - u^{n-1}), \quad (1.36)$$

where B is the iteration operator given by

$$B = (I - \omega P_H)(I - \omega P_h), \quad (1.37)$$

with I denoting the identity operator in V_{Hh} . Hence

$$u_{Hh} - u^n = B^n(u_{Hh} - u^0). \quad (1.38)$$

It is readily seen that this algorithm introduced in [36, 37] is a Schwarz type domain decomposition method [58] without any conformity between the meshes \mathcal{T}_H and \mathcal{T}_h (see for instance the work by Chan et al. [27]). In the multi-scale situation depicted on Figure 1.1 we have a complete overlapping.

It is similar to the Chimera, or overset grid method, investigated by Steger et al. [60] and Brezzi et al. [25]. However in its original formulation the latter is an additive method: in fact, from [25, eqn. 2], we derive that the iteration operator for the Chimera method is $I - P_h - P_H$. The difference, when $\omega = 1$, stems from (1.34) where we consider in the residual the intermediate solution $u^{n-\frac{1}{2}}$ and not u^{n-1} . The Chimera method can easily be rewritten to be multiplicative: it suffices to use in the second line of [25, eqn. 2] the updated u^{n+1} instead of u^n . This change makes it a multiplicative method equivalent to Algorithm 1.3 with $\omega = 1$.

Our multiplicative Schwarz method is also similar to a Gauss-Seidel method and can be put in the framework of the successive subspace correction method studied by Xu [64, 65, 66] and Xu and Zikatanov [67].

The spaces V_H and V_h defined on the arbitrary triangulations \mathcal{T}_H and \mathcal{T}_h are not necessary orthogonal nor share the only element zero as intersection. Note in particular that the sum which defines V_{Hh} is a priori not a direct sum. This property makes the above algorithm different from most iterative schemes (see for example the scheme by Laydi [46]).

For structured grid constellations, the algorithm resembles the Fast Adaptive Composite (FAC) grid method, see for example the works from McCormick et al. [49, 50, 51], and after Lee et al. [12]. It also resembles possibly a hierarchical method (see for example the papers from Yserentant [69, 70], Bank et al. [14] and Bank and Smith [15]) with a mortar method (see [4]).

We underline that the new aspect we introduce is to link the speed of convergence of Algorithm 1.3 to the parameter $\tilde{\gamma}$, introduced here below in (1.107), corresponding to the cosine of an abstract angle between the spaces V_h and V_H . Furthermore, an optimal relaxation through the choice of the parameter ω keeps the method competitive in cases where the problem is badly conditioned (see Section 2.2 and in particular Table 2.3(c)).

1.4 Some properties of vector spaces

As discussed earlier in the Introduction, one objective of this work is to analyze the convergence behavior of the iterative method (Algorithm 1.3) introduced in the previous section. The algorithm is described by an iteration operator B (1.37) on the discretization spaces V_H (1.7) and V_h (1.8). The objective of this section is to prepare in an abstract setting the analysis of the spectral properties of the operator B (see Section 1.5).

Generally, let V be a Hilbert space with scalar product (\cdot, \cdot) and denote by $\|\cdot\|$ the induced norm. Consider V_1, V_2 two closed subspaces of V . This section contains two main ideas: On one hand, we introduce the quantities which describe most completely the relation between the spaces V_1 and V_2 . As projection operators appear in the definition of the iteration operator it is important to find the right quantities to describe the “angle” between the two spaces. On the other hand, we consider the case where V is of finite dimension and $V_1 + V_2 = V$, i.e. the two subspaces span the whole space. This sum being not necessarily a direct sum, we decompose V (Proposition 1.7) in terms of summands, mutually orthogonal subspaces of V and invariant with respect to the orthogonal projection operators from V onto V_1 and V_2 .

We introduce the number

$$\gamma = \begin{cases} \sup_{\substack{v_1 \in V_1, v_1 \neq 0 \\ v_2 \in V_2, v_2 \neq 0}} \frac{(v_1, v_2)}{\|v_1\| \|v_2\|}, & \text{if } V_1 \neq \{0\} \text{ and } V_2 \neq \{0\}, \\ 0, & \text{otherwise,} \end{cases} \quad (1.39)$$

which is the optimal constant for the corresponding strengthened Cauchy-Buniakowski-Schwarz (C.B.S.) inequality

$$(v_1, v_2) \leq C \|v_1\| \|v_2\|, \quad \forall v_1 \in V_1, \forall v_2 \in V_2. \quad (1.40)$$

The constant γ is the cosine of the abstract angle between the two subspaces V_1 and V_2 if $V_1 \cap V_2 = \{0\}$. From the definition (1.39), we have the following obvious properties for γ .

Properties 1.4.

- (i) Constant γ is necessarily included in the interval $[0; 1]$.
- (ii) If $V_1 \cap V_2 \neq \{0\}$, then we have $\gamma = 1$.
- (iii) Constant $\gamma = 0$ if and only if V_1 is orthogonal to V_2 .

We set $V_0 = V_1 \cap V_2$ and V_0^\perp the orthogonal complement of V_0 in V . The above property (ii) implies that when $V_0 \neq \{0\}$, the parameter γ is not very informative as it remains equal to one for a large set of spaces V_1 and V_2 .

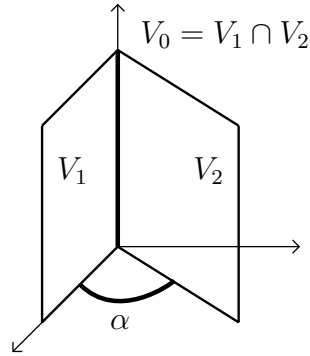


Figure 1.5: Illustration for γ and $\tilde{\gamma}$ in a case with 2D spaces V_1 and V_2 and $V_0 \neq \{0\}$: the parameter $\gamma = 1$, while $\tilde{\gamma}$ corresponds to the cosine of the angle α between V_1 and V_2 .

This suggests to introduce the number

$$\tilde{\gamma} = \begin{cases} \sup_{\substack{v_1 \in V_1 \cap V_0^\perp, v_1 \neq 0 \\ v_2 \in V_2 \cap V_0^\perp, v_2 \neq 0}} \frac{(v_1, v_2)}{\|v_1\| \|v_2\|}, & \text{if } V_1 \neq V_0 \text{ and } V_2 \neq V_0, \\ 0, & \text{otherwise.} \end{cases} \quad (1.41)$$

Figure 1.5 shows in a particular case that if $V_0 \neq \{0\}$, while γ equals 1, the parameter $\tilde{\gamma}$ corresponds to the cosine of the abstract angle of the two subspaces V_1 and V_2 .

In the next section we recall (Proposition 1.8) and discuss existing work based on [24] for the analysis of the iteration operator. To our knowledge all theories assume that the following hypothesis is satisfied:

Hypothesis 1.5. There exists a constant C_0 such that for all $v \in V$ there exist $v_1 \in V_1$, $v_2 \in V_2$ satisfying

$$v = v_1 + v_2, \quad (1.42)$$

and

$$\|v_1\|^2 + \|v_2\|^2 \leq C_0^2 \|v\|^2. \quad (1.43)$$

Let us observe that:

1. If Hypothesis (1.5) is satisfied, we have necessarily $V = V_1 + V_2$.
2. If $V_1 \neq V_2$, we have necessarily $C_0 \geq 1$.
3. In the case $V_1 = V_2 = V$ the optimal constant C_0 in (1.43) is equal to $1/\sqrt{2}$ (it suffices to take $v_1 = v_2 = \frac{1}{2}v$, $v \in V$).
4. If V_1 is orthogonal to V_2 , we can take $C_0 = 1$ from Pythagore's Theorem.

As the constant C_0 of (1.43) appears in unsharp results of earlier works (see Proposition 1.8) it is useful for comparison to inquire about the optimal constant that can be chosen. The answer is given by the following Proposition 1.6 which gives a simple condition for Hypothesis 1.5 to be satisfied and links the constant C_0 to the above introduced number $\tilde{\gamma}$.

Proposition 1.6 (see [36]). *If $V = V_1 + V_2$ then Hypothesis 1.5 is satisfied and $\tilde{\gamma} < 1$. If, moreover, $V_1 \neq V_2$ then*

$$C_0^{opt} = \sqrt{\frac{1}{1 - \tilde{\gamma}}}, \quad (1.44)$$

is the optimal constant in (1.43).

For the convenience of the reader we give here the proof.

Proof. Let us denote $\tilde{V}_j = V_j \cap V_0^\perp$, $j = 1, 2$, then $V_0^\perp = \tilde{V}_1 \oplus \tilde{V}_2$ and $V = V_0 \oplus \tilde{V}_1 \oplus \tilde{V}_2$. The Corollary of the Open Mapping Theorem (see Yosida [68, §II.5]) for the one-to-one mapping $(\tilde{v}_1, \tilde{v}_2) \in \tilde{V}_1 \times \tilde{V}_2 \rightarrow \tilde{v}_1 + \tilde{v}_2 \in V_0^\perp$ yields the existence of $\tilde{C}_0 < +\infty$ such that $\forall \tilde{v}_j \in \tilde{V}_j$, $j = 1, 2$, we have $\|\tilde{v}_1\|^2 + \|\tilde{v}_2\|^2 \leq \tilde{C}_0^2 \|\tilde{v}_1 + \tilde{v}_2\|^2$. We can take $\tilde{C}_0 \geq 1$.

For all $v \in V$ we have a unique decomposition

$$v = v_0 + \tilde{v}_1 + \tilde{v}_2 \text{ with } v_0 \in V_0, \tilde{v}_j \in \tilde{V}_j, j = 1, 2. \quad (1.45)$$

Hence, we can put

$$v_1 = v_0 + \tilde{v}_1 \in V_1 \text{ and } v_2 = \tilde{v}_2 \in V_2, \quad (1.46)$$

so that $v = v_1 + v_2$ and

$$\|v_1\|^2 + \|v_2\|^2 = \|v_0\|^2 + \|\tilde{v}_1\|^2 + \|\tilde{v}_2\|^2 \quad (1.47)$$

$$\leq \tilde{C}_0^2 (\|v_0\|^2 + \|\tilde{v}_1 + \tilde{v}_2\|^2) = \tilde{C}_0^2 \|v\|^2, \quad (1.48)$$

i.e. Hypothesis 1.5 is satisfied with $C_0 = \tilde{C}_0$.

Let us now consider the case $V_1 \neq V_0$ and $V_2 \neq V_0$. Using Definition (1.41), there exists a sequence $v^m = \tilde{v}_1^m + \tilde{v}_2^m$ with $\tilde{v}_1^m \in \tilde{V}_1$, $\tilde{v}_2^m \in \tilde{V}_2$ and $\|\tilde{v}_1^m\| = \|\tilde{v}_2^m\| = 1$ such that

$$(\tilde{v}_1^m, \tilde{v}_2^m) \rightarrow -\tilde{\gamma}. \quad (1.49)$$

Suppose *ad absurdum* that $\tilde{\gamma} = 1$. Thus

$$\frac{\|\tilde{v}_1^m\|^2 + \|\tilde{v}_2^m\|^2}{\|v^m\|^2} = \frac{1}{1 + (\tilde{v}_1^m, \tilde{v}_2^m)} \rightarrow +\infty, \quad (1.50)$$

which contradicts Hypothesis 1.5. Hence $\tilde{\gamma} < 1$.

Using again the decomposition (1.45) for any $v \in V$ and setting $v_1 \in V_1$, $v_2 \in V_2$ as in (1.46), we have $\|v_1\|^2 + \|v_2\|^2 \leq \|v_0\|^2 + \|\tilde{v}_1 + \tilde{v}_2\|^2 + 2|(\tilde{v}_1, \tilde{v}_2)| \leq \|v\|^2 + 2\tilde{\gamma}\|\tilde{v}_1\|\|\tilde{v}_2\|$. Since $2\|\tilde{v}_1\|\|\tilde{v}_2\| \leq \|\tilde{v}_1\|^2 + \|\tilde{v}_2\|^2 \leq \|v_1\|^2 + \|v_2\|^2$, we get

$$\|v_1\|^2 + \|v_2\|^2 \leq \frac{1}{1 - \tilde{\gamma}} \|v\|^2. \quad (1.51)$$

Thus we can choose $C_0 = \sqrt{\frac{1}{1-\tilde{\gamma}}}$ in (1.43). It suffices to use (1.49) in (1.41) to show that $\sqrt{\frac{1}{1-\tilde{\gamma}}}$ is the best constant we can choose.

In the case $V_1 = V_0$ or $V_2 = V_0$ we have that $\tilde{\gamma} = 0$ and if, moreover, $V_1 \neq V_2$, then $C_0^{\text{opt}} = 1$, i.e. (1.44) is also valid. \square

In order to perform the analysis of the iteration operator in cases where $V = V_1 + V_2$ is not a direct sum, we need to decompose V in terms of direct summands. This is the objective of the following Proposition 1.7. We introduce $P_j : V \rightarrow V_j \subset V$ the orthogonal projectors from V onto V_j , $j = 1, 2$, and call V_j^\perp the orthogonal complement of V_j in V .

Proposition 1.7 ([36]). *Let V be of finite dimension and $V = V_1 + V_2$. There exist $2p$ ($p \geq 0$) vectors $v_1^{(m)} \in V_1$ and $v_2^{(m)} \in V_2$, $m = 1, \dots, p$, such that*

$$\|v_1^{(m)}\| = \|v_2^{(m)}\| = 1, \quad (v_1^{(m)}, v_2^{(m)}) = \gamma_m, \quad m = 1, \dots, p, \quad (1.52)$$

with

$$1 > \gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_p > 0, \quad (1.53)$$

and V can be decomposed into the direct sum

$$V = (V_1 \cap V_2) \oplus (V_1^\perp \cap V_2) \oplus (V_1 \cap V_2^\perp) \oplus L_1 \oplus \dots \oplus L_p, \quad (1.54)$$

where $L_m = \text{span}\{v_1^{(m)}, v_2^{(m)}\}$, $m = 1, \dots, p$, and all the summands in (1.54) are mutually orthogonal subspaces of V , which are invariant with respect to both operators P_1 and P_2 , i.e. $P_j L_m \subset L_m$, $j = 1, 2$.

For reader convenience we repeat the proof from [36].

Proof. Let us prove that for any integer k , $0 \leq k \leq p$ with p to be identified later in the proof, the space V can be decomposed into a direct sum with mutually orthogonal summands

$$V = V_0 \oplus W_k \oplus L_1 \oplus \dots \oplus L_k, \quad (1.55)$$

where $V_0 = V_1 \cap V_2$, the spaces L_m are the two-dimensional subspaces of V appearing in (1.54) and all the subspaces V_0 and $L_1, \dots, L_k, W_k \subset V_0^\perp$ are invariant with respect to both operators P_1 and P_2 . The decomposition (1.55) will be constructed by induction on k .

We start with $k = 0$ and set $W_0 = V_0^\perp$. Note that V_0 and W_0 are invariant subspaces of operators P_1 and P_2 . On the k -th step of our construction ($k \geq 1$) we suppose that (1.55) is established for $k - 1$. Let $V_1^{(k)} = V_1 \cap W_{k-1}$, $V_2^{(k)} = V_2 \cap W_{k-1}$ and define

$$\gamma_k = \begin{cases} \max_{\substack{v_1 \in V_1^{(k)}, v_2 \in V_2^{(k)} \\ \|v_1\| = \|v_2\| = 1}} (v_1, v_2), & \text{if } V_1^{(k)} \neq \{0\} \text{ and } V_2^{(k)} \neq \{0\}, \\ 0, & \text{otherwise.} \end{cases} \quad (1.56)$$

If $\gamma_k = 0$ we stop the induction and set $p = k - 1$. Indeed, it is easy to see that in this case, any vector from $V_1^{(k)}$ is orthogonal to V_2 and any vector from $V_2^{(k)}$ is orthogonal to V_1 , i.e.

$$W_{k-1} \subseteq (V_1^\perp \cap V_2) \oplus (V_1 \cap V_2^\perp), \quad (1.57)$$

which gives in combination with (1.55) the desired decomposition (1.54).

Assume now $\gamma_k \neq 0$ and let us construct L_k and W_k . Note that $0 < \gamma_k < 1$. Indeed, if $\gamma_k = 1$ there would exist a non-zero vector $v \in V_1^{(k)} \cap V_2^{(k)} = V_1 \cap V_2 \cap W_{k-1} \subseteq V_0 \cap V_0^\perp$, which is impossible. Let $v_1^{(k)} \in V_1^{(k)}$ and $v_2^{(k)} \in V_2^{(k)}$, $\|v_1^{(k)}\| = \|v_2^{(k)}\| = 1$, be the vectors that give the maximum in (1.56) and $L_k = \text{span}\{v_1^{(k)}, v_2^{(k)}\}$. The vector $P_1 v_2^{(k)}$ belongs to $V_1^{(k)}$ since $v_2^{(k)} \in W_{k-1}$ and W_{k-1} is the invariant subspace of P_1 by induction hypothesis. Suppose that $P_1 v_2^{(k)}$ is not parallel to $v_1^{(k)}$. We have then the inequality

$$(v_1^{(k)}, v_2^{(k)}) = (v_1^{(k)}, P_1 v_2^{(k)}) < \|P_1 v_2^{(k)}\| = \left(\frac{P_1 v_2^{(k)}}{\|P_1 v_2^{(k)}\|}, v_2^{(k)} \right), \quad (1.58)$$

which contradicts the definition of $v_1^{(k)}$ and $v_2^{(k)}$. This means that $P_1 v_2^{(k)}$ is parallel to $v_1^{(k)}$, hence $P_1 L_k \subset L_k$. One can prove in the same manner that $P_2 v_1^{(k)}$ is parallel to $v_2^{(k)}$, hence $P_2 L_k \subset L_k$. Let $W_k = (V_0 \oplus W_{k-1} \oplus L_1 \oplus \dots \oplus L_k)^\perp$. The subspace $V_0 \oplus W_{k-1} \oplus L_1 \oplus \dots \oplus L_k$ is invariant with respect to P_1 and P_2 and so is the subspace W_k since operators P_1 and P_2 are symmetric.

Note at last that $W_{k-1} = W_k \oplus L_k$ hence for $k > 1$, $V_1^{(k)} \subset V_1^{(k-1)}$ and $V_2^{(k)} \subset V_2^{(k-1)}$, i.e. $\gamma_k \leq \gamma_{k-1}$ according to (1.56). Thus we have result (1.53). \square

At this point we have introduced the necessary tools to investigate the spectral properties of the iteration operator. This is the topic of the next section.

1.5 Spectral analysis of the iteration operator

The analysis of an algorithm described by its iteration operator is efficiently done by studying the spectral radius and norm of the latter. These properties give information about the characteristics of the method. The spectral radius gives the convergence speed in some norm, and the spectral norm is an upper bound for the factor of the reduction of the error in the spectral norm.

In this section we first recall results from earlier works (Proposition 1.8). Then, using the results from Section 1.4, we establish results (Proposition 1.9) recently published in [36], which we compare to the existing ones. Finally we consider a relaxation alternative for Algorithm 1.3 and analyze it through Proposition 1.10.

If $\mathcal{L}(V)$ is the space of linear and continuous operators from V into V , we denote by $\|B\| = \sup_{v \in V, \|v\|=1} \|Bv\|$ the norm of $B \in \mathcal{L}(V)$. If I denotes the identity operator in V

and ω is a real parameter, we define the operator $B \in \mathcal{L}(V)$ by

$$B = (I - \omega P_2)(I - \omega P_1). \quad (1.59)$$

We formulate first Proposition 1.8 for the norm of the operator B in order to get an estimate as presented in [37]. The idea of Proposition 1.8 and its proof come originally from Bramble et al. [24]. In their work, an abstract analysis of product iterative methods is presented and similar convergence estimates are given.

Comparable results proved using the technique from [24] can be found, for example, in early papers from Xu [64, 65] and Yserentant [71] appended by the work of Griebel and Oswald [40], in the article of Cai and Widlund [26] or Wang [61], and in an abstract theory presented by Widlund in [62]. More recent reports include the framework of the successive subspace correction algorithm by Xu and Zikatanov [67] and Xu [66]. Some estimates in the framework of an abstract convergence analysis of Schwarz methods are presented in textbooks, e.g., by Quarteroni and Valli [53, §4.6], Smith et al. [59, §5.2] and Wohlmuth [63, §2.1].

Proposition 1.8. *If Hypothesis 1.5 is satisfied and if $0 < \omega < 2$, then the norm of the operator B given by (1.59) verifies*

$$\|B\| \leq \left(1 - \frac{(2 - \omega)\omega}{C_0^2(1 + \omega\gamma)^2}\right)^{\frac{1}{2}} < 1. \quad (1.60)$$

Proof. The proof is adapted from [24] to the present setting and we establish it for the convenience of the reader. Introduce $R_1 = I - \omega P_1$ and $R_2 = (I - \omega P_2)(I - \omega P_1) = B$. We begin by proving

$$(2 - \omega)\omega (\|P_1 v\|^2 + \|P_2 R_1 v\|^2) = \|v\|^2 - \|Bv\|^2, \quad \forall v \in V. \quad (1.61)$$

As $v = R_1 v + \omega P_1 v$, $\|v\|^2 = \|R_1 v\|^2 + \omega^2 \|P_1 v\|^2 + 2\omega(R_1 v, P_1 v)$, and by definition $(R_1 v, P_1 v) = ((I - \omega P_1)v, P_1 v) = (1 - \omega)\|P_1 v\|^2$. Hence

$$\|v\|^2 - \|R_1 v\|^2 = [\omega^2 + 2\omega(1 - \omega)] \|P_1 v\|^2 = (2 - \omega)\omega \|P_1 v\|^2. \quad (1.62)$$

Furthermore, $R_1 v = R_2 v + \omega P_2 R_1 v$ implies $\|R_1 v\|^2 = \|R_2 v\|^2 + \omega^2 \|P_2 R_1 v\|^2 + 2\omega(R_2 v, P_2 R_1 v)$ and by definition $(R_2 v, P_2 R_1 v) = ((I - \omega P_2)R_1 v, P_2 R_1 v) = (1 - \omega)\|P_2 R_1 v\|^2$. Hence

$$\|R_1 v\|^2 - \|R_2 v\|^2 = (2 - \omega)\omega \|P_2 R_1 v\|^2. \quad (1.63)$$

Summing (1.62) and (1.63), we get (1.61).

We next prove

$$\|P_1 v\|^2 + \|P_2 v\|^2 \leq (1 + \gamma\omega)^2 (\|P_1 v\|^2 + \|P_2 R_1 v\|^2), \quad \forall v \in V. \quad (1.64)$$

Starting from $I - R_1 = \omega P_1$, we get

$$(P_2v, v) - (P_2v, R_1v) = \omega(P_2v, P_1v), \quad (1.65)$$

which implies that $\|P_2v\|^2 = (P_2v, R_1v) + \omega(P_2v, P_1v)$. Hence

$$\|P_1v\|^2 + \|P_2v\|^2 = (P_1v, P_1v) + (P_2v, P_2R_1v) + \omega(P_2v, P_1v) \quad (1.66)$$

$$\leq (\|P_1v\|^2 + \|P_2v\|^2)^{\frac{1}{2}} (\|P_1v\|^2 + \|P_2R_1v\|^2)^{\frac{1}{2}} + \omega(P_1v, P_2v). \quad (1.67)$$

From the definition (1.39) of γ we get

$$|(P_1v, P_2v)| \leq \gamma \|P_1v\| \|P_2v\| \leq \gamma (\|P_2v\| \|P_1v\| + \|P_1v\| \|P_2R_1v\|) \quad (1.68)$$

$$\leq \gamma (\|P_1v\|^2 + \|P_2v\|^2)^{\frac{1}{2}} (\|P_1v\|^2 + \|P_2R_1v\|^2)^{\frac{1}{2}}. \quad (1.69)$$

Thus we have

$$\|P_1v\|^2 + \|P_2v\|^2 \leq (1 + \omega\gamma) (\|P_1v\|^2 + \|P_2v\|^2)^{\frac{1}{2}} (\|P_1v\|^2 + \|P_2R_1v\|^2)^{\frac{1}{2}}, \quad (1.70)$$

which leads to (1.64).

Finally, we show that Hypothesis 1.5 implies

$$\|v\|^2 \leq C_0^2 (\|P_1v\|^2 + \|P_2v\|^2), \quad \forall v \in V. \quad (1.71)$$

When $v \in V$, there exist $v_1 \in V_1, v_2 \in V_2$ such that $v = v_1 + v_2$ and $\|v_1\|^2 + \|v_2\|^2 \leq C_0^2 \|v\|^2$ (see Hypothesis 1.5). Hence $\|v\|^2 = (v_1, v) + (v_2, v) = (v_1, P_1v) + (v_2, P_2v)$. Result (1.71) thus follows from:

$$\|v\|^2 \leq \|v_1\| \|P_1v\| + \|v_2\| \|P_2v\| \quad (1.72)$$

$$\leq (\|v_1\|^2 + \|v_2\|^2)^{\frac{1}{2}} (\|P_1v\|^2 + \|P_2v\|^2)^{\frac{1}{2}} \quad (1.73)$$

$$\leq C_0 \|v\| (\|P_1v\|^2 + \|P_2v\|^2)^{\frac{1}{2}}. \quad (1.74)$$

The proof of Proposition 1.8 is now straightforward. Combining (1.61) and (1.64), we get for all $v \in V$,

$$\frac{(2 - \omega)\omega}{(1 + \gamma\omega)^2} (\|P_1v\|^2 + \|P_2v\|^2) \leq \|v\|^2 - \|Bv\|^2, \quad (1.75)$$

and finally, (1.71) yields

$$\frac{(2 - \omega)\omega}{C_0^2(1 + \gamma\omega)^2} \|v\|^2 \leq \|v\|^2 - \|Bv\|^2. \quad (1.76)$$

Thus $\|Bv\|^2 \leq \left(1 - \frac{(2-\omega)\omega}{C_0^2(1+\gamma\omega)^2}\right) \|v\|^2$, i.e. $\|B\| \leq \left(1 - \frac{(2-\omega)\omega}{C_0^2(1+\gamma\omega)^2}\right)^{\frac{1}{2}}$ which is strictly bounded by one if $0 < \omega < 2$. \square

It is readily seen that the estimate of Proposition 1.8 is not optimal even in the case where $V = V_1 \oplus V_2$. In particular, if the space V is two-dimensional and V_1 and V_2 are one-dimensional subspaces of V , then $\|B\| = \gamma$ for $\omega = 1$. Indeed, $\forall v \in V$ we have in this case $\|Bv\|^2 = |(Bv, (I - P_1)v)| = \gamma\|Bv\|\|(I - P_1)v\|$ since $(I - P_1)v \in V_1^\perp$, $Bv \in V_2^\perp$ and the angle between V_1^\perp and V_2^\perp is equal to the angle between V_1 and V_2 . However, estimate (1.60) with the best choice of C_0 (1.44) gives only $\|B\| \leq \sqrt{\gamma(\gamma + 3)}/(1 + \gamma)$, which is optimal only if $\gamma = 0$. The non-optimality of (1.60) is also discussed, for example, by Griebel and Oswald in the concluding remarks of [40].

In the case where V_1 and V_2 are of finite dimension, an analysis of the spectral properties of B leads to exact formulas for its spectral radius and its norm. Hereafter we present these results published in [36].

For $\tilde{\gamma}$ and $\omega \in (0; 2)$ we define the functions

$$\rho(\tilde{\gamma}, \omega) = \begin{cases} \frac{1}{2}\omega^2\tilde{\gamma}^2 - \omega + 1 + \frac{1}{2}\omega\tilde{\gamma}\sqrt{\omega^2\tilde{\gamma}^2 - 4\omega + 4}, & \text{if } \omega \leq \omega_0(\tilde{\gamma}), \\ \omega - 1, & \text{otherwise,} \end{cases} \quad (1.77)$$

where

$$\omega_0(\tilde{\gamma}) = \begin{cases} \frac{2-2\sqrt{1-\tilde{\gamma}^2}}{\tilde{\gamma}^2}, & \text{for } \tilde{\gamma} \in (0; 1], \\ 1, & \text{for } \tilde{\gamma} = 0, \end{cases} \quad (1.78)$$

and

$$N(\tilde{\gamma}, \omega) = \frac{1}{2}\omega(2 - \omega)\tilde{\gamma} + \sqrt{\frac{1}{4}\omega^2(2 - \omega)^2\tilde{\gamma}^2 + (\omega - 1)^2}. \quad (1.79)$$

Proposition 1.9 (see [36]). *Let V be of finite dimension, $V = V_1 + V_2$ and $\tilde{\gamma}$ be defined by (1.41). The spectral radius of operator B given by (1.59) is a function of $\tilde{\gamma}$ and $\omega \in (0; 2)$ given by $\rho(B) = \rho(\tilde{\gamma}, \omega)$. The norm of B is a function of $\tilde{\gamma}$ and $\omega \in (0; 2)$ given by $\|B\| = N(\tilde{\gamma}, \omega)$.*

We repeat here the proof from [36].

Proof. The idea of the proof is to establish first all the results in the two-dimensional case and to use then decomposition (1.54) to extend the results to the general case. Therefore, we assume first that the space V is two-dimensional and V_1 and V_2 are one-dimensional subspaces of V spanned by the vectors v_1 and v_2 , respectively. Figure 1.6 illustrates this situation and the construction of Bv for $v \in V$ with $\omega = 0.25$.

Without loss of generality, we can assume that $\|v_1\| = \|v_2\| = 1$ and $(v_1, v_2) = \tilde{\gamma}$. We can verify that the linear operator B is represented in the basis $\{v_1, v_2\}$ by the matrix

$$\mathbf{B} = \begin{pmatrix} 1 - \omega & -\omega\tilde{\gamma} \\ \omega(\omega - 1)\tilde{\gamma} & \omega^2\tilde{\gamma}^2 + 1 - \omega \end{pmatrix}. \quad (1.80)$$

The characteristic polynomial of this matrix is

$$p(\lambda) = \lambda^2 - (\omega^2\tilde{\gamma}^2 - 2\omega + 2)\lambda + (\omega - 1)^2. \quad (1.81)$$

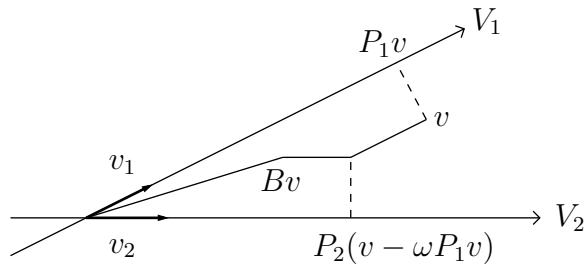


Figure 1.6: Illustration of the construction of Bv for $v \in V$ ($\dim V = 2$, $\omega = 0.25$).

If $\tilde{\gamma} > 0$ and $\omega \in (\omega_0(\tilde{\gamma}); 2)$, $p(\lambda)$ has two complex conjugate roots λ_{\pm} such that $|\lambda_{\pm}| = \omega - 1$. If $\tilde{\gamma} > 0$ and $\omega \in (0; \omega_0(\tilde{\gamma}))$, $p(\lambda)$ has two real roots λ_{\pm} given by

$$\lambda_{\pm} = \frac{1}{2}\omega^2\tilde{\gamma}^2 - \omega + 1 \pm \frac{1}{2}\omega\tilde{\gamma}\sqrt{\omega^2\tilde{\gamma}^2 - 4\omega + 4}. \quad (1.82)$$

If $\tilde{\gamma} = 0$, $p(\lambda)$ has the only double root $\lambda = 1 - \omega$. Identity $\rho(B) = \rho(\tilde{\gamma}, \omega)$ is thus proved in the two-dimensional case.

Let us consider now the norm of operator B that can be written as

$$\|B\|^2 = \max_{x \in \mathbb{R}^2, x \neq 0} \frac{x^T \mathbf{B}^T \mathbf{\Gamma} \mathbf{B} x}{x^T \mathbf{\Gamma} x}, \quad (1.83)$$

where $\mathbf{\Gamma}$ is the Gramm matrix of the basis $\{v_1, v_2\}$,

$$\mathbf{\Gamma} = \begin{pmatrix} 1 & \tilde{\gamma} \\ \tilde{\gamma} & 1 \end{pmatrix}. \quad (1.84)$$

By making the substitution $y = \mathbf{\Gamma}^{1/2}x$, we can rewrite (1.83) as

$$\|B\|^2 = \max_{y \in \mathbb{R}^2, y \neq 0} \frac{y^T \mathbf{\Gamma}^{-1/2} \mathbf{B}^T \mathbf{\Gamma} \mathbf{B} \mathbf{\Gamma}^{-1/2} y}{y^T y}. \quad (1.85)$$

Since the matrix $\mathbf{C} = \mathbf{\Gamma}^{-1/2} \mathbf{B}^T \mathbf{\Gamma} \mathbf{B} \mathbf{\Gamma}^{-1/2}$ is symmetric positive definite, (1.85) implies that $\|B\|^2$ is equal to the spectral radius of \mathbf{C} . Let μ^2 be an eigenvalue of \mathbf{C} , then

$$\det(\mathbf{C} - \mu^2 \mathbf{I}) = 0. \quad (1.86)$$

But

$$\begin{aligned} & \det(\mathbf{C} - \mu^2 \mathbf{I}) \\ &= \det(\mathbf{B}^T \mathbf{\Gamma} \mathbf{B} \mathbf{\Gamma}^{-1} - \mu^2 \mathbf{I}) \end{aligned} \quad (1.87)$$

$$= \mu^4 - \mu^2 \text{tr}(\mathbf{B}^T \mathbf{\Gamma} \mathbf{B} \mathbf{\Gamma}^{-1}) + \det(\mathbf{B}^T \mathbf{\Gamma} \mathbf{B} \mathbf{\Gamma}^{-1}) \quad (1.88)$$

$$= \mu^4 - \mu^2[(2 - \omega)^2 \omega^2 \tilde{\gamma}^2 + 2(\omega - 1)^2] + (\omega - 1)^4 \quad (1.89)$$

$$= (\mu^2 - \omega(2 - \omega)\tilde{\gamma}\mu - (\omega - 1)^2)(\mu^2 + \omega(2 - \omega)\tilde{\gamma}\mu - (\omega - 1)^2). \quad (1.90)$$

The roots of (1.86) are thus given by

$$\mu = \pm \frac{1}{2} \omega (2 - \omega) \tilde{\gamma} \pm \sqrt{\frac{1}{4} \omega^2 (2 - \omega)^2 \tilde{\gamma}^2 + (\omega - 1)^2}, \quad (1.91)$$

and the largest among them gives $\|B\|$, i.e. identity $\|B\| = N(\tilde{\gamma}, \omega)$ is proved in the two-dimensional case.

Let us turn now to the general case. According to Proposition 1.7, V can be decomposed into the direct sum (1.54) where all the summands are invariant subspaces of projectors P_1 and P_2 , and hence of B . Hence the spectrum of B is given by the set of all eigenvalues of the operators $B_0 = B|_{V_0}$, $B_{12} = B|_{V_1^\perp \cap V_2}$, $B_{21} = B|_{V_1 \cap V_2^\perp}$ and $B_m = B|_{L_m}$, $m = 1, 2, \dots, p$, where here $B|_W$ is the restriction of B to W . We verify easily that $\rho(B_0) = (1 - \omega)^2$, $\rho(B_{12}) = \rho(B_{21}) = |1 - \omega|$, and concerning the two-dimensional spaces L_m , $m = 1, 2, \dots, p$, we have proved just above that $\rho(B_m) = \rho(\gamma_m, \omega)$ where $\rho(\gamma, \omega)$ is defined by (1.77). Hence

$$\rho(B) = \max \left((1 - \omega)^2, |1 - \omega|, \rho(\gamma_1, \omega), \dots, \rho(\gamma_p, \omega) \right). \quad (1.92)$$

It is easy to verify that $\omega_0(\gamma)$ is an increasing function and for fixed ω , $\rho(\gamma, \omega)$ is a non-decreasing function. It follows that we have $\rho(\gamma_1, \omega) \geq \dots \geq \rho(\gamma_p, \omega) > \rho(0, \omega) = |1 - \omega|$. Since $\tilde{\gamma} = \gamma_1$ if $p > 0$ and $\tilde{\gamma} = 0$ if $p = 0$, we conclude that $\rho(B) = \rho(\tilde{\gamma}, \omega)$. Analogously, since all the subspaces in (1.54) are mutually orthogonal, Pythagore's Theorem implies

$$\|B\| = \max \left((1 - \omega)^2, |1 - \omega|, N(\gamma_1, \omega), \dots, N(\gamma_p, \omega) \right), \quad (1.93)$$

where $N(\gamma, \omega)$ is defined by (1.79). Noting that $N(0, \omega) = |1 - \omega|$, we conclude that $\|B\| = N(\tilde{\gamma}, \omega)$. \square

Finally, let us observe that:

1. The spectral radius $\rho(B)$ is less than one for $\omega \in (0; 2)$ and, for $\tilde{\gamma}$ given by (1.41), attains the minimum value $\rho(B) = \omega_0(\tilde{\gamma}) - 1$ at $\omega = \omega_0(\tilde{\gamma}) \in [1; 2)$. We have $\rho(B) = \tilde{\gamma}^2$ at $\omega = 1$.
2. The norm $\|B\|$ is less than one for $\omega \in (0; 2)$ and, for $\tilde{\gamma}$ given by (1.41), attains the minimum value $\|B\| = \tilde{\gamma}$ at $\omega = 1$. This last result is given by Blaheta in [17].
3. The functions $\rho(\tilde{\gamma}, \omega)$ and $N(\tilde{\gamma}, \omega)$ are non-decreasing with respect to $\tilde{\gamma}$ for any fixed value of $\omega \in (0; 2)$.
4. Both formulas (1.77) and (1.79) can be rewritten in the case $V_1 \neq V_2$ as functions only of C_0^{opt} and ω due to the relation (1.44).

These properties are illustrated in the Figures 1.7–1.10. The plots in Figures 1.7 and 1.9 illustrate the functions $\rho(\tilde{\gamma}, \omega)$ and $N(\tilde{\gamma}, \omega)$ respectively for given $\tilde{\gamma} = 0.3, 0.6$,

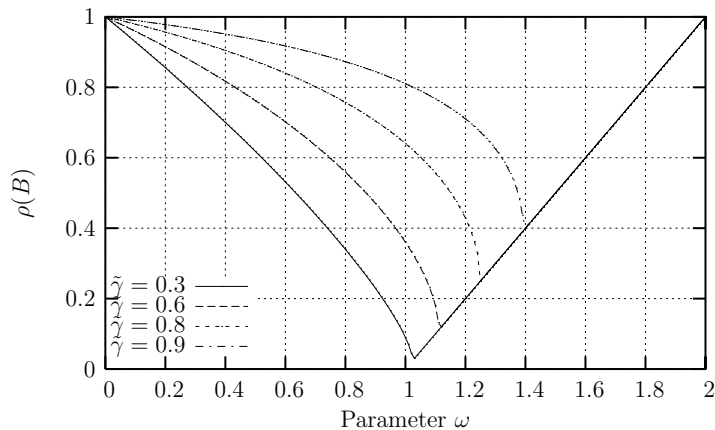


Figure 1.7: Illustration of $\rho(\tilde{\gamma}, \omega)$ for $\tilde{\gamma} = 0.3, 0.6, 0.8$ and 0.9 .

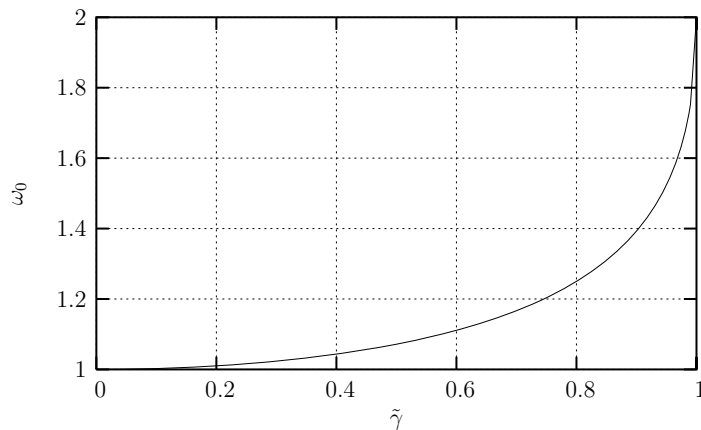


Figure 1.8: Illustration of $\omega_0(\tilde{\gamma})$.

0.8 and 0.9. Figure 1.8 depicts the relation $\omega_0(\tilde{\gamma})$. Note that as $\tilde{\gamma}$ tends to one, the optimal parameter ω_0 yields 2. On Figure 1.10 we compare the non-optimal bound for $\|B\|$ given by (1.60) with its exact value (1.79) for $\gamma = \tilde{\gamma} = 0.5$. Remark in particular that the bound (1.60) does not provide the optimal value for the parameter ω . The bound (1.60) suggests to choose $\omega < 1$ while the exact expression of the spectral radius (1.77) yields the optimal value $\omega = \omega_0 \geq 1$ given by (1.78) for the best convergence speed.

At this point it is natural to inquire about an alternative to the proposed relaxation in Algorithm 1.3.

Let us consider Algorithm 1.3 with two relaxation parameters ω_h and ω_H . For this, we replace equations (1.33) and (1.35) as follows: In the fine correction step (1.33) we write

$$u^{n-\frac{1}{2}} = u^{n-1} + \omega_h w_h, \quad (1.94)$$

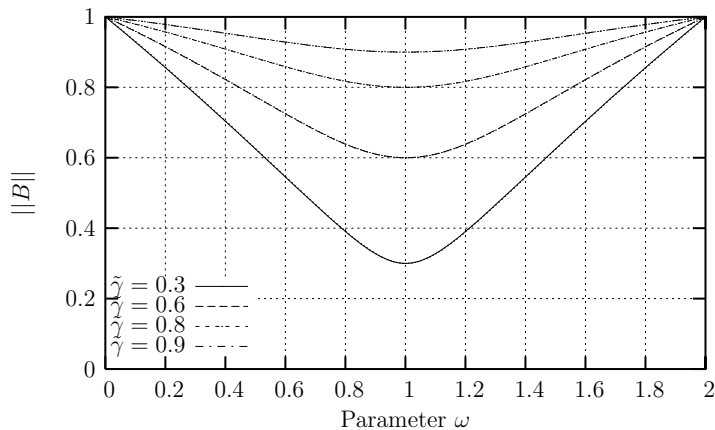


Figure 1.9: Illustration of $N(\tilde{\gamma}, \omega)$ for $\tilde{\gamma} = 0.3, 0.6, 0.8$ and 0.9 .

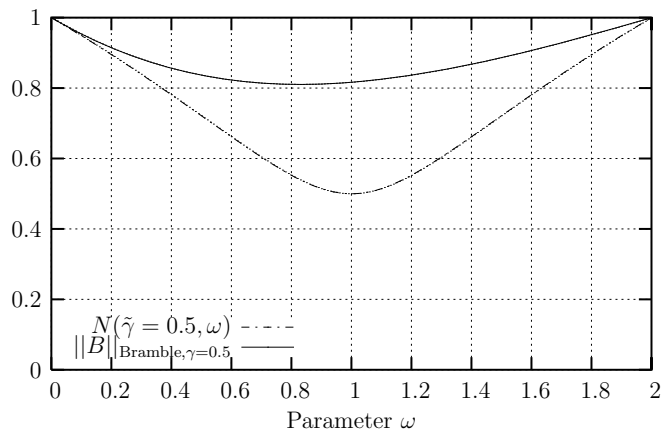


Figure 1.10: Comparison of the bound (1.60) and the exact value $N(\tilde{\gamma}, \omega)$ given by (1.79) for the norm $\|B\|$.

and the coarse correction (1.35) becomes

$$u^n = u^{n-\frac{1}{2}} + \omega_H w_H. \quad (1.95)$$

The evaluation of w_h and w_H through equations (1.32) and (1.34) remains unchanged. Hence the iteration operator, to be compared with (1.37), becomes

$$(I - \omega_H P_H)(I - \omega_h P_h). \quad (1.96)$$

Following the above introduced notation, and comparing with (1.59), we introduce the operator $B_2 \in \mathcal{L}(V)$, defined by

$$B_2 = (I - \omega_2 P_2)(I - \omega_1 P_1), \quad (1.97)$$

where ω_1, ω_2 are real parameters.

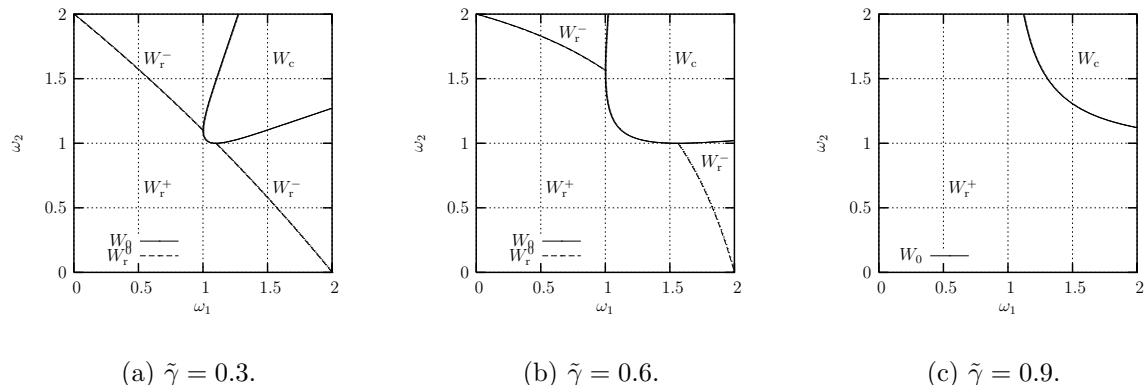


Figure 1.11: Illustration of domains $W_r(\tilde{\gamma}) = W_r^-(\tilde{\gamma}) \cup W_r^+(\tilde{\gamma})$ and $W_c(\tilde{\gamma})$ for different $\tilde{\gamma}$.

For $\tilde{\gamma}$ and $(\omega_1, \omega_2) \in (0; 2) \times (0; 2)$ we define the function

$$\rho_2(\tilde{\gamma}, \omega_1, \omega_2) = \begin{cases} \left| \frac{1}{2}\omega_1\omega_2\tilde{\gamma}^2 - \frac{1}{2}(\omega_1 + \omega_2) + 1 \right| \\ \quad + \frac{1}{2}\sqrt{(\omega_1 - \omega_2)^2 + \tilde{\gamma}^2\omega_1\omega_2 [\tilde{\gamma}^2\omega_1\omega_2 - 2(\omega_1 + \omega_2) + 4]}, & \text{if } (\omega_1, \omega_2) \in W_r(\tilde{\gamma}) \cup W_0(\tilde{\gamma}), \tilde{\gamma} > 0, \text{ or if } \tilde{\gamma} = 0, \\ \sqrt{(\omega_1 - 1)(\omega_2 - 1)}, & \text{if } (\omega_1, \omega_2) \in W_c(\tilde{\gamma}), \tilde{\gamma} > 0, \end{cases} \quad (1.98)$$

where $W_0(\tilde{\gamma})$ is the arc defined by

$$W_0(\tilde{\gamma}) = \left\{ (\omega_1, \omega_2) \in (0; 2) \times (0; 2) : (\omega_1 - \omega_2)^2 + \tilde{\gamma}^2\omega_1\omega_2 [\tilde{\gamma}^2\omega_1\omega_2 - 2(\omega_1 + \omega_2) + 4] = 0 \right\}, \quad (1.99)$$

splitting the domain $(0; 2) \times (0; 2)$, when $\tilde{\gamma} > 0$, into two disjoint open sub-domains, $W_r(\tilde{\gamma})$ the closure of which contains $(\omega_1, \omega_2) = (0, 0)$, and $W_c(\tilde{\gamma})$ the closure of which contains $(2, 2)$.

Furthermore, in order to write out $\left| \frac{1}{2}\omega_1\omega_2\tilde{\gamma}^2 - \frac{1}{2}(\omega_1 + \omega_2) + 1 \right|$, it is useful to introduce, for $0 < \tilde{\gamma} < \sqrt{2}/2$, the arc

$$W_r^0(\tilde{\gamma}) = \left\{ (\omega_1, \omega_2) \in (0; 2) \times (0; 2) : \frac{1}{2}\omega_1\omega_2\tilde{\gamma}^2 - \frac{1}{2}(\omega_1 + \omega_2) + 1 = 0 \right\}, \quad (1.100)$$

which splits $W_r(\tilde{\gamma})$ into disjoint open sub-domains, $W_r^+(\tilde{\gamma})$ the closure of which contains $(\omega_1, \omega_2) = (0, 0)$, and $W_r^-(\tilde{\gamma}) = W_r(\tilde{\gamma}) \setminus W_r^+(\tilde{\gamma})$. For $\sqrt{2}/2 < \tilde{\gamma} \leq 1$, we identify $W_r^+(\tilde{\gamma}) = W_r(\tilde{\gamma})$. Figure 1.11 illustrates the above introduced domains for different $\tilde{\gamma}$.

Proposition 1.10. *Let V be of finite dimension, $V = V_1 + V_2$ and $\tilde{\gamma}$ defined by (1.41). The spectral radius of operator B_2 given by (1.97) is a function of $\tilde{\gamma}$ and $(\omega_1, \omega_2) \in (0; 2) \times (0; 2)$ given by $\rho(B_2) = \rho_2(\tilde{\gamma}, \omega_1, \omega_2)$. For given $\tilde{\gamma}$, the spectral radius of B_2 is minimum when $\omega_1 = \omega_2 = \omega_0(\tilde{\gamma})$ where the function ω_0 is given by (1.78).*

Proof. Following the proof of Proposition 1.9, we write, when V is a two-dimensional space and $V = V_1 \oplus V_2$, the operator (1.97) in matrix-form

$$\begin{pmatrix} 1 - \omega_1 & -\omega_1 \tilde{\gamma} \\ \omega_2 (\omega_1 - 1) \tilde{\gamma} & \omega_1 \omega_2 \tilde{\gamma}^2 + 1 - \omega_2 \end{pmatrix}, \quad (1.101)$$

and study its characteristic polynomial

$$p(\lambda) = \lambda^2 - \lambda (\omega_1 \omega_2 \tilde{\gamma}^2 - \omega_1 - \omega_2 + 2) + (1 - \omega_1)(1 - \omega_2). \quad (1.102)$$

If $\tilde{\gamma} > 0$ and $(\omega_1, \omega_2) \in W_c(\tilde{\gamma})$, $p(\lambda)$ has two complex conjugate roots λ_{\pm} such that $|\lambda_{\pm}| = \sqrt{(\omega_1 - 1)(\omega_2 - 1)}$. If $\tilde{\gamma} > 0$ and $(\omega_1, \omega_2) \in W_r(\tilde{\gamma})$, $p(\lambda)$ has two real roots λ_{\pm} given by

$$\begin{aligned} \lambda_{\pm}(\tilde{\gamma}, \omega_1, \omega_2) &= \frac{1}{2} \omega_1 \omega_2 \tilde{\gamma}^2 - \frac{1}{2} (\omega_1 + \omega_2) + 1 \\ &\quad \pm \frac{1}{2} \sqrt{(\omega_1 - \omega_2)^2 + \tilde{\gamma}^2 \omega_1 \omega_2 [\tilde{\gamma}^2 \omega_1 \omega_2 - 2(\omega_1 + \omega_2) + 4]}. \end{aligned} \quad (1.103)$$

If $\tilde{\gamma} = 0$, $p(\lambda)$ has the two real roots $\lambda = 1 - \omega_1$ and $\lambda = 1 - \omega_2$. Identity $\rho(B_2) = \rho_2(\tilde{\gamma}, \omega_1, \omega_2)$ is thus proved when V is a two-dimensional space.

In the general case, we follow again the proof of Proposition 1.9. With the notation of the latter, in particular with $\rho(B_{2,0}) = \rho(B_2|_{V_0}) = |(1 - \omega_1)(1 - \omega_2)|$, $\rho(B_{2,12}) = \rho(B_2|_{V_1^{\perp} \cap V_2}) = |1 - \omega_2|$, and $\rho(B_{2,21}) = \rho(B_2|_{V_1 \cap V_2^{\perp}}) = |1 - \omega_1|$, we conclude that $\rho(B_2) = \rho_2(\tilde{\gamma}, \omega_1, \omega_2)$. Note that $\rho(B_2)$ is less than one for $(\omega_1, \omega_2) \in (0; 2) \times (0; 2)$.

We are now left with proving that, for given $\tilde{\gamma}$, $\rho(B_2)$ is minimum when $\omega_1 = \omega_2 = \omega_0(\tilde{\gamma})$ given by (1.78). For this we show that for given $\tilde{\gamma}$ and for any $(\omega_1, \omega_2) \in (0; 2) \times (0; 2)$ we have $\rho_2(\tilde{\gamma}, \omega_1, \omega_2) \geq w_0(\tilde{\gamma}) - 1$, the last expression being an equality if and only if $\omega_1 = \omega_2 = \omega_0(\tilde{\gamma})$.

If $\omega_1 = \omega_2$, the result is proved by Proposition 1.9. If $\tilde{\gamma} = 0$, $\rho_2(\tilde{\gamma}, \omega_1, \omega_2) = \max(|1 - \omega_1|, |1 - \omega_2|)$, which is minimal, i.e. reduces to zero for $\omega_1 = \omega_2 = 1 = \omega_0(\tilde{\gamma} = 0)$.

Suppose now $\tilde{\gamma} > 0$. For $(\omega_1, \omega_2) \in W_0(\tilde{\gamma})$,

$$\rho_2(\tilde{\gamma}, \omega_1, \omega_2) = \left| \frac{1}{2} \omega_1 \omega_2 \tilde{\gamma}^2 - \frac{1}{2} (\omega_1 + \omega_2) + 1 \right| = \sqrt{(\omega_1 - 1)(\omega_2 - 1)}. \quad (1.104)$$

Without loss of generality due to the symmetry in ω_1 and ω_2 , suppose that $w_2 \leq w_1$. For $(\omega_1, \omega_2) \in W_0(\tilde{\gamma})$, $w_2 \leq w_1$, using the definition (1.99), we write ω_2 as a function of ω_1 , namely

$$\omega_2(\omega_1) = \frac{\omega_1}{(1 - \tilde{\gamma}^2 \omega_1)^2} \left(1 - 2\tilde{\gamma}^2 + \tilde{\gamma}^2 \omega_1 - 2\tilde{\gamma} \sqrt{1 - \tilde{\gamma}^2} \sqrt{\omega_1 - 1} \right). \quad (1.105)$$

For a given parameter $\tilde{\gamma}$, we explicit now $\rho_2(\tilde{\gamma}, \omega_1, \omega_2)$ from (1.104) with $\omega_2 = \omega_2(\omega_1)$ from (1.105), and obtain $\rho_{\tilde{\gamma}}(\omega_1) \doteq \rho_2(\tilde{\gamma}, \omega_1, \omega_2(\omega_1))$. Minimizing $\rho_{\tilde{\gamma}}(\omega_1)$ with respect to ω_1 , yields that the minimum is reached at $\omega_1 = \omega_0(\tilde{\gamma})$ given by (1.78) and $\omega_2(\omega_1) = \omega_1$. Hence the minimum of $\rho_2(\tilde{\gamma}, \omega_1, \omega_2)$ on $W_0(\tilde{\gamma})$ is given by $w_0(\tilde{\gamma}) - 1$.

Let us now consider $(\omega_1, \omega_2) \notin W_0(\tilde{\gamma})$. For $(\omega_1, \omega_2) \in W_c(\tilde{\gamma})$, $\rho_2(\tilde{\gamma}, \omega_1, \omega_2)$ is an increasing function in ω_1 and ω_2 and hence $\rho_2(\tilde{\gamma}, \omega_1, \omega_2) > w_0(\tilde{\gamma}) - 1$. Furthermore, if $(\omega_1, \omega_2) \in W_r(\tilde{\gamma})$, we introduce the vector $\mathbf{n} = (-1, 1)$ in the plane (ω_1, ω_2) . Hence, evaluating $\nabla \rho_2(\tilde{\gamma}, \omega_1, \omega_2) \doteq \left(\frac{\partial \rho_2(\tilde{\gamma}, \omega_1, \omega_2)}{\partial \omega_1}, \frac{\partial \rho_2(\tilde{\gamma}, \omega_1, \omega_2)}{\partial \omega_2} \right)$ yields

$$\nabla \rho_2(\tilde{\gamma}, \omega_1, \omega_2) \cdot \mathbf{n} = \begin{cases} \frac{(\omega_1 - \omega_2)(\rho_2(\tilde{\gamma}, \omega_1, \omega_2)\tilde{\gamma}^2 - 1)}{\sqrt{(\omega_1 - \omega_2)^2 + \tilde{\gamma}^2 \omega_1 \omega_2 [\tilde{\gamma}^2 \omega_1 \omega_2 - 2(\omega_1 + \omega_2) + 4]}}, & \text{if } (\omega_1, \omega_2) \in W_r^+(\tilde{\gamma}), \tilde{\gamma} > 0, \\ \frac{(\omega_1 - \omega_2)(-\rho_2(\tilde{\gamma}, \omega_1, \omega_2)\tilde{\gamma}^2 - 1)}{\sqrt{(\omega_1 - \omega_2)^2 + \tilde{\gamma}^2 \omega_1 \omega_2 [\tilde{\gamma}^2 \omega_1 \omega_2 - 2(\omega_1 + \omega_2) + 4]}}, & \text{if } (\omega_1, \omega_2) \in W_r^-(\tilde{\gamma}), 0 < \tilde{\gamma} \leq \sqrt{2}/2, \end{cases} \quad (1.106)$$

which is strictly negative for $\omega_1 > \omega_2$, and strictly positive for $\omega_2 > \omega_1$. Hence the minimum of $\rho_2(\tilde{\gamma}, \omega_1, \omega_2)$ in $W_r(\tilde{\gamma})$ is to be found for $\omega_2 = \omega_1$, reducing our problem to the result of Proposition 1.9. \square

Given the result of Proposition 1.10, we stick to Algorithm 1.3 with one relaxation parameter ω .

1.6 Convergence of the algorithm

After the foregoing study of some properties of vector spaces (Section 1.4) and the abstract analysis of the iteration operator B (Section 1.5), we are now able to give a new convergence result for the two-scale algorithm.

We set $V_{Hh0} = V_H \cap V_h$ and V_{Hh0}^\perp the orthogonal complement of V_{Hh0} in V_{Hh} . Setting $V_1 = V_h$, $V_2 = V_H$ and $V_0 = V_{Hh0}$, the definition (1.41) of $\tilde{\gamma}$ rewrites

$$\tilde{\gamma} = \begin{cases} \sup_{\substack{v_h \in V_h \cap V_{Hh0}^\perp, v_h \neq 0 \\ v_H \in V_H \cap V_{Hh0}, v_H \neq 0}} \frac{a(v_h, v_H)}{\|v_h\| \|v_H\|}, & \text{if } V_h \neq V_{Hh0} \text{ and } V_H \neq V_{Hh0}, \\ 0, & \text{otherwise.} \end{cases} \quad (1.107)$$

Recalling definitions (1.77–1.79) and Proposition 1.9, we have the following proposition.

Proposition 1.11. *If $\omega \in (0; 2)$, then Algorithm 1.3 converges, i.e. $\lim_{n \rightarrow \infty} \|u^n - u_{Hh}\| = 0$. The spectral radius of the iteration operator defined by (1.37) is given by $\rho(B) = \rho(\tilde{\gamma}, \omega)$. The convergence factor in the norm induced by the scalar product $a(\cdot, \cdot)$ is bounded by $\|B\| = N(\tilde{\gamma}, \omega)$.*

Proof. Proposition 1.11 is readily proved by applying Proposition 1.9 to $V = V_{Hh}$, $V_1 = V_h$ and $V_2 = V_H$ using the form $a(\cdot, \cdot)$ as scalar product. \square

The convergence speed in some norm is given by $\rho(B)$ and the factor of reduction of the error in the norm $a(\cdot, \cdot)^{\frac{1}{2}}$ is bounded by $\|B\|$. The new aspect we have introduced here

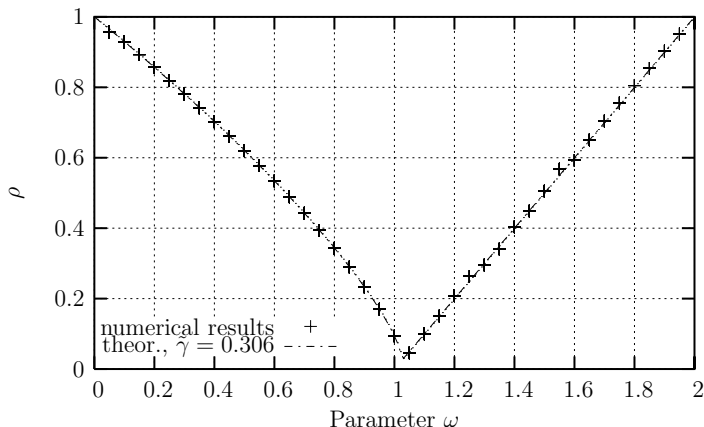


Figure 1.12: Comparison of numerical estimates and analytical results for ρ for different parameter ω in the case of the nested grids in Figure 1.3(a) and the norm induced by the scalar product (1.18).

is to link the speed of convergence of the algorithm to one only parameter corresponding to an abstract angle between the spaces V_h and V_H . This also leads to a method to find the optimal relaxation parameter ω (see Section 2.2). Hence the parameter for optimal relaxation is solely linked to the grid constellation.

Since Proposition 1.10 we know that the introduction of one only relaxation parameter is optimal with regard to the alternative of a coarse and a fine relaxation parameter studied at the end of Section 1.5.

Since Proposition 1.11 we have an exact expression for the spectral radius $\rho(B)$. For given $\tilde{\gamma}$, the function $\rho(\tilde{\gamma}, \omega)$ is plotted in Figure 1.7. At this point, it is interesting to compare, for given meshes and a given scalar product, this algebraic result with numerical estimates of the spectral radius.

In Figure 1.12 we compare the plot of $\rho(\tilde{\gamma}, \omega)$ to the results obtained numerically for the spectral radius in the case of the nested grids depicted in Figure 1.3(a) and the norm induced by the scalar product (1.18). First we evaluate the spectral radius for $\omega = 1$ which enables us to get an approximation of $\tilde{\gamma}$ (see Section 2.2, Proposition 2.3 and the description of the method). We obtain that $\tilde{\gamma} = 0.306$. Thus we can plot the function $\rho(\tilde{\gamma} = 0.306, \omega)$ given by (1.77). On the same plot we superpose the corresponding numerical results for some values of $\omega \in (0; 2)$.

The following chapter consists in analyzing several aspects of the method and results presented. We consider estimates of the parameter $\tilde{\gamma}$ (Section 2.1) and numerical evaluations of ω_0 (Section 2.2). After a discussion of the implementation issues (Section 2.3) we refer the reader to Chapters 3 and 4 for applications of the algorithm in two dimensions.

Chapter 2

Parameter discussion and computational considerations

In the previous chapter, we have introduced and proved the convergence of an iterative correction method (Algorithm 1.3) on two discretization spaces. The results show that the speed of convergence depends on two parameters: the cosine of the abstract “angle” $\tilde{\gamma}$ (1.107) between the spaces and the relaxation parameter ω . We have established a formula (1.78) giving the optimal relaxation parameter for a given $\tilde{\gamma}$. The first objective of this chapter is to discuss some estimates for the parameter $\tilde{\gamma}$. We give a practical method which allows to evaluate the latter and hence the optimum value ω^{opt} for ω . Finally, we care about implementation issues and the usage of memory, in particular concerning integration of the scalar products.

The outline of this chapter is the following:

2.1	Estimates of the C.B.S. constant	36
2.2	Numerical evaluation of $\tilde{\gamma}$ and optimal relaxation	42
2.3	Implementation issues and memory and CPU-time usage .	48

In Section 2.1 we discuss estimates of the optimal constant of the Cauchy-Buniakowski-Schwarz inequality which is related to the parameter $\tilde{\gamma}$. We illustrate the influence of nested and non-nested grid constellations to the parameter in one dimension. We give estimates in some particular two-dimensional cases and briefly consider a particular 2D or 3D situation (Proposition 2.2) where the patch is entirely included in one element of the coarse grid. Section 2.2 introduces a general method, based on Algorithm 1.3, how to approximate $\tilde{\gamma}$ numerically. We discuss the evaluation of the optimal value for the relaxation parameter. In Section 2.3 we consider computational issues and assess the convergence of the method in practice with respect to the usage of memory and CPU-time.

2.1 Estimates of the C.B.S. constant

In the previous chapter we have seen that the optimal constant of the Cauchy-Buniakowski-Schwarz (C.B.S.) inequality (1.40) is closely related to the spectral analysis of the iteration operator. In particular situations the constant $\tilde{\gamma}$ equals γ , and hence the optimal C.B.S. constant. Our objective is now to discuss this constant and give some estimates in simple cases.

In the following we give a small survey of existing works on the constant and give some new estimates. We limit ourselves on presenting a couple of particular situations to get an idea of the involvement of the finite elements and triangulation used. We first consider some 1D situations where we discuss the constants γ and $\tilde{\gamma}$. Then we derive a general upper bound for γ (Proposition 2.2) for polynomial spaces of order 1 and a particular problem in d dimensions ($d = 2, 3$) where the patch is entirely included in one element of the coarse triangulation. We specify our result for the situation where the coefficients of the differential operator (2.8) are constant over the patch. Next, we come back to a 2D situation where the patch is included in the union of two coarse elements: in some particular cases (see Figure 2.4) with first order polynomials, we give an upper bound for the parameter γ (Table 2.1). Finally, we conclude with a numerical method using Algorithm 1.3 and the spectral properties of the iteration operator reported in Proposition 1.11 to evaluate the constant $\tilde{\gamma}$.

Estimates and upper bounds for the constant from the C.B.S. inequality are abundant in the literature as it is the main tool in the convergence analysis of many methods. The C.B.S. inequality has been used in two-level methods by Axelsson [6], Axelsson and Gustavson [9], Braess [20, 21], Maître and Musy [47]. A survey of the role of this constant is reported by Axelsson and Vassilevski [10, 11] and by Eijkhout and Vassilevski [32]. The constant is also used in local refinement preconditioning methods, e.g., by McCormick [49] and Bramble et al. [23]. The latest papers present estimates of γ depending generally on the scalar product, i.e. the bilinear form a , the problem coefficients, and the type and shape of the finite element used. In some cases it is possible to have universal bounds [8]. Margenov [48] gives estimates of the 2D elasticity problem on a triangular mesh and piecewise linear approximation. More recently and for the same problem, Achchab and Maître [3] and Axelsson [7] proved that the constant γ^2 is bounded from above by $3/4$. Numerical experiments by Jung and Maître [45] generalize the latter to more choices of finite elements. General estimates in 3D have been developed over the last years, see, e.g., the papers by Achchab et al. [1, 2].

The one-dimensional case.

Recall the introductory discussion of Section 1.1 where we first mentioned the problem of exhibiting a finite element-type basis for the space $V_{Hh} = V_H + V_h$ (given by equations 1.7 and 1.8). We also accounted that depending on the relative nestedness of the underlying subspaces the dimension of V_{Hh} changes. The relative grid constellation is determinant for the “angle” between the subspaces. In the following we present some finite element considerations for the evaluation of the constants γ and $\tilde{\gamma}$ in 1D. We consider the interval $(0; 1)$, the scalar product $a(u, v) = \int_{\Omega} u'v' dx$, and the induced norm $\|\cdot\| = a(\cdot, \cdot)^{1/2}$ in $H_0^1(0; 1)$.

We study a first case with a fine 1D mesh \mathcal{T}_h over $\bar{\Lambda} = [a; b] \subset (0; 1)$. Let $\{x_H^j\}_{j=0}^{N+1}$ and $\{x_h^j\}_{j=0}^{M+1}$ denote the set of regularly distributed nodes of a coarse mesh \mathcal{T}_H over $[0; 1]$ and the fine mesh \mathcal{T}_h respectively: set $H = 1/(N + 1)$ and $x_H^j = jH$, $j = 0, \dots, N + 1$, and $h = |b - a|/(M + 1)$ and $x_h^j = a + jh$, $j = 0, \dots, M + 1$. We consider \mathcal{T}_h such that the coarse nodes x_H^j that are in Λ coincide with fine grid points x_h^j . We construct the “hat” finite element functions φ_H^j , $j = 1, \dots, N$, and φ_h^j , $j = 1, \dots, M$, such that $\varphi_H^j(x_H^i) = \delta_{ij}$ ($j = 1, \dots, N$, $i = 0, \dots, N + 1$) and $\varphi_h^j(x_h^i) = \delta_{ij}$ ($j = 1, \dots, M$, $i = 0, \dots, M + 1$). We call $V_H = \text{span}\{\varphi_H^j\}_{j=1}^N$ and $V_h = \text{span}\{\varphi_h^j\}_{j=1}^M$. Figure 2.1 illustrates the situation for $M = 5$. Plain lines represent the graph of the coarse finite element functions, dotted lines represent the graph of the fine functions.

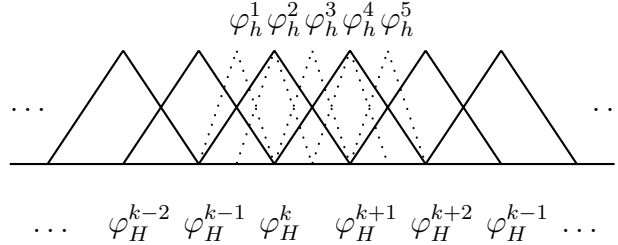


Figure 2.1: Nested grids in 1D.

We introduce $V_{Hh0} = V_H \cap V_h$, i.e.

$$V_{Hh0} = \text{span}\{\varphi_H^j \text{ with } j \text{ s.t. } \exists i, 1 \leq i \leq M, \text{ s.t. } \varphi_h^i(x_H^j) = 1\}. \quad (2.1)$$

In our example, we have $V_{Hh0} = \{\varphi_H^k, \varphi_H^{k+1}\}$. We call $V_{Hh} = V_H + V_h$ and from the above we deduce that

$$\dim V_{Hh} = N + M - \dim V_{Hh0}. \quad (2.2)$$

We introduce the orthogonal complement of V_{Hh0} in V_H and V_h . We label $V_{Hh0}^{\perp H}$ resp. $V_{Hh0}^{\perp h}$ the spaces $V_H \cap V_{Hh0}^{\perp}$ and $V_h \cap V_{Hh0}^{\perp}$. They can be written

$$V_{Hh0}^{\perp H} = \text{span}\{\varphi_H^j \text{ s.t. } \varphi_H^j \perp V_{Hh0}\} \oplus \text{span}\{\chi_H^l\}_{l=1}^{\dim V_{Hh0}} \equiv V_{Hh0}^{\perp H1} \oplus V_{Hh0}^{\perp H2}, \quad (2.3)$$

where the χ_H^l are linear combinations of the $\varphi_H^j \notin V_{Hh0}$ such that $\chi_H^l \perp V_{Hh0}$, and

$$V_{Hh0}^{\perp h} = \text{span}\{\varphi_h^j \text{ s.t. } \nexists i, 1 \leq i \leq N, \text{ s.t. } \varphi_H^i(x_h^j) = 1\}. \quad (2.4)$$

We have $V_{Hh0}^{\perp} = V_{Hh0}^{\perp H} \oplus V_{Hh0}^{\perp h}$. We note that

$$a(v_H, v_h) = 0, \quad \forall v_H \in V_{Hh0}^{\perp H}, \forall v_h \in V_{Hh0}^{\perp h}. \quad (2.5)$$

In fact $v_H = v_H^1 + v_H^2$, with $v_H^1 \in V_{Hh0}^{\perp H1}$ and $v_H^2 \in V_{Hh0}^{\perp H2}$, and we have the result (2.5) from $\text{supp}(v_H^1) \cap \text{supp}(v_h) = \emptyset$ and $v_H^2|_{\text{supp}(v_h)} = \text{constant}$. Hence with the definition (1.107) of $\tilde{\gamma}$, we have

$$\tilde{\gamma} = \sup_{\substack{v_h \in V_{Hh0}^{\perp h}, v_h \neq 0 \\ v_H \in V_{Hh0}^{\perp H}, v_H \neq 0}} \frac{a(v_h, v_H)}{\|v_h\| \|v_H\|} = 0. \quad (2.6)$$

Note that in this case $\gamma = 1$.

Since $\tilde{\gamma} = 0$, Algorithm 1.3 converges in only one iteration. Note that this can be investigated in the present case in another way. First we recall the following known result using the above notation:

Lemma 2.1. *Consider $f \in L^2(0; 1)$ and let $u \in H_0^1(0; 1)$ be such that $\int_0^1 u'v' dx = \int_0^1 fv dx$, $\forall v \in H_0^1(0; 1)$. Let $u_H \in V_H$ be such that $\int_0^1 u_H'v_H' dx = \int_0^1 fv_H dx$, $\forall v_H \in V_H$. Then $u(x_H^j) = u_H(x_H^j)$, $\forall j = 0, 1, 2, \dots, N+1$.*

Proof. Let $G(x, y) = (1-x)y$, $0 \leq y \leq x \leq 1$, $G(x, y) = x(1-y)$, $0 \leq x < y \leq 1$, denote Green's kernel of the problem. Then we have that $G(\cdot, x_H^j) \in V_H$, $\forall j = 0, \dots, N+1$. Since $u(x) = \int_0^1 f(y)G(x, y) dy$, we have $u(x_H^j) = \int_0^1 f(y)G(x_H^j, y) dy = \int_0^1 f(y)G(y, x_H^j) dy$, $\forall j$, and it suffices to take $\{G(\cdot, x_H^j)\}_{j=1}^N$ as a basis for V_H to conclude the proof. \square

Lemma 2.1 states that the approximation u_H on the coarse grid is exact on the nodes of \mathcal{T}_H . Algorithm 1.3 uses this approximation as initial condition. Hence, if the coarse nodes of \mathcal{T}_H in Λ coincide with fine grid points of \mathcal{T}_h (nested grids, see Figure 2.1), at the first half-step of the algorithm the residual is zero on the coarse nodes in Λ . Applying now Lemma 2.1 to the computation of the correction on \mathcal{T}_h , we conclude that adding this correction ($\omega = 1$) to the initial coarse solution yields the discrete solution in V_{Hh} , i.e. the exact solution on the coarse and fine nodes. Thus we conclude that the algorithm gives the solution in only one iteration in the case of nested grids in 1D.

Examine now briefly the case where \mathcal{T}_h and \mathcal{T}_H are not nested but such that the extremities of the interval Λ are nodes of \mathcal{T}_H . This is the situation for example when $M+1$ is an odd number and all nodes are equidistant, see Figure 2.2 for an illustration with $M = 4$. Then $V_{Hh0} = \{0\}$, $\dim V_{Hh} = N+M$ and $\tilde{\gamma} = \gamma \neq 0$.

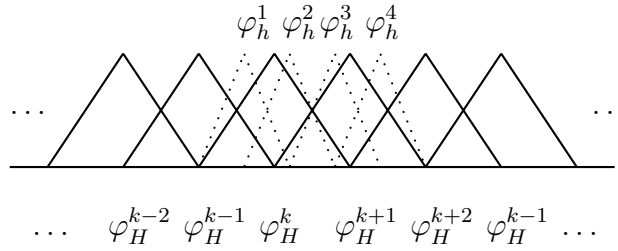


Figure 2.2: Boundary conforming non-nested grids in 1D.

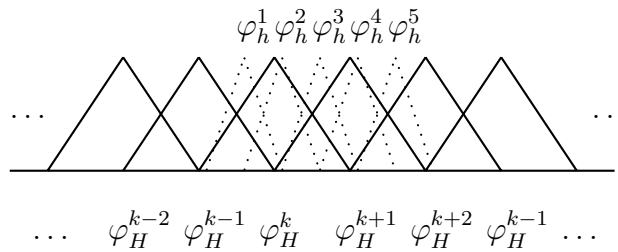


Figure 2.3: Non-nested grids in 1D.

Finally, we analyze the non-nested case obtained by a translation of $\epsilon \ll h$ of the patch Λ , starting from the nested situation of the first case. The situation is illustrated in Figure 2.3.

We compare the present non-nested case where $\dim V_{Hh} = N + M$ with the nested case (2.2). The translation to the non-nested case augments the dimension of V_{Hh} .

We call H and h the distance between two consecutive nodes of the coarse resp. fine triangulation. A straightforward calculation gives for any elements $1 \leq i \leq N, 1 \leq j \leq M$, $\|\varphi_H^j\| = \sqrt{2/H}$, $\|\varphi_h^i\| = \sqrt{2/h}$ and $a(\varphi_H^j, \varphi_h^i) \geq 2(h - \epsilon)/Hh$. Hence $\tilde{\gamma} \geq (h - \epsilon)/\sqrt{Hh}$.

Remark that for $\epsilon \rightarrow 0$, we obtain $\tilde{\gamma} \geq \sqrt{h/H} \neq 0$. This shows that $\tilde{\gamma}(\epsilon)$ is discontinuous. This discontinuity stems from the change of dimension of V_{Hh} once the spaces V_H and V_h become non-nested.

The two- or three-dimensional case.

Let us now go over to the analysis of some properties in d dimensions ($d = 2, 3$). We start with a case where the patch is entirely included in one element of the coarse grid. This situation is most relevant when analyzing two-scale problems. Here the dimension of the patch is at the scale of the grid-size of the coarse grid. Furthermore, large variations of the coefficients a_{ij} of an elliptic operator, as defined in equation (2.8), are of importance and give rise to multi-scale situations. The latter situation will also be discussed as an

implementation issue in Section 2.3.

Let $a_{ij} \in W^{1,\infty}(\Omega)$, $1 \leq i, j \leq d$, verifying $a_{ij} = a_{ji}$ and the hypothesis of strong ellipticity,

$$\sum_{i,j=1}^d a_{ij}(x) \xi_i \xi_j \geq \alpha \sum_{i=1}^d \xi_i^2, \quad \forall (\xi_1, \xi_2) \in \mathbb{R}^d, \text{ a.e. in } \Omega, \quad (2.7)$$

where α is a positive constant. If \mathcal{L} is the elliptic operator given by

$$\mathcal{L}(u) = - \sum_{i,j=1}^d \frac{\partial}{\partial x_i} \left(a_{ij} \frac{\partial u}{\partial x_j} \right), \quad (2.8)$$

the associated bilinear form is given by

$$a(u, v) = \sum_{i,j=1}^d \int_{\Omega} a_{ij} \frac{\partial u}{\partial x_j} \frac{\partial v}{\partial x_i} dx. \quad (2.9)$$

We consider the case when $\bar{\Lambda} \subset K$, for $K \in \mathcal{T}_H$. Let $\tilde{\Lambda} \supseteq \Lambda$ be a rectangle or parallelepiped with dimensions L_i , $1 \leq i \leq d$, and define

$$\tilde{\lambda} = \min_{v \in H_0^1(\tilde{\Lambda}), v \neq 0} \|\nabla v\|_{L^2(\tilde{\Lambda})}^2 / \|v\|_{L^2(\tilde{\Lambda})}^2. \quad (2.10)$$

We have $\tilde{\lambda} = \pi^2 \sum_{i=1}^d 1/L_i^2$ and we introduce $\delta = \sqrt{1/\tilde{\lambda}}$. We set

$$\beta = \left[\sum_{j=1}^d \left(\sum_{i=1}^d \left\| \frac{\partial a_{ij}}{\partial x_i} \right\|_{L^\infty(\Lambda)} \right)^2 \right]^{\frac{1}{2}}. \quad (2.11)$$

Proposition 2.2. *If (2.7) is satisfied and if there exists $K \in \mathcal{T}_H$ such that $\bar{\Lambda} \subset K$ and if $r = 1$, then $\gamma \leq \frac{\beta\delta}{\alpha}$. If furthermore the a_{ij} 's are constant over Λ , $1 \leq i, j \leq d$, Algorithm 1.3 converges in only one iteration when $\omega = 1$.*

Proof. We shall first prove that $\gamma \leq \frac{\beta\delta}{\alpha}$. For any $u_H \in V_H, v_h \in V_h$, we have

$$|a(u_H, v_h)| = \left| \sum_{i,j=1}^d \int_{\Lambda} a_{ij} \frac{\partial u_H}{\partial x_j} \frac{\partial v_h}{\partial x_i} dx \right|, \quad (2.12)$$

as $v_h = 0$ in $\Omega \setminus \bar{\Lambda}$. Since $\bar{\Lambda} \subset K \in \mathcal{T}_H$, $\frac{\partial u_H}{\partial x_j}$ is constant over $\bar{\Lambda}$ so that

$$|a(u_H, v_h)| = \left| \sum_{i,j=1}^d \frac{\partial u_H}{\partial x_j} \Big|_K \int_{\Lambda} \frac{\partial a_{ij}}{\partial x_i} v_h dx \right|, \quad (2.13)$$

where we have applied the divergence theorem taking into account that $v_h = 0$ on $\partial\Lambda$. By the Cauchy-Schwarz inequality we have

$$|a(u_H, v_h)| \leq \sum_{i,j=1}^d \left\| \frac{\partial a_{ij}}{\partial x_i} \right\|_{L^\infty(\Lambda)} \left| \frac{\partial u_H}{\partial x_j} \right|_K \int_{\Lambda} |v_h| \, dx \quad (2.14)$$

$$\leq \beta \left(\sum_{j=1}^d \left\| \frac{\partial u_H}{\partial x_j} \right\|_{L^2(\Lambda)}^2 \|v_h\|_{L^2(\Lambda)}^2 \right)^{\frac{1}{2}} \quad (2.15)$$

$$= \beta \|\nabla u_H\|_{L^2(\Lambda)} \|v_h\|_{L^2(\Lambda)}. \quad (2.16)$$

At this point we need to bound $\|v_h\|_{L^2(\Lambda)}$ from above with $\|\nabla v_h\|_{L^2(\Lambda)}$. We introduce $\lambda = \min_{v \in H_0^1(\Lambda), v \neq 0} \|\nabla v\|_{L^2(\Lambda)}^2 / \|v\|_{L^2(\Lambda)}^2$, the smallest value of the Rayleigh quotient for the Laplacian operator on Λ . In order to estimate λ , we consider the rectangle or parallelepiped $\tilde{\Lambda}$ and $\tilde{\lambda}$ as introduced above. As $\Lambda \subseteq \tilde{\Lambda}$ we have $\lambda \geq \tilde{\lambda} = 1/\delta^2$, i.e. we get $\|v_h\|_{L^2(\Lambda)} \leq \delta \|\nabla v_h\|_{L^2(\Lambda)}$. Hence combining the previous results,

$$|a(u_H, v_h)| \leq \beta \delta \|\nabla u_H\|_{L^2(\Lambda)} \|\nabla v_h\|_{L^2(\Lambda)}. \quad (2.17)$$

The hypothesis of strong ellipticity (2.7) implies that, $\forall u \in H_0^1(\Omega)$,

$$a(u, u) = \int_{\Omega} \sum_{i,j=1}^d a_{ij} \frac{\partial u}{\partial x_j} \frac{\partial u}{\partial x_i} \, dx \geq \alpha \|\nabla u\|_{L^2(\Omega)}^2, \quad (2.18)$$

i.e. $\alpha \|\nabla u\|_{L^2(\Lambda)}^2 \leq \alpha \|\nabla u\|_{L^2(\Omega)}^2 \leq a(u, u) = \|u\|^2$. Applying this inequality to u_H and v_h , we obtain $|a(u_H, v_h)| \leq \frac{\beta \delta}{\alpha} \|u_H\| \|v_h\|$, i.e. $\gamma \leq \frac{\beta \delta}{\alpha}$.

If the a_{ij} 's are constant over Λ , $1 \leq i, j \leq d$, we clearly have $\beta = 0$, thus $\gamma = 0$. Hence the constant C_0 introduced in Hypothesis 1.5 (equation (1.43)) is $C_0 = 1$. Furthermore, in this case V_H and V_h are orthogonal (Properties 1.4(iii)) and, since the iteration operator $B = (I - \omega P_H)(I - \omega P_h)$ introduced in (1.37) yields $B = 0$ for $\omega = 1$, the Algorithm 1.3 converges in only one iteration. \square

In two dimensions, when $\bar{\Lambda} \subset K_1 \cup K_2$, with $K_1, K_2 \in \mathcal{T}_H$, the analysis gets more involved. In the sequel we present some upper bounds for γ in the case where $a_{ij} = \delta_{ij}$, i.e. $a(u, v) = \int_{\Omega} \nabla u \cdot \nabla v \, dx$, and with Λ the union of two triangles K_1 and K_2 of \mathcal{T}_H , \mathcal{T}_h being a refinement of \mathcal{T}_H and $r = s = 1$. We consider the situations illustrated in Figure 2.4 by the triangulations of the patch Λ .

Estimates can be obtained by splitting $v \in V_{Hh}$ into $v = v_h + v_H$, where $v_H = r_H v$ is the interpolant of v in V_H and $v_h = v - r_H v \in V_h$. The nodes in $\bar{\Lambda}$ corresponding to the underlying finite element functions on which v_h and v_H are based on are depicted in Figure 2.4. Using the fact that $v_h = 0$ in $\Omega \setminus \bar{\Lambda}$ and the divergence theorem, we have that

$$a(v_H, v_h) \leq \left| \left[\frac{\partial v_H}{\partial n} \right]_{\Gamma} \right| \int_{\Gamma} |v_h| \, ds, \quad (2.19)$$

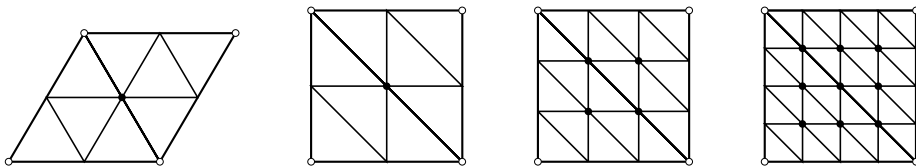


Figure 2.4: Illustration of the triangulations of Λ considered in Table 2.1. White dots refer to the degrees of freedom of $r_H v$, black dots refer to those of $v - r_H v$.

where $\Gamma = \partial K_1 \cap \partial K_2$, $[\cdot]_\Gamma$ denotes the jump on Γ in the direction of a normal unit vector \mathbf{n} on Γ . We have for on K_i , $i = 1, 2$,

$$\frac{\partial v_H}{\partial n} = \nabla v_H \cdot \mathbf{n} \leq |\nabla v_H| = \frac{\int_{K_i} |\nabla v_H| dx}{\text{area}(K_i)} \leq \frac{\sqrt{\text{area}(K_i)} \|\nabla v_H\|_{L^2(K_i)}}{\text{area}(K_i)}, \quad (2.20)$$

and we consider the non-optimal bound

$$\left| \left[\frac{\partial u_H}{\partial n} \right]_\Gamma \right| \leq \left| \frac{\partial u_H}{\partial n_1} \right| + \left| \frac{\partial u_H}{\partial n_2} \right|, \quad (2.21)$$

where n_i , $i = 1, 2$, denotes the normal direction outward of K_i on Λ . Hence the first factor of the right-hand side of (2.19) can be bounded by

$$\left| \left[\frac{\partial v_H}{\partial n} \right]_\Gamma \right| \leq \sum_{i=1}^2 \frac{\|\nabla v_H\|_{L^2(K_i)}}{\sqrt{\text{area}(K_i)}} \leq \frac{\sqrt{2}}{\min_{i=1,2} \sqrt{\text{area}(K_i)}} \|\nabla v_H\|_{L^2(\Lambda)}, \quad (2.22)$$

and $\|\nabla v_H\|_{L^2(\Lambda)} \leq \|\nabla v_H\|_{L^2(\Omega)}$. As the dimension of V_h is small in our cases, we evaluate $\int_\Gamma |v_h| ds$ explicitly and do the same with $\|\nabla v_h\|_{L^2(\Lambda)}$. Hence we can express $\int_\Gamma |v_h| ds$, i.e. the second factor of the right-hand side of (2.19) in relation to $\|\nabla v_h\|_{L^2(\Lambda)}$ which is equal to $\|\nabla v_h\|_{L^2(\Omega)}$. This way get a non-optimal estimate of the constant γ (refer to Chapter 1 for its definition in (1.39)) from (2.19): $a(v_H, v_h) \leq C \|\nabla v_h\|_{L^2(\Omega)} \|\nabla v_H\|_{L^2(\Omega)}$, where $\gamma \leq C$.

Thus, applying the above procedure to our situations (Figure 2.4), we get $a(v_H, v_h) \leq C \|v_H\| \|v_h\|$ (see equation 1.40) and hence we have $\gamma \leq C$. The upper bounds found for γ are reported in Table 2.1. Note that the bound for γ on right isosceles triangles with $H/h = 2$ is reported by Axelsson and Gustafsson in [9].

For the cases with right isosceles triangles presented in Figure 2.4 we estimate $\gamma = \tilde{\gamma}$ numerically using the method described later in this chapter (see Section 2.2). These results, reported in Table 2.2, are to be compared with the unsharp estimated upper bounds presented in Table 2.1.

2.2 Numerical evaluation of $\tilde{\gamma}$ and optimal relaxation

In this section we firstly present a method to numerically evaluate $\tilde{\gamma}$ for given discretization spaces. This produces then with (1.78) a good approximation for the optimal relax-

Triangles	H/h	Upper bound for γ
equilateral	2	$\sqrt{3}/3 \approx 0.577$
right isosceles	2	$\sqrt{2}/2 \approx 0.707$
right isosceles	3	$2/3 \approx 0.667$
right isosceles	4	$\sqrt{2}/2 \approx 0.707$

Table 2.1: Upper bounds for γ .

Triangles	H/h	Numerical estimate for γ
right isosceles	2	0.426
right isosceles	3	0.464
right isosceles	4	0.476

Table 2.2: Numerical estimates for γ .

ation parameter ω in Algorithm 1.3.

A crucial question for running the algorithm is to know how to choose the relaxation parameter ω . We refer to Figure 2.7 where we compare the algorithm convergence for $\omega = 1$ and $\omega = \omega^{\text{opt}}$. In fact, the spectral radius ρ of the iteration operator giving the speed of convergence strongly depends on the relaxation of the method. Since Proposition 1.9, the spectral radius is given by the algebraic relationship (1.77): ρ is function of $\tilde{\gamma}$ and ω . An illustration of the functional relation $\rho(\omega)$ for given $\tilde{\gamma}$ is depicted in Figure 1.7. Furthermore, equation (1.78) establishes a formula for calculating the optimal relaxation parameter once $\tilde{\gamma}$ is known. Hence, a good approximation of the parameter $\tilde{\gamma}$ is the key for an estimate of the optimal relaxation parameter ω^{opt} .

The result of Proposition 1.9 with (1.77) gives an algebraic relationship for the spectral radius ρ of the operator B as a function of $\tilde{\gamma}$ and ω . This leads to a very convenient application to determine numerically a good approximation for $\tilde{\gamma}$. Since (1.77) with $\omega = 1$ (see also the first observation after Proposition 1.9), we have the relation $\tilde{\gamma}^2 = \rho$.

Running Algorithm 1.3 with zero right-hand side and for given $\omega = 1$ or $\omega \neq 1$, our objective is to evaluate numerically an estimate of ρ . Hence we find an estimate of the parameter $\tilde{\gamma}$, either directly with $\tilde{\gamma}^2 = \rho$ or through (1.77).

If the largest eigenvalue of the iteration operator B given by (1.37) is real we can obtain its spectral radius via the known result $\rho(B) = \lim_{n \rightarrow \infty} \|Bu^n\|/\|u^n\|$ for $f = 0$. But in general we cannot assume that the largest eigenvalue is real. We do not use the standard power method as it does not apply most generally, and in particular when $\rho(B)$

corresponds, e.g., to complex conjugated eigenvalues.

Hence we need Proposition 2.3. Consider $\mathcal{B} \in \mathcal{L}(U)$ a linear operator over a finite dimensional complex normed vector space U with a set of N eigenvalues $\lambda_i \in \mathbb{C}$, $i = 1, \dots, N$, such that

$$|\lambda_1| = \dots = |\lambda_k| > |\lambda_{k+1}| \geq \dots \geq |\lambda_N|, \quad 1 \leq k < N. \quad (2.23)$$

Let v_i , $i = 1, \dots, k$, resp. w_i , $i = k + 1, \dots, N$, denote the generalized eigenvectors associated to $\lambda_1, \dots, \lambda_k$ resp. $\lambda_{k+1}, \dots, \lambda_N$. Call $V = \text{span}(v_1, \dots, v_k)$ and $W = \text{span}(w_{k+1}, \dots, w_N)$. We have $U = V \oplus W$.

Proposition 2.3. *Let $\mathcal{B} \in \mathcal{L}(U)$ be such that its eigenvalues verify (2.23). For any $u = v + w \in U$ with $v \in V$, $w \in W$, such that $v \neq 0$, the spectral radius of \mathcal{B} is given by $\rho(\mathcal{B}) = |\lambda_1| = \lim_{n \rightarrow \infty} \sqrt[n]{\|\mathcal{B}^n u\|}$.*

Proof. We remark that $\rho(\mathcal{B}) = |\lambda_1| = \dots = |\lambda_k|$. If we set $\rho^* = |\lambda_{k+1}| = \max_{i \geq k+1} |\lambda_i|$, we have $\rho^* < \rho(\mathcal{B})$ since (2.23). Let ρ_1 , ρ_2 and ρ_3 be constants such that $\rho^* < \rho_1 < \rho_2 < \rho(\mathcal{B}) < \rho_3$.

We show that there exist constants C_1 , C_2 and C_3 , independent of n , such that

$$|C_2 \rho_2^n - C_1 \rho_1^n| \leq \|\mathcal{B}^n u\| \leq C_3 \rho_3^n. \quad (2.24)$$

In fact, since Yoshida [68, §VIII.2], we have $\rho(\mathcal{B}) = \lim_{n \rightarrow \infty} \sqrt[n]{\|\mathcal{B}^n\|}$, where $\|\cdot\|$ denotes the operator-norm induced by $\|\cdot\|$. Hence, since $\rho(\mathcal{B}) < \rho_3$, there exists a constant \tilde{C}_3 independent of n such that $\|\mathcal{B}^n\| \leq \tilde{C}_3 \rho_3^n$. This implies that $\|\mathcal{B}^n u\| \leq \|\mathcal{B}^n\| \|u\| \leq C_3 \rho_3^n$ for any $u \in U$. Similarly, we have $\|\mathcal{B}^n w\| \leq C_1 \rho_1^n$ for any $w \in W$. Next we show that $\|\mathcal{B}^n v\| \geq C_2 \rho_2^n$ for $v \in V$, $v \neq 0$. For this, we introduce \mathcal{B}_V , the restriction of \mathcal{B} on V . We have $\lim_{n \rightarrow \infty} \sqrt[n]{\|(\mathcal{B}_V^{-1})^n\|} = \rho(\mathcal{B}_V^{-1}) = 1/\rho(\mathcal{B})$ and thus there exists a constant \tilde{C}_2 such that $\|(\mathcal{B}_V^{-1})^n\| \leq \tilde{C}_2 / \rho_2^n$. Furthermore, for any $v \in V$, $v \neq 0$,

$$\|v\| = \|(\mathcal{B}_V^{-1})^n \mathcal{B}_V^n v\| \leq \|(\mathcal{B}_V^{-1})^n\| \|\mathcal{B}_V^n v\| \leq \tilde{C}_2 / \rho_2^n \|\mathcal{B}_V^n v\|, \quad (2.25)$$

i.e. $\|\mathcal{B}_V^n v\| \geq C_2 \rho_2^n$. Finally, we have

$$\|\mathcal{B}^n u\| = \|\mathcal{B}^n v + \mathcal{B}^n w\| \geq \|\mathcal{B}^n v\| - \|\mathcal{B}^n w\| \geq |C_2 \rho_2^n - C_1 \rho_1^n|, \quad (2.26)$$

for any $u = v + w$ with $v \in V$, $w \in W$ and $v \neq 0$.

With (2.24) we are able to conclude. In fact from (2.24) we have $\rho_2 |C_2 - C_1(\rho_1/\rho_2)^n|^{1/n} \leq \sqrt[n]{\|\mathcal{B}^n u\|} \leq C_3^{1/n} \rho_3$. Hence $\limsup_{n \rightarrow \infty} \sqrt[n]{\|\mathcal{B}^n u\|} \leq \rho_3$ and $\liminf_{n \rightarrow \infty} \sqrt[n]{\|\mathcal{B}^n u\|} \geq \rho_2$, for any ρ_2 and ρ_3 such that $\rho_2 < \rho(\mathcal{B}) < \rho_3$. Thus we conclude that $\lim_{n \rightarrow \infty} \sqrt[n]{\|\mathcal{B}^n u\|} = \rho(\mathcal{B})$. \square

Proposition 2.3 can be applied straightforwardly on the iteration operator B when V_{Hh} is considered as a complex Hilbert space. In this case we have

$$\rho(B) = \lim_{n \rightarrow \infty} \sqrt[n]{\|B^n v\|}, \quad (2.27)$$

and consequently, when $\omega = 1$,

$$\tilde{\gamma} = \sqrt{\lim_{n \rightarrow \infty} \sqrt[n]{\|B^n v\|}}. \quad (2.28)$$

For implementation we set $\omega = 1$ and the right-hand side $f \equiv 0$, and perform m steps of the algorithm, starting in practice from any non zero initial condition v^0 , to obtain some $v^m = B^m v^0$. Following (2.28) we use the approximation

$$\rho \approx \sqrt[m]{\|v^m\|}, \quad (2.29)$$

for m large, and obtain with (1.78) and $\rho = \tilde{\gamma}^2$ that

$$\omega^{\text{opt}} = \omega_0(\tilde{\gamma} = \sqrt{\rho}) = \frac{2 - 2\sqrt{1 - \rho}}{\rho}. \quad (2.30)$$

This is the optimal relaxation parameter in the sense that it gives the minimum value for $\rho(B)$ which is most relevant for the speed of convergence. The speed of convergence is asymptotically given by the ratio e^{n+1}/e^n for large n , where e^n is the relative error at iteration n defined in the paragraph here below. The evolution of this error through the iterative process gives information about the speed of convergence of the algorithm (see Figure 2.7).

In general, for running the algorithm, we use the following stopping criteria and errors. First we define the variation of the discrepancy between two iterations and require that $\|u^n - u^{n-1}\|/\|u^n\| < \epsilon_1$ where ϵ_1 is a given tolerance. If this criterion yields true at iteration $n = n_{\text{cvg}}$, we define $u_{Hh} = u^{n_{\text{cvg}}}$. To verify that the algorithm has well converged, we check that u_{Hh} satisfies a second criterion, namely $\|\bar{u}_{Hh} - u_{Hh}\|/\|\bar{u}_{Hh}\| < \epsilon_2$, where $\bar{u}_{Hh} = u^{n_{\text{cvg}}+p}$, $p = 20$. We have chosen $\epsilon_1 = 10^{-4}$ and $\epsilon_2 = 10\epsilon_1$ for the results reported in this section. We define the relative error at iteration n by $e^n = \|\bar{u}_{Hh} - u^n\|/\|\bar{u}_{Hh}\|$.

Up to here we have only introduced the tools to assess the convergence of the algorithm, i.e. obtaining the approximation u_{Hh} to the exact solution u . To assess the convergence of u_{Hh} to u , e.g. in H and h given by the *a priori* estimate of Proposition 1.1, or with regard to the memory usage, we introduce the relative error $e_{Hh} = \|u - u_{Hh}\|/\|u\|$. Results of this are reported in Section 1.2 (Figures 1.3 and 1.4) and below at the end of Section 2.3 (Figures 2.8 and 2.9).

We proceed now to a study of $\tilde{\gamma}$ for various spaces V_h and V_H and introduce for this the following model problem. Consider again the two-dimensional Poisson-Dirichlet

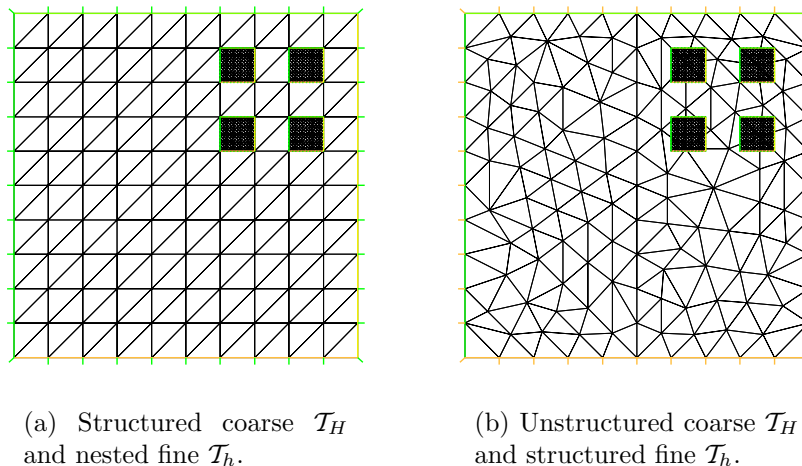
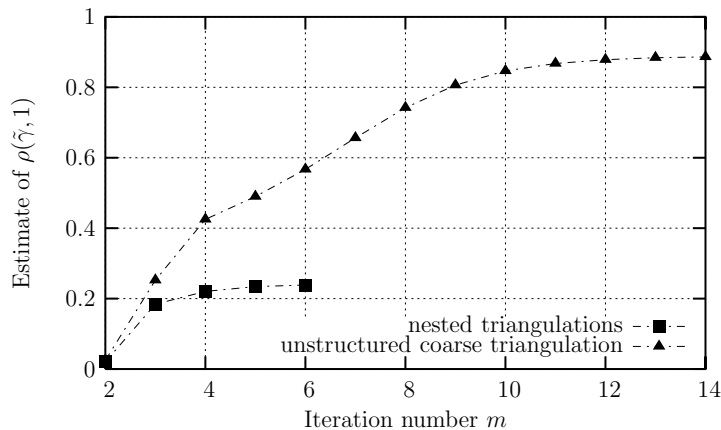


Figure 2.5: Illustration of the considered grid constellations for $N = M = 10$ and ratio $H/h = 10$.

problem (1.17) in $\Omega = (-1; 1)^2$ introduced earlier. Now we take f such that the exact solution to the problem is given by $u = u_0 + \sum_{i=1}^4 u_i$, $u_0(x, y) = \cos(\frac{\pi}{2}x) \cos(\frac{\pi}{2}y)$ and $u_i(x, y) = \eta \chi(R_i) \exp \epsilon_f^{-2} \exp(-1/|\epsilon_f^2 - R_i^2|)$, where $R_i(x, y) = \sqrt{(x - x_i)^2 + (y - y_i)^2}$ and $\chi(R_i) = 1$ if $R_i \leq \epsilon_f$, $\chi(R_i) = 0$ if $R_i > \epsilon_f$; η , ϵ_f and (x_i, y_i) , $i = 1, 2, 3, 4$ are parameters. Hence the right-hand side of (1.17) is given by $f = f_0 + \sum_{i=1}^4 f_i$, where $f_0 = -\Delta u_0$ and $f_i = -\Delta u_i$, $i = 1, 2, 3, 4$. We choose $\eta = 10$, $\epsilon_f = 0.3$ and $(x_1, y_1) = (0.3, 0.3)$, $(x_2, y_2) = (0.7, 0.3)$, $(x_3, y_3) = (0.3, 0.7)$, $(x_4, y_4) = (0.7, 0.7)$.

In the following, we refer to Section 1.1 for the definition of the notation used. For the triangulation of $\bar{\Omega}$, we use a coarse uniform grid with mesh size H and $r = 1$. We consider the patches Λ_i , $i = 1, 2, 3, 4$, with a fine uniform triangulation of size h and $s = 1$. Choose $\Lambda_i = (x_i - \epsilon; x_i + \epsilon) \times (y_i - \epsilon; y_i + \epsilon)$, with $\epsilon = 0.1$. We set $H = 2/N$ and $h = 2\epsilon/M$, N, M being the number of discretization intervals on one side of the squares Ω and Λ_i respectively. Non-nested and nested situations are illustrated in Figure 2.5.

We consider different situations including structured nested and non-nested as well as unstructured grids on the domain Ω . We always use the same structured grids for the patches. Our objective is to assess the convergence of our method with regard to the influence of the grids used. Our goal is to show that the algorithm performs well when $h \rightarrow 0$ for fixed H , and when each patch covers only a small number of coarse elements. It is particularly competitive when used with the optimal relaxation parameter ω^{opt} given by (2.30) in initially ill-conditioned situations like those presented in Table 2.3(c): small displacement of the coarse nodes of \mathcal{T}_H with regard to a nested \mathcal{T}_h .

Figure 2.6: Convergence to ρ over iterations m .

We first illustrate that obtaining an estimate of the optimal relaxation parameter is fast. Only a small number m of iterations are necessary to get the approximation of ρ through equation (2.29). In Figure 2.6 we plot the estimate of ρ obtained for increasing m when $N = M = 20$. We conclude that a couple of iterations are sufficient to get a good estimate and hence the optimal relaxation parameter ω^{opt} .

Recall that first estimates for the spectral radius have already been reported through the situation considered in Section 1.6 in Figure 1.12 to illustrate equation (1.77) and verify the algebraic fitting to the numerical results. Further numerical estimates for the parameter γ corresponding to the situations depicted in Figure 2.4 with right isosceles triangles are reported in the Table 2.2.

Let us study now the algorithm on the above introduced situation.

Main results are reported in the following table (Table 2.3). In each part we depict the considered situation by small graphics showing first the whole triangulation \mathcal{T}_H with the patches, then a zoom to emphasize the region around one corner of a patch to show how \mathcal{T}_h and \mathcal{T}_H are related. First we set $\omega = 1$ and run our method to obtain an estimate of $\tilde{\gamma}$ through (2.28) and hence of the spectral radius $\rho = \tilde{\gamma}^2$ of the iteration operator. Then we run the algorithm on the problem till convergence (see the text above for the stopping criteria used) and report the number of iterations n_{cvg} . These values are respectively reported in the first rows of Tables 2.3(a)–2.3(c). Given the approximation for $\tilde{\gamma}$ we determine the optimal relaxation parameter ω^{opt} with (1.78) and give the spectral radius. The last line in the tables reports the required iterations needed by the method to converge under optimal relaxation.

In a first test, we choose N and M such that the ratio H/h is of magnitude 10. In

these first cases, the patches cover a small number of triangles of \mathcal{T}_H leading to small coefficients $\tilde{\gamma}$ and ρ . Hence convergence is reached after a small number of iterations.

When doubling the number of fine triangles, see Table 2.3(b), the situation remains similar. A slight over-relaxation realizes a gain of a couple of iterations. This suggests that the method is efficient in multi-scale situations where the size of the applied patch is of the order of a coarse element.

In the examples of Table 2.3(c) we increase the precision of the coarse triangulation. These cases show that the algorithm is best-suited to situations with patches covering a small number of coarse triangles. Here again, this shows that the method is efficient when the size (i.e., the diameter or the length of a side) of the patch is of the order of the coarse grid size H . In fact, increasing the number of coarse triangles covered by the patches, i.e. increasing the size of the patches, leads to bad condition numbers (ρ close to 1). Nevertheless optimal relaxation allows to divide by a factor two the number of iterations necessary to obtain convergence. This shows that optimal relaxation is a key ingredient in our method.

These basic results show that the method is very well adapted for multi-scale situations when applying small patches $|\Lambda| \ll |\Omega|$ in the regions with sharp data.

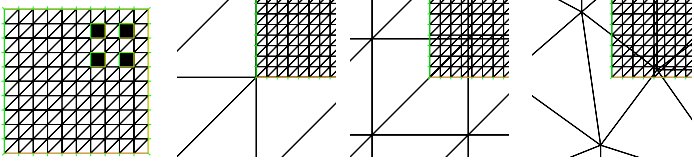
In Table 2.3 we have seen that optimizing the relaxation optimizes the convergence of the algorithm, i.e. the rate of the error reduction through the iterations. In Figure 2.7 we plot the evolution of the a -norm error of u^n to u_{Hh} for the nested and the unstructured cases of Table 2.3(c). Note that the values of ρ give an upper bound for the slope of the error reduction (see Figure 2.7).

We refer the reader to [36] for more examples of \mathcal{T}_H and \mathcal{T}_h . In [36, Section 6], we study $\tilde{\gamma}$ and the convergence to u_{Hh} through the algorithm iterations in various situations (see Table 3 and Figure 5 therein).

2.3 Implementation issues and memory and CPU-time usage

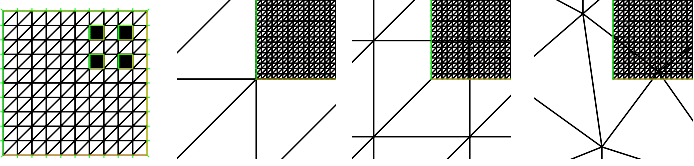
In this section we describe how we have implemented the method and what critical points need special care with regard to the usage of memory, to integration and the grids used. Finally we illustrate the efficiency of the method in particular versus memory and CPU-time usage.

We start with discussing practical aspects to construct an efficient computer program



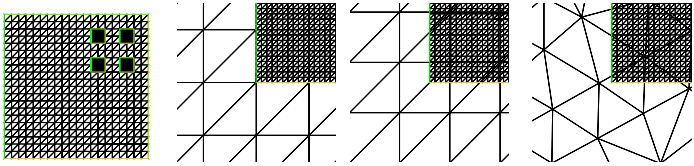
$H/h = 10$	nested $N = M = 10$	non-nested $N = 11, M = 10$	unstructured $N = M = 10$
$\rho(\tilde{\gamma}, 1) = \tilde{\gamma}^2$	0.28	0.30	0.34
n_{cvg}	6	8	9
ω^{opt}	1.08	1.09	1.10
$\rho(\tilde{\gamma}, \omega^{\text{opt}})$	0.08	0.09	0.10
n_{cvg}	5	6	9

 (a) Algorithm properties for $H/h = 10$, $N = 10$.



$H/h = 20$	nested $N = 10, M = 20$	non-nested $N = 11, M = 20$	unstructured $N = 10, M = 20$
$\rho(\tilde{\gamma}, 1) = \tilde{\gamma}^2$	0.28	0.31	0.38
n_{cvg}	6	8	9
ω^{opt}	1.08	1.09	1.12
$\rho(\tilde{\gamma}, \omega^{\text{opt}})$	0.08	0.09	0.12
n_{cvg}	5	6	6

 (b) Algorithm properties for $H/h = 20$, $N = 10$.



$H/h = 10$	nested $N = M = 20$	non-nested $N = 21, M = 20$	unstructured $N = M = 20$
$\rho(\tilde{\gamma}, 1) = \tilde{\gamma}^2$	0.24	0.89	0.91
n_{cvg}	6	24	27
ω^{opt}	1.07	1.50	1.54
$\rho(\tilde{\gamma}, \omega^{\text{opt}})$	0.07	0.50	0.54
n_{cvg}	5	13	15

 (c) Algorithm properties for $H/h = 20$, $N = 20$.

Table 2.3: Comparison of the algorithm properties.

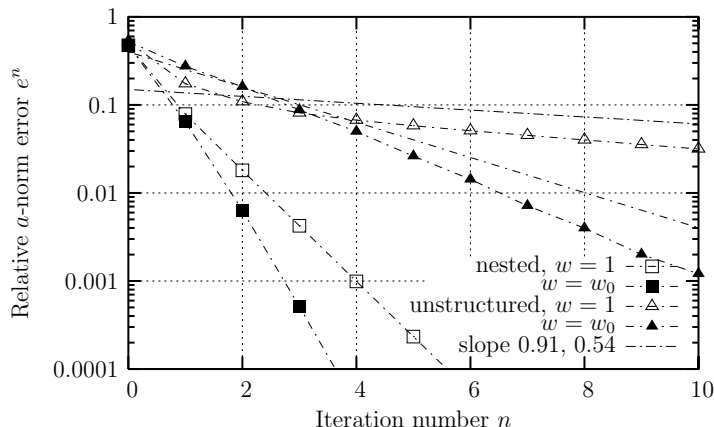


Figure 2.7: Convergence of u^n to u_{Hh} with respect to the iteration number for cases of Table 2.3(c). Comparison of the convergence for the non-relaxed ($\omega = 1$) and the optimally relaxed ($\omega = \omega_0 = \omega^{\text{opt}}$) method.

for implementing Algorithm 1.3. All results reported have been obtained using a basic implementation with the software **Freefem++** [43].

Handling two domains with a priori non-conforming triangulations raises a couple of practical issues. At any stage the coarse and the fine parts of the solution u^n are stored separately, that is to say $u^{n-1} = u_H^{n-1} + u_h^{n-1}$ with $u_H^{n-1} \in V_H$, $u_h^{n-1} \in V_h$. We write the first step of the n -th iteration of the algorithm as follows:

$$\begin{aligned} &\text{Find } v_h \in V_h \text{ s.t. } a(v_h, \varphi) = \langle f | \varphi \rangle - a(u_H^{n-1}, \varphi), \forall \varphi \in V_h . \\ &\text{Set } u_h^n = (1 - \omega)u_h^{n-1} + \omega v_h. \end{aligned}$$

The same holds for the second step which writes out explicitly:

$$\begin{aligned} &\text{Find } v_H \in V_H \text{ s.t. } a(v_H, \varphi) = \langle f | \varphi \rangle - a(u_h^n, \varphi), \forall \varphi \in V_H. \\ &\text{Set } u_H^n = (1 - \omega)u_H^{n-1} + \omega v_H. \end{aligned}$$

Comparing this formulation with the initial writing of Algorithm 1.3 we have $v_h = w_h + u_h^{n-1}$ and $v_H = w_H + u_H^{n-1}$. Hence the respective right-hand sides are shorter by one term.

Note here that the above formulation makes immediate the comparison with the Chimera method [25] made already in Section 1.3. When $\omega = 1$, we are left with the following two steps:

- (i) Find $u_h^n \in V_h$ s.t. $a(u_h^n, \varphi) = \langle f | \varphi \rangle - a(u_H^{n-1}, \varphi), \forall \varphi \in V_h$;
- (ii) find $u_H^n \in V_H$ s.t. $a(u_H^n, \varphi) = \langle f | \varphi \rangle - a(u_h^n, \varphi), \forall \varphi \in V_H$.

Recall that the multiplicative character of our method versus the Chimera method is immediately seen in the second step when using the updated fine grid solution u_h^n in the right-hand side.

At this point we need to discuss the numerical integration and restrict ourselves to linear finite elements ($r = s = 1$).

Two difficulties are to be taken into account whether sharp data, i.e. data needing fine integration, of the problem comes from the right-hand side f or originates from the form a . In the first case the evaluation of $\langle f|\varphi \rangle$ needs particular attention. In the second case scalar products evaluated on the coarse grid must be considered with care. Another issue is the treatment of mixed term scalar products wherein finite element functions of both V_H and V_h appear.

In the sequel, we consider these problems and illustrate our proposals with the scalar product given by (1.5). The evaluation of the different terms appearing in the algorithm is conforming to the following guidelines:

- If the coefficients a_{ij} defining the scalar product a are “smooth” in Λ and in Ω , the homogeneous terms $a(\varphi_H, \psi_H)$ with $\varphi_H, \psi_H \in V_H$, and $a(\varphi_h, \psi_h)$ with $\varphi_h, \psi_h \in V_h$, of support in Ω resp. Λ are integrated using the grid \mathcal{T}_H on Ω resp. \mathcal{T}_h in Λ . Numerical integration in 2D is done with the standard three-point formula (in 3D we use a four-point formula). In the case of (1.5) this writes out, $\forall \varphi_H, \psi_H \in V_H$,

$$a(\varphi_H, \psi_H) \approx \sum_{K \in \mathcal{T}_H} \frac{|K|}{d+1} \sum_{\alpha=1}^{d+1} \sum_{i,j=1}^d a_{ij}(\mathbf{x}_K^\alpha) \frac{\partial \varphi_H}{\partial x_j} \Big|_K \frac{\partial \psi_H}{\partial x_i} \Big|_K, \quad (2.31)$$

where $|K|$ denotes the area or volume, and \mathbf{x}_K^α , $\alpha = 1, \dots, d+1$, the vertices of the element K . We use the same formula for $a(\varphi_h, \psi_h)$ where $\varphi_h, \psi_h \in V_h$ with $K \in \mathcal{T}_h$ in (2.31).

The mixed term $a(\varphi_h, \psi_H)$, $\varphi_h \in V_h, \psi_H \in V_H$, of support in Λ , is approximated by $a(\varphi_h, r_h \psi_H)$, i.e.

$$a(\varphi_h, \psi_H) \approx \sum_{K \in \mathcal{T}_h} \frac{|K|}{d+1} \sum_{\alpha=1}^{d+1} \sum_{i,j=1}^d a_{ij}(\mathbf{x}_K^\alpha) \frac{\partial \varphi_h}{\partial x_j} \Big|_K \frac{\partial (r_h \psi_H)}{\partial x_i} \Big|_K, \quad (2.32)$$

where r_h is the standard interpolant to the space V_h . When implementing, our technique consists in introducing, besides \mathcal{T}_H and \mathcal{T}_h , a transmission grid, i.e. a fine structured grid considered over the patch Λ . This enables handling of the elements of V_H and V_h in the mixed scalar products. The transmission grid helps associating fine and coarse triangles and vertices of the grids \mathcal{T}_H and \mathcal{T}_h and gives information on, e.g., which nodes of \mathcal{T}_h are in a given triangle of \mathcal{T}_H .

- If the coefficients a_{ij} are sharp in Λ , the above approximations are sufficient. In fact all a -products appearing in a right-hand side of the algorithm are integrated on the

fine grid \mathcal{T}_h (see (2.32)). Furthermore, as our algorithm is a correction algorithm with corrections tending to zero, the left-hand side $a(v_H, \varphi)$, $\varphi \in V_H$, in the second step, is not to be rewritten.

- The term $\langle f|\varphi \rangle$, $\varphi \in V_h$ or V_H , is approximated with

$$\begin{aligned} \langle f|\varphi_H \rangle &\approx \sum_{K \in \mathcal{T}_H} \frac{|K|}{d+1} \sum_{\alpha=1}^{d+1} f^1(\mathbf{x}_K^\alpha) \varphi_H(\mathbf{x}_K^\alpha) \\ &\quad + \sum_{K \in \mathcal{T}_h} \frac{|K|}{d+1} \sum_{\alpha=1}^{d+1} f^2(\mathbf{x}_K^\alpha) (r_h \varphi_H)(\mathbf{x}_K^\alpha), \quad \forall \varphi_H \in V_H, \end{aligned} \quad (2.33)$$

and

$$\langle f|\varphi_h \rangle \approx \sum_{K \in \mathcal{T}_h} \frac{|K|}{d+1} \sum_{\alpha=1}^{d+1} f^2(\mathbf{x}_K^\alpha) \varphi_h(\mathbf{x}_K^\alpha), \quad \forall \varphi_h \in V_h, \quad (2.34)$$

where $f = f^1 + f^2$ with $f^1 = \begin{cases} f & \text{in } \Omega \setminus \Lambda \\ 0 & \text{in } \Lambda \end{cases}$, and $f^2 = \begin{cases} 0 & \text{in } \Omega \setminus \Lambda \\ f & \text{in } \Lambda \end{cases}$.

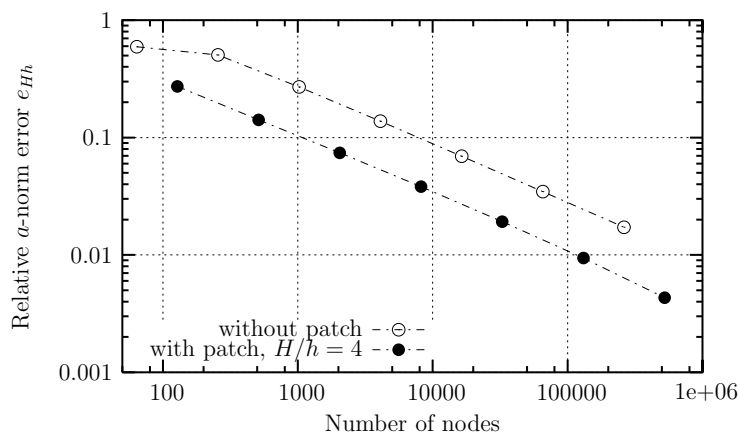
Finally, let us assess the efficiency of our method with respect to memory and CPU-time usage. Using a well-situated patch reduces the number of nodes necessary to obtain a given accuracy on the solution. Recalling problem (1.17) defined in Section 1.2 and the nested grid constellation introduced in Figure 1.3(a), we illustrate our point in Figures 2.8 and 2.9.

Using the convergence results reported in Figure 1.3(b), it is readily done to convert them to data with respect to the number of nodes used. This is illustrated in Figures 2.8(a) and 2.8(b).

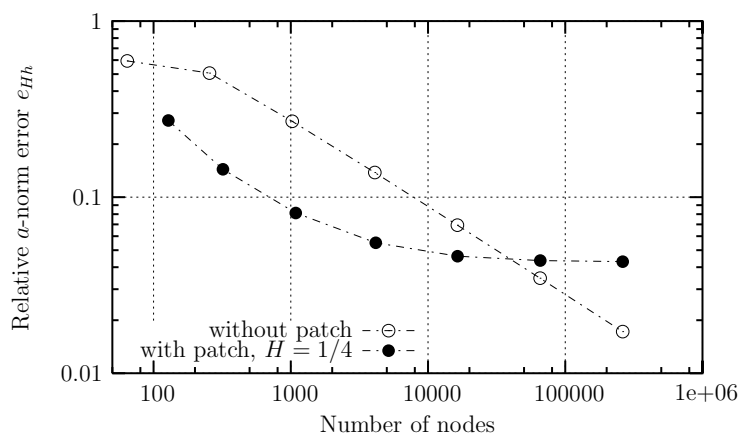
In Figure 2.8(a) we keep the ratio H/h constant. In Figure 2.8(b) we report results keeping the number of nodes of the coarse mesh fixed (i.e. $H = 1/4$) and divide h by two, keeping the size of the patch constant. Of course, after some refinements, the error in the domain $\Omega \setminus \Lambda$ with coarse discretization dominates and hence the global error stagnates, meaning that further refinement of the patch by reducing the mesh size h is not useful.

We note that the error reduction with respect to the number of degrees of freedom is concluding when refining the grid over the patch. This holds as long as the solution is well approached outside the patch by the coarse grid, i.e. as the error in the patch dominates.

In parallel to the analysis with respect to memory usage we report results of the use of CPU-time. These are shown in Figures 2.9(a) and 2.9(b). The CPU-time is reported in seconds as used by the basic implementation with **Freefem++**. When considering the experience as reported in Figure 2.8(a), i.e. increasing precision while decreasing H and



(a) Reduction of H and h with $H/h = 4$ fixed.

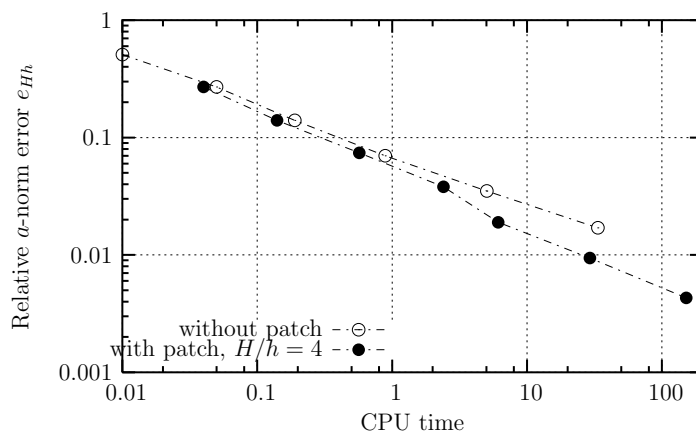


(b) Reduction of h with $H = 1/4$ fixed.

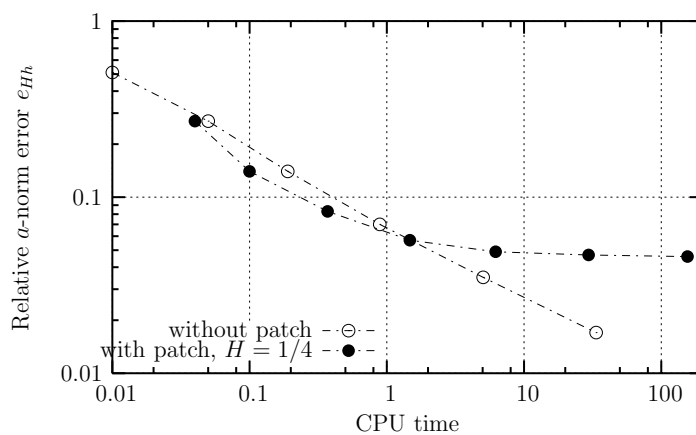
Figure 2.8: Convergence of u_{Hh} to u with respect to the number of nodes on structured and nested grids.

h with constant ratio H/h , we have a slight advantage in CPU-time with the correction method, this advantage growing when requiring more precision. In Figure 2.9(b) we plot the CPU-time needed for the results reported in Figure 2.8(b) (keeping H and the patch fixed and dividing h by two to increase precision). We conclude, as here above, that the refinement of the patch needs to be chosen in accordance with the underlying coarse grid in order to minimize the global error optimally.

Hence, we conclude that the key for an efficient use of the algorithm lies in adapted choice of the patch and its grid. In the CPU usage illustration from Figure 2.9(a) we note that the correction method becomes more efficient in terms of CPU-time as more precision is required, iterating getting less time consuming than solving larger systems.



(a) Reduction of H and h with $H/h = 4$ fixed.



(b) Reduction of h with $H = 1/4$ fixed.

Figure 2.9: Convergence of u_{Hh} to u with respect to the CPU-time usage on structured and nested grids.

Chapter 3

Analysis of the algorithm on two Poisson problems

In this chapter we take inspiration from the correction method (Algorithm 1.3) and consider regularity and convergence results for problems with singularities due to changing Dirichlet-Neumann boundary conditions and domains with entrant corners. Such problems have been studied for example by Grisvard [41, 42]. We discuss how patches can improve the quality of the solution and the convergence order in the grid size of the method on a model problem. In particular, we assess the efficiency of the application of patches with regard to the usage of memory.

The outline of this chapter is the following:

3.1	Preliminary results	58
3.2	Problem with change in boundary conditions	60
3.3	Problem in a domain with entrant corner	71

In Section 3.1 we introduce some preliminary results used for the *a priori* error analysis to follow. In Section 3.2, we consider a Laplace problem with changing Dirichlet-Neumann boundary conditions. First we study the regularity of solutions to this type of problems. Once the regularity result established (with inspiration taken from [41, 42]), we present *a priori* error estimates. The objective is to improve the latter through the use of the correction algorithm with chosen patches. Concluding results are presented, as well for the improvement of the convergence order in the mesh size and the precision of the approximation with regard to economic usage of memory. In Section 3.3, following the same structure then above, we assess our method on a Poisson-Dirichlet problem on a polygonal domain with entrant corner.

3.1 Preliminary results

In this section we introduce results leading to an *a priori* error estimate (Proposition 3.5) for the finite element approximation in a two-dimensional polygonal domain Ω of functions $u \in W^{2,p}(\Omega)$, $p \in (1; 2)$. Here $W^{2,p}(\Omega)$ denotes the usual Sobolev space of functions $f \in L^p(\Omega)$ with first and second derivatives in $L^p(\Omega)$.

We introduce \mathcal{T}_H a regular family of triangulations (see Ciarlet [29, Sect. 17]) with triangles K over $\bar{\Omega}$ and call $H = \max_{K \in \mathcal{T}_H} \text{diam}(K)$.

Lemma 3.1. *Let $q \in (1; +\infty)$ and $p \in (1; q)$. Then there exists $C = C(p, q)$ such that*

$$\|v\|_{L^p(K)}^p \leq \text{meas}(K)^{(q-p)/q} \|v\|_{L^q(K)}^p, \quad \forall v \in L^q(K), \forall K \in \mathcal{T}_H. \quad (3.1)$$

Proof. If $v \in L^q(K)$, we have

$$\|v\|_{L^p(K)}^p = \int_K |v|^p \, dx = \int_K 1 \cdot (|v|^q)^{p/q} \, dx = \int_K 1 \cdot w^{p/q} \, dx, \quad (3.2)$$

where $w = |v|^q$. Using Hölder's inequality with $s = q/p$ and $s^* = q/(q-p)$ (verifying $1/s + 1/s^* = 1$), we obtain

$$\|v\|_{L^p(K)}^p \leq \|1\|_{L^{s^*}(K)} \|w^{p/q}\|_{L^s(K)} \quad (3.3)$$

$$= \text{meas}(K)^{1/s^*} \left(\int_K w \, dx \right)^{p/q} \quad (3.4)$$

$$= \text{meas}(K)^{(q-p)/q} \|v\|_{L^q(K)}^p, \quad (3.5)$$

what concludes the proof. \square

Lemma 3.2. *Let $v \in L^q(\Omega)$ with $q \in (1; +\infty)$. Then, for any $p \in (1; q)$, there exists $H_0 = H_0(p, q, v)$ such that, $\forall H \leq H_0$,*

$$\|v\|_{L^p(K)} \leq 1, \quad \forall K \in \mathcal{T}_H. \quad (3.6)$$

Proof. With $v \in L^q(\Omega)$, we have

$$\|v\|_{L^q(\Omega)}^q = \sum_{K \in \mathcal{T}_H} \|v\|_{L^q(K)}^q = \sum_{K \in \mathcal{T}_H} \left(\|v\|_{L^p(K)}^p \right)^{q/p}, \quad (3.7)$$

and with Lemma 3.1:

$$\|v\|_{L^q(\Omega)}^q \geq \sum_{K \in \mathcal{T}_H} \left(\text{meas}(K)^{(p-q)/q} \|v\|_{L^p(K)}^p \right)^{q/p} \quad (3.8)$$

$$= \sum_{K \in \mathcal{T}_H} \text{meas}(K)^{(p-q)/p} \|v\|_{L^p(K)}^q \quad (3.9)$$

$$\geq C/H^{2(q-p)/p} \sum_{K \in \mathcal{T}_H} \|v\|_{L^p(K)}^q, \quad (3.10)$$

where C does not depend on K . Finally, we obtain

$$\sum_{K \in \mathcal{T}_H} \|v\|_{L^p(K)}^q \leq C^{-1} H^{2(q-p)/p} \|v\|_{L^q(\Omega)}^q, \quad (3.11)$$

which proves that

$$\lim_{H \rightarrow 0} \sum_{K \in \mathcal{T}_H} \|v\|_{L^p(K)}^q = 0. \quad (3.12)$$

Thus we have necessarily $\|v\|_{L^p(K)} \leq 1$ for $H \leq H_0$ sufficiently small. \square

Lemma 3.3. *Let $v \in L^q(\Omega)$ for all $q \in (1; 2)$. Then, for any $p \in (1; 2)$, there exists $H_0 = H_0(p, v)$ such that*

$$\sum_{K \in \mathcal{T}_H} \|v\|_{L^p(K)}^2 \leq \|v\|_{L^p(\Omega)}^p, \quad \text{if } H \leq H_0. \quad (3.13)$$

Proof. Let $p \in (1; 2)$ and consider $q \in (p; 2)$. By Lemma 3.2, there exists H_0 such that $\|v\|_{L^p(K)} \leq 1, \forall K \in \mathcal{T}_H, \forall H \leq H_0$. Hence

$$\sum_{K \in \mathcal{T}_H} \|v\|_{L^p(K)}^2 \leq \sum_{K \in \mathcal{T}_H} \|v\|_{L^p(K)}^p = \|v\|_{L^p(\Omega)}^p, \quad (3.14)$$

which completes the proof. \square

Let now $u \in W^{2,p}(\Omega)$, $p \in (1; 2)$, and consider the approximation of u by its interpolant $r_H u$ using a finite element method of order 1 on \mathcal{T}_H . Interpolation results by Ciarlet [29, Section 16], give the following *a priori* error estimate.

Proposition 3.4. *Let $u \in W^{2,p}(\Omega)$ with $p \in (1; 2)$. Then the interpolant $r_H u$ to u satisfies the a priori error estimate*

$$\|u - r_H u\|_{H^1(K)} \leq C H^{2-2/p} |u|_{W^{2,p}(K)}, \quad \forall K \in \mathcal{T}_H, \quad (3.15)$$

where C is a constant independent of H and u but depending on p , and $|\cdot|_{W^{2,p}(K)}$ is the semi-norm in $W^{2,p}(K)$.

Proof. The proof is straightforward by applying the result of equation (16.4) from Ciarlet [29, Section 16] to the present situation. \square

The idea of Proposition 3.5 is to give an estimate of $\|u - r_H u\|_{H^1(\Omega)}$.

Proposition 3.5. *Let $u \in W^{2,p}(\Omega)$ with $p \in (1; 2)$. Then there exists $H_0 = H_0(p, u)$ such that the interpolant $r_H u$ to u satisfies the a priori error estimate*

$$\|u - r_H u\|_{H^1(\Omega)} \leq C H^{2-2/p} |u|_{W^{2,p}(\Omega)}^{p/2}, \quad \forall H \leq H_0, \quad (3.16)$$

where C is a constant independent of H and u but depending on p .

Proof. We have $\|u - r_H u\|_{H^1(\Omega)}^2 = \sum_{K \in \mathcal{T}_H} \|u - r_H u\|_{H^1(K)}^2$. Since Ciarlet's result recalled in Proposition 3.4, we obtain

$$\|u - r_H u\|_{H^1(\Omega)}^2 \leq C H^{2(2-2/p)} \sum_{K \in \mathcal{T}_H} |u|_{W^{2,p}(K)}^2, \quad (3.17)$$

where C denotes a constant independent of H and u . With the result from Lemma 3.3 we can write

$$\sum_{K \in \mathcal{T}_H} |u|_{W^{2,p}(K)}^2 \leq |u|_{W^{2,p}(\Omega)}^p, \quad \forall H \leq H_0 = H_0(p, u). \quad (3.18)$$

Thus combining (3.17) and (3.18), we obtain

$$\|u - r_H u\|_{H^1(\Omega)}^2 \leq C H^{2(2-2/p)} |u|_{W^{2,p}(\Omega)}^p, \quad \forall H \leq H_0, \quad (3.19)$$

which leads to the result. \square

3.2 Problem with change in boundary conditions

The objective of this discussion is to assess the correction algorithm (Algorithm 1.3) on a Laplace problem with changing Dirichlet-Neumann boundary conditions. We first study the regularity of the solution to such a problem and give the *a priori* convergence results. Next we implement the problem numerically, and after comparing the theoretical orders with the ones we obtain numerically, we use different types of patches in order to improve the convergence order and the precision on the solution, economically with respect to the usage of memory.

Regularity result.

Before introducing a model problem (see problem (3.31)), we consider the situation here below to analyze the singular behavior of solutions in a domain with changing boundary conditions. Such an analysis is not new, a short development can be found in the books by Grisvard, see [41, Section 4.4] and [42, Pages 49–51 and Section 2.4]. Nevertheless it is useful to explicit the reasoning here.

Consider the domain $\Omega_\infty = (-\infty; +\infty) \times (0; +\infty)$ and the problem of finding the functions $v \in H_{\text{loc}}^1(\overline{\Omega}_\infty)$ verifying $\Delta v = 0$ in $L_{\text{loc}}^2(\overline{\Omega}_\infty)$ and obeying the following set of boundary conditions:

$$\begin{aligned} \frac{\partial v}{\partial n} &= 0 && \text{on } (-\infty; 0) \times \{0\}, \text{ and} \\ v &= 0 && \text{on } (0; +\infty) \times \{0\}. \end{aligned} \quad (3.20)$$

Note that $H_{\text{loc}}^1(\overline{\Omega}_\infty)$ can be extended to $H_{\text{loc}}^1(\mathcal{O})$ where \mathcal{O} is an open domain such that $\mathcal{O} \supset \overline{\Omega}_\infty$.

We denote x and y the two space variables and (r, θ) are the polar coordinates. The situation as well as the notation used are illustrated in Figure 3.1.

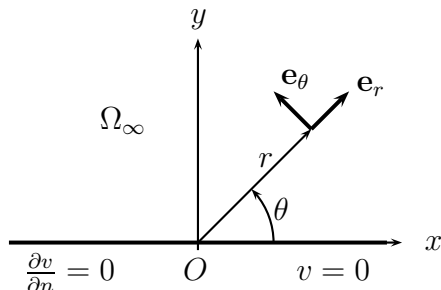


Figure 3.1: Illustration of the situation and notation.

To analyze these solutions, we consider the function $v = v(r, \theta)$ in polar coordinates. Taking into account the boundary conditions, we write

$$v(r, \theta) = \sum_{m \geq 0} \rho_m(r) \sin\left(\frac{2m+1}{2}\theta\right), \quad (3.21)$$

and calculate its gradient and Laplacian. By denoting \mathbf{e}_r , \mathbf{e}_θ the vectors of the tangential reference frame, the gradient in polar coordinates is given by $\nabla v = \partial_r v \mathbf{e}_r + \frac{1}{r} \partial_\theta v \mathbf{e}_\theta$, i.e.,

$$\nabla v = \sum_{m \geq 0} \rho'_m(r) \sin\left(\frac{2m+1}{2}\theta\right) \mathbf{e}_r + \sum_{m \geq 0} \frac{1}{r} \frac{2m+1}{2} \rho_m(r) \cos\left(\frac{2m+1}{2}\theta\right) \mathbf{e}_\theta. \quad (3.22)$$

The Laplacian, $\Delta v = \frac{1}{r} [\partial_r(r \partial_r v) + \frac{1}{r} \partial_{\theta\theta}^2 v]$, yields

$$\Delta v = \frac{1}{r} \left[\sum_{m \geq 0} \left(\partial_r(r \rho'_m(r)) - \left(\frac{2m+1}{2}\right)^2 \frac{1}{r} \rho_m(r) \right) \sin\left(\frac{2m+1}{2}\theta\right) \right]. \quad (3.23)$$

Hence, requiring $\Delta v = 0$ implies

$$r \rho''_m(r) + \rho'_m(r) - \left(\frac{2m+1}{2}\right)^2 \frac{1}{r} \rho_m(r) = 0, \quad m = 0, 1, 2, \dots \quad (3.24)$$

If we assume $\rho_m(r) = r^\gamma$, we obtain

$$\gamma(\gamma - 1) + \gamma - \left(\frac{2m+1}{2}\right)^2 = 0, \quad (3.25)$$

i.e.

$$\gamma = \pm \frac{2m+1}{2}, \quad m = 0, 1, 2, \dots \quad (3.26)$$

Hence the harmonic functions in Ω_∞ with boundary conditions (3.20) are expressed in polar coordinates by

$$v(r, \theta) = \sum_{m \geq 0} (c_m r^{(2m+1)/2} + c_{-m} r^{-(2m+1)/2}) \sin\left(\frac{2m+1}{2}\theta\right), \quad (3.27)$$

where c_m, c_{-m} are real coefficients. Note that c_0 and c_{-0} are different a priori.

In the sequel, using the general expression (3.27) in polar coordinates, we analyze the regularity of these solutions. We study first local integrability of $|\nabla v|^2$ in $\overline{\Omega}_\infty$. We calculate

$$\begin{aligned} |\nabla v|^2 &= \left[\sum_{m \geq 0} \left(\frac{2m+1}{2} c_m r^{(2m-1)/2} - \frac{2m+1}{2} c_{-m} r^{-(2m+3)/2} \right) \sin\left(\frac{2m+1}{2}\theta\right) \right]^2 \\ &+ \left[\sum_{m \geq 0} \frac{1}{r} \frac{2m+1}{2} (c_m r^{(2m+1)/2} + c_{-m} r^{-(2m+1)/2}) \cos\left(\frac{2m+1}{2}\theta\right) \right]^2. \end{aligned} \quad (3.28)$$

For $|\nabla v|^2$ to be locally integrable in $\overline{\Omega}_\infty$ a priori, we need to impose that, if $c_m \neq 0$, $\frac{2m-1}{2}2 + 1 > -1$, and if $c_{-m} \neq 0$, $-\frac{2m+3}{2}2 + 1 > -1$. The first condition is always verified for $m \geq 0$. The second implies $m < 0$, and hence $c_{-0} = c_{-1} = c_{-2} = \dots = 0$. Thus, the functions of the form (3.27) that are $H_{\text{loc}}^1(\overline{\Omega}_\infty)$ are expressed by

$$v(r, \theta) = \sum_{m \geq 0} c_m r^{(2m+1)/2} \sin\left(\frac{2m+1}{2}\theta\right). \quad (3.29)$$

Furthermore, considering the second derivatives of v , we note that, if $c_0 \neq 0$, then v does not belong to $H_{\text{loc}}^2(\overline{\Omega}_\infty)$.

Thus we are interested in calculating p such that v of the form (3.29) with $c_0 \neq 0$ is in $W_{\text{loc}}^{2,p}(\overline{\Omega}_\infty)$. For finding such p , we evaluate the second derivative of $r^{1/2}$ ($c_0 \neq 0$) and require it to be p -integrable. We obtain the relation $-\frac{3}{2}p + 1 > -1$, and hence $p < \frac{4}{3}$.

In conclusion, if $v \in H_{\text{loc}}^1(\overline{\Omega}_\infty)$ is an harmonic function in Ω_∞ verifying the boundary conditions (3.20), then $v \in W_{\text{loc}}^{2,p}(\overline{\Omega}_\infty)$ with $p \in [1; \frac{4}{3})$. We denote by $\varphi(r, \theta)$ the component $m = 0$ of v ,

$$\varphi(r, \theta) = c_0 \sqrt{r} \sin(\theta/2). \quad (3.30)$$

Model problem and a priori error estimate.

Let now $\Omega = (-L; L) \times (0; l) \subset \mathbb{R}^2$ be a rectangular domain (see Figure 3.2). We consider the following Poisson problem with homogeneous Dirichlet-Neumann boundary conditions:

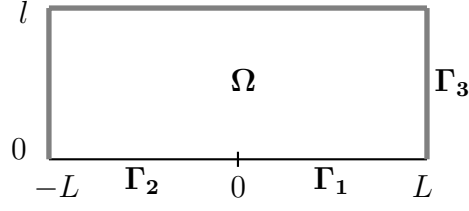


Figure 3.2: Illustration of the domain Ω and its boundaries Γ_i , $i = 1, 2, 3$.

For given $f \in L^2(\Omega)$, find $u \in H^1(\Omega)$ such that

$$\begin{cases} -\Delta u = f & \text{in } \Omega, \\ u = 0 & \text{on } \Gamma_1 = \{(x, 0) : 0 < x < L\}, \\ \frac{\partial u}{\partial n} = 0 & \text{on } \Gamma_2 = \{(x, 0) : -L < x < 0\}, \\ u = 0 & \text{on } \Gamma_3 = \partial\Omega \setminus (\Gamma_1 \cup \Gamma_2). \end{cases} \quad (3.31)$$

Let us remark that this problem means that, if $\tilde{H}^1(\Omega) = \{v \in H^1(\Omega) : v = 0 \text{ on } \Gamma_1 \cup \Gamma_3\}$, then $u \in \tilde{H}^1(\Omega)$ satisfies

$$\int_{\Omega} \nabla u \cdot \nabla v \, dx = \int_{\Omega} f v \, dx, \quad \forall v \in \tilde{H}^1(\Omega). \quad (3.32)$$

Since Grisvard [42, Section 2.4] complemented by the above paragraph, and since the four corners of Ω are right angles, we know that the unique solution u of (3.32) (Lax-Milgram Theorem) can be written as $u = w + \varphi$ where $w \in H^2(\Omega)$ and $\varphi \in W^{2,p}(\Omega)$, $p \in [1, \frac{4}{3})$, given by (3.30).

Recall that \mathcal{T}_H denotes a regular triangulation over $\bar{\Omega}$ with triangles K . We call $H = \max_{K \in \mathcal{T}_H} \text{diam}(K)$. We will assume that O is a node of the triangulation.

Let $u_H \in V_H = \{\psi \in C^0(\bar{\Omega}) : \psi|_K \in \mathbb{P}_1(K), \forall K \in \mathcal{T}_H \text{ and } \psi = 0 \text{ on } \Gamma_1 \cup \Gamma_3\}$ be the approximation of u such that

$$\int_{\Omega} \nabla u_H \cdot \nabla v_H \, dx = \int_{\Omega} f v_H \, dx, \quad \forall v_H \in V_H. \quad (3.33)$$

Using Céa's Lemma we have

$$\int_{\Omega} |\nabla(u - u_H)|^2 \, dx = \int_{\Omega} \nabla(u - u_H) \cdot \nabla(u - v_H) \, dx, \quad (3.34)$$

and consequently

$$|u - u_H|_{H^1(\Omega)} \leq |u - v_H|_{H^1(\Omega)}, \quad (3.35)$$

where $|\cdot|_{H^1(\Omega)}$ is the semi-norm in $H^1(\Omega)$. Since the Poincaré inequality it satisfies in $\tilde{H}^1(\Omega)$, we have by taking $v_H = r_H u$ (possible with $u \in C^0(\bar{\Omega})$),

$$|u - u_H|_{H^1(\Omega)} \leq C|u - r_H u|_{H^1(\Omega)}. \quad (3.36)$$

With $u \in W^{2,p}(\Omega)$, $p \in [1; \frac{4}{3})$, and by Proposition 3.5 we have the following *a priori* error estimate: For $p \in (1; 4/3)$, there exists $H_0 = H_0(p, u)$ such that the approximation u_H given by (3.33) to u satisfies

$$\|u - u_H\|_{H^1(\Omega)} \leq CH^{2-2/p}|u|_{W^{2,p}(\Omega)}^{p/2}, \quad \forall H \leq H_0, \quad (3.37)$$

where C is a constant independent of H and u but depending on p .

Hence, the *a priori* convergence order in H in the H^1 -norm is smaller than $2 - \frac{2}{4/3} = 1/2$.

The objective of the next part will be to apply the correction method introduced in Chapter 1 to the present situation. We use a patch Λ with a finer triangulation to augment the precision around the origin where the first derivative of the solution u explodes in O . By strategically choosing the patch we try to optimize the convergence order.

Improving the *a priori* convergence order through using patches.

In this paragraph we keep considering the model problem (3.31) introduced above with its approximation (3.33). The aim is to use the correction algorithm to obtain a better *a priori* convergence order.

As described in Chapter 1 (Section 1.1), we consider two families of regular triangulations \mathcal{T}_H over $\overline{\Omega}$ and \mathcal{T}_h over a patch $\overline{\Lambda}_\epsilon$, the size of which depends on $\epsilon > 0$. Then the idea is to use linear elements and approximate the solution u of (3.31) with $u_{Hh} = u_H + u_h$, u_H and u_h defined on \mathcal{T}_H and \mathcal{T}_h by Algorithm 1.3. In order to develop *a priori* error estimates, we need to introduce a particular setting where it is valid. At the end of this paragraph we conjecture a generalization of the latter.

Consider the domain $\Omega = (-L; L) \times (0; l)$. To simplify the discussion and to fix the ideas we take $L = l = 1$. We introduce a regular triangulation \mathcal{T}_H over $\overline{\Omega}$ with triangles K such that, if \mathcal{N}_H denotes the set of nodes of the triangulation \mathcal{T}_H and \mathcal{C}_ϵ is the half-circle in $\overline{\Omega}$ centered at the origin and of radius $\epsilon \in (0; 1/2)$, we have $O \in \mathcal{N}_H$, $\{(\pm\epsilon, 0)\} \subset \mathcal{N}_H$, and $\forall K \in \mathcal{T}_H$, $\partial K \cap \mathcal{C}_\epsilon \subset \mathcal{N}_H$. We call $H = \max_{K \in \mathcal{T}_H} \text{diam}(K)$ and suppose furthermore that ϵ is such that $H/\epsilon \rightarrow 0$.

If \mathcal{D}_ϵ denotes the half-disk in $\overline{\Omega}$ centered at the origin and of radius ϵ , we introduce the patch $\overline{\Lambda}_\epsilon = \cup_{K \in \mathcal{T}_H, K \subset \mathcal{D}_\epsilon} K$. Over $\overline{\Lambda}_\epsilon$ we consider a regular triangulation \mathcal{T}_h . We call \mathcal{N}_h the set of nodes of the triangulation \mathcal{T}_h and $h = \max_{K \in \mathcal{T}_h} \text{diam}(K)$. We suppose that the triangulations \mathcal{T}_H and \mathcal{T}_h are nested, i.e. $\forall K \in \mathcal{T}_h$, $\exists \tilde{K} \in \mathcal{T}_H$ s.t. $K \subset \tilde{K}$.

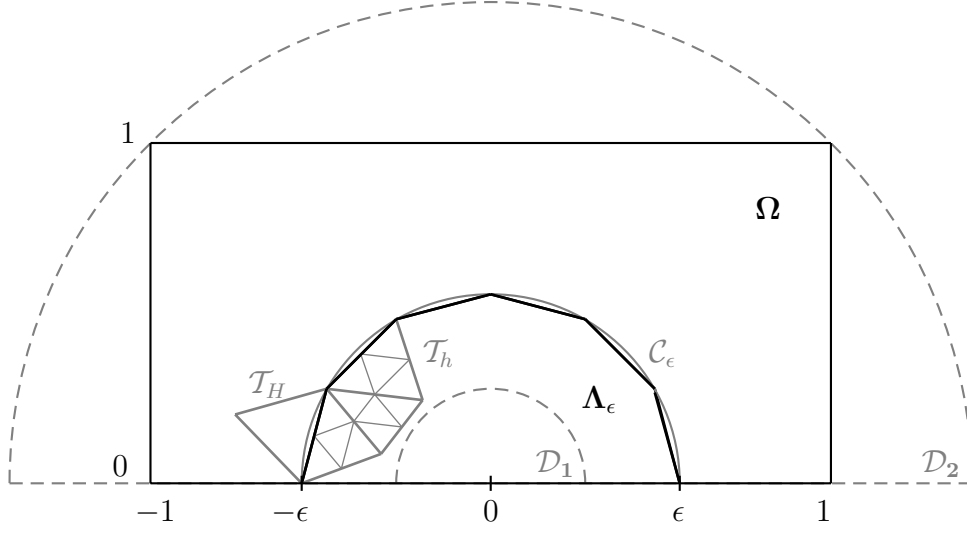


Figure 3.3: Illustration of the setting and notation.

Since our idea is to use linear finite elements to approximate u on \mathcal{T}_H and \mathcal{T}_h , we need to introduce the *ad hoc* spaces. Let $V_H = \{\psi \in H^1(\Omega) : \psi|_K \in \mathbb{P}_1(K), \forall K \in \mathcal{T}_H \text{ and } \psi = 0 \text{ on } (0; L) \times \{0\}\}$ and $V_h = \{\psi \in H^1(\Omega) : \psi|_K \in \mathbb{P}_1(K), \forall K \in \mathcal{T}_h \text{ and } \psi = 0 \text{ in } \bar{\Omega} \setminus \bar{\Lambda}_\epsilon \text{ and } \psi = 0 \text{ on } \partial\Lambda_\epsilon \setminus ([-\epsilon; 0] \times \{0\})\}$. We denote by r_H and r_h the standard interpolants to the space V_H and V_h respectively.

We call $\tilde{\mathcal{N}}_h = \mathcal{N}_h \cap (\partial\Lambda_\epsilon \setminus (-\epsilon; \epsilon) \times \{0\})$. We consider $\tilde{V}_h = \{\psi \in H^1(\Lambda_\epsilon) : \psi|_K \in \mathbb{P}_1(K), \forall K \in \mathcal{T}_h \text{ and } \psi(P) = 0, \forall P \in \mathcal{N}_h \setminus \tilde{\mathcal{N}}_h\}$, and call \tilde{r}_h the standard interpolant to the space \tilde{V}_h . Call $\bar{V}_h = V_h + \tilde{V}_h$ and \bar{r}_h the interpolant to \bar{V}_h .

An illustration of the introduced setting is given in Figure 3.3.

Lemma 3.6. *Consider the setting and the notation introduced above. For φ given by (3.30) we have*

$$|\varphi|_{H^2(\Omega \setminus \bar{\Lambda}_\epsilon)} \leq C/\sqrt{\epsilon}, \quad (3.38)$$

where C is a constant independent of ϵ .

Proof. Since (3.30) we have

$$|\varphi|_{H^2(\Omega \setminus \bar{\Lambda}_\epsilon)}^2 \leq C \int_{\Omega \setminus \bar{\Lambda}_\epsilon} |x|^{-3} dx. \quad (3.39)$$

Consider the half-disk $\mathcal{D}_1 \subset \bar{\Lambda}_\epsilon$ of radius $\epsilon/2$ centered at the origin, and the smallest half-disk $\mathcal{D}_2 \supset \bar{\Omega}$ centered at the origin (see Figure 3.3). The radius of \mathcal{D}_1 being of order ϵ

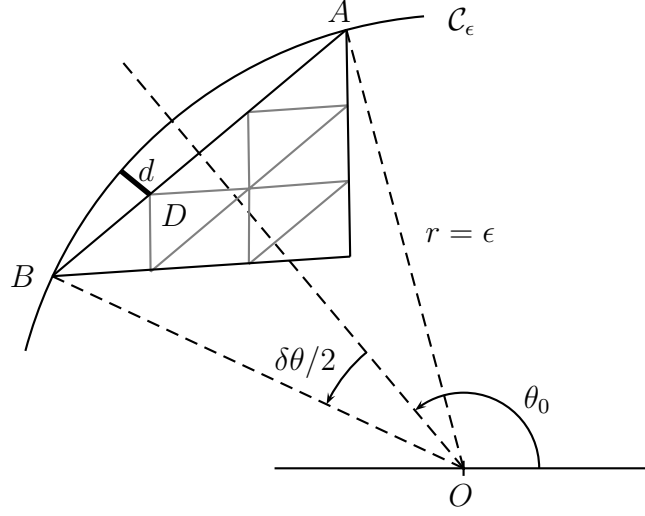


Figure 3.4: Illustration of the notation for Lemma 3.7.

and the radius of \mathcal{D}_2 of order 1, we can write

$$\int_{\Omega \setminus \bar{\Lambda}_\epsilon} |x|^{-3} dx \leq \int_{\mathcal{D}_2 \setminus \bar{\mathcal{D}}_1} |x|^{-3} dx \leq C \int_{\epsilon/2}^1 r^{-2} dr \leq C/\epsilon, \quad (3.40)$$

and hence conclude. \square

Lemma 3.7. *Consider the setting and the notation introduced above and recall in particular the hypothesis $H/\epsilon \rightarrow 0$. If $\chi_h = \tilde{r}_h(\varphi - r_H\varphi)|_{\bar{\Lambda}_\epsilon} \in \tilde{V}_h$, then we have*

$$|\chi_h|_{H^1(\Lambda_\epsilon)} \leq C \frac{H^2}{\epsilon h^{1/2}}, \quad (3.41)$$

where C is a constant independent of H , h and ϵ .

Proof. Let $A, B \in \mathcal{N}_H \cap \mathcal{C}_\epsilon$ be two vertices of a triangle $K \in \mathcal{T}_H$, $K \subset \bar{\Lambda}_\epsilon$, and $D \in \mathcal{N}_h \cap [A; B]$. The polar coordinates of A and B are respectively $(r_A = \epsilon, \theta_A = \theta_0 - \delta\theta/2)$ and $(r_B = \epsilon, \theta_B = \theta_0 + \delta\theta/2)$ where θ_0 is the angle bisecting the arc \widehat{AB} in two equal parts and $\delta\theta = H/\epsilon + O((H/\epsilon)^2)$ is the measure of the angle \widehat{AOB} . See Figure 3.4 for an illustration of the situation.

The parameter $t \in [-1/2; 1/2]$ defines the position of $D(t)$ given by $\overrightarrow{OD}(t) = \overrightarrow{OA} + (t + 1/2)\overrightarrow{AB}$. We write $D(t = -1/2) = A$ and $D(t = 1/2) = B$.

Taking $c_0 = 1$ in (3.30), we have $\varphi(A) = \sqrt{\epsilon} \sin(\theta_0 - \delta\theta/2)$ and $\varphi(B) = \sqrt{\epsilon} \sin(\theta_0 + \delta\theta/2)$. Thus $(r_H\varphi)(D(t)) = (1/2 - t)\varphi(A) + (1/2 + t)\varphi(B)$ and with the development $\sin(\theta_0 + \delta\theta) = \sin \theta_0 + \cos \theta_0 \delta\theta + O(\delta\theta^2)$, $\delta\theta \ll 1$ since $H/\epsilon \rightarrow 0$, we get

$$(r_H\varphi)(D(t)) = \sqrt{\epsilon} (\sin \theta_0 + t \cos \theta_0 \delta\theta + O(\delta\theta^2)). \quad (3.42)$$

Furthermore, as the arc \widehat{AB} is described by the points $(r = \epsilon, \theta = \theta_0 + t\delta\theta)$ with $t \in [-1/2; 1/2]$, we can explicitly write $\varphi(D(t)) = \sqrt{\epsilon - d(t)} \sin(\theta_0 + t\delta\theta)$ where

$$d(t) = \epsilon - \sqrt{\epsilon^2 - |AB|^2(1/4 - t^2)} \leq d(t=0) \leq CH^2/\epsilon, \quad (3.43)$$

with $|AB| \leq H$ denoting the length of the segment $[A; B]$ and C a generic constant. Note that $\epsilon - d(t) = |OD|$. We develop $\varphi(D(t))$ as follows:

$$\varphi(D(t)) = \sqrt{\epsilon - d(t)} (\sin(\theta_0) + t \cos \theta_0 \delta\theta + O(\delta\theta^2)). \quad (3.44)$$

Finally, combining (3.42) and (3.44) with the inequality (3.43), we can write

$$|\chi_h|_{H^1(\Lambda_\epsilon)} \leq C \left[\left(\frac{\sqrt{\epsilon} - \sqrt{\epsilon - d(0)}}{h} \right)^2 \epsilon h \right]^{1/2} \quad (3.45)$$

$$\leq C \left(\sqrt{\epsilon} \frac{d(0)}{\epsilon} \right) \epsilon^{1/2} h^{-1/2} \quad (3.46)$$

$$\leq CH^2 \epsilon^{-1} h^{-1/2}, \quad (3.47)$$

concluding the proof. \square

Proposition 3.8. *Suppose that u is the solution of (3.31). Let $p \in (1; 4/3)$ and consider the setting introduced above. Then there exist C and h_0 such that the approximation u_{Hh} to u satisfies the a priori error estimate*

$$\|u - u_{Hh}\|_{H^1(\Omega)} \leq C \left(\frac{H}{\sqrt{\epsilon}} + h^{2-2/p} + \frac{H^2}{\epsilon h^{1/2}} \right), \quad \forall h \leq h_0, H/\epsilon \rightarrow 0, \quad (3.48)$$

where C and h_0 are constants independent of H , h and ϵ but depending on p and u .

Proof. Since u is the solution of (3.31) we have that $u = w + \varphi$, where $w \in H^2(\Omega)$ and φ is given by (3.30), $\varphi = c_0 \sqrt{r} \sin(\theta/2)$. Hence

$$|u - u_{Hh}|_{H^1(\Omega)}^2 \leq |u - v_{Hh}|_{H^1(\Omega)}^2 \quad (3.49)$$

$$\leq C \left(|w - v_{Hh}^1|_{H^1(\Omega)}^2 + |\varphi - v_{Hh}^2|_{H^1(\Omega)}^2 \right), \quad \forall v_{Hh}^1, v_{Hh}^2 \in V_{Hh}, \quad (3.50)$$

where $v_{Hh} = v_{Hh}^1 + v_{Hh}^2$ and C denotes a generic constant independent of H and h .

We choose $v_{Hh}^1 = r_H w$. Standard interpolation results yield

$$|w - v_{Hh}^1|_{H^1(\Omega)} = |w - r_H w|_{H^1(\Omega)} \leq CH |w|_{H^2(\Omega)}. \quad (3.51)$$

In the second term of (3.50) we write $v_{Hh}^2 = v_H + v_h$ with $v_H \in V_H$ and $v_h \in V_h$. We have

$$|\varphi - v_{Hh}^2|_{H^1(\Omega)}^2 = |\varphi - v_H - v_h|_{H^1(\Omega \setminus \overline{\Lambda_\epsilon})}^2 + |\varphi - v_H - v_h|_{H^1(\Lambda_\epsilon)}^2. \quad (3.52)$$

Since $\varphi \in C^0(\overline{\Omega})$, we can choose $v_H = r_H\varphi$ and $v_h = r_h(\varphi - r_H\varphi)$ in Λ_ϵ , $v_h = 0$ in $\Omega \setminus \overline{\Lambda_\epsilon}$ and on all nodes in $\tilde{\mathcal{N}}_h$. We have

$$|\varphi - v_H - v_h|_{H^1(\Omega \setminus \overline{\Lambda_\epsilon})} = |\varphi - r_H\varphi|_{H^1(\Omega \setminus \overline{\Lambda_\epsilon})}, \quad (3.53)$$

and hence, since $\varphi \in H^2(\Omega \setminus \overline{\Lambda_\epsilon})$, by standard interpolation results, and furthermore with Lemma 3.6, equation (3.38), we get

$$|\varphi - v_H - v_h|_{H^1(\Omega \setminus \overline{\Lambda_\epsilon})} \leq CH|\varphi|_{H^2(\Omega \setminus \overline{\Lambda_\epsilon})} \leq CH/\sqrt{\epsilon}. \quad (3.54)$$

Let us now turn to the second term in the right-hand side of (3.52). We have

$$|\varphi - v_H - v_h|_{H^1(\Lambda_\epsilon)}^2 = |\varphi - r_H\varphi - r_h(\varphi - r_H\varphi)|_{H^1(\Lambda_\epsilon)}^2 \quad (3.55)$$

$$= |\varphi - r_H\varphi - r_h(\varphi - r_H\varphi) - \chi_h|_{H^1(\Lambda_\epsilon)}^2 + |\chi_h|_{H^1(\Lambda_\epsilon)}^2 \quad (3.56)$$

where $\chi_h = \tilde{r}_h(\varphi - r_H\varphi)|_{\overline{\Lambda_\epsilon}}$. Hence $r_h(\varphi - r_H\varphi) + \chi_h$ is equal to $\bar{r}_h(\varphi - r_H\varphi)|_{\overline{\Lambda_\epsilon}}$, and with Propositions 3.4 and 3.3, there exists h_0 such that

$$|\varphi - r_H\varphi - r_h(\varphi - r_H\varphi) - \chi_h|_{H^1(\Lambda_\epsilon)}^2 = \sum_{K \in \mathcal{T}_h} |\varphi - r_H\varphi - \bar{r}_h(\varphi - r_H\varphi)|_{H^1(K)}^2 \quad (3.57)$$

$$\leq Ch^{2(2-2/p)} \sum_{K \in \mathcal{T}_h} |\varphi|_{W^{2,p}(K)}^2 \quad (3.58)$$

$$\leq Ch^{2(2-2/p)} \sum_{K \in \mathcal{T}_h} |\varphi|_{W^{2,p}(K)}^p, \quad \forall h \leq h_0 \quad (3.59)$$

Finally,

$$|\varphi - r_H\varphi - r_h(\varphi - r_H\varphi) - \chi_h|_{H^1(\Lambda_\epsilon)} \leq Ch^{2-2/p} |\varphi|_{W^{2,p}(\Lambda_\epsilon)}^{p/2}, \quad \forall h \leq h_0. \quad (3.60)$$

Furthermore, with Lemma 3.7, we conclude that

$$|\varphi - v_H - v_h|_{H^1(\Lambda_\epsilon)} \leq C \left(h^{2-2/p} |\varphi|_{W^{2,p}(\Lambda_\epsilon)}^{p/2} + \frac{H^2}{\epsilon h^{1/2}} \right), \quad \forall h \leq h_0. \quad (3.61)$$

Finally, combining (3.54) and (3.61) in (3.52), introduced with (3.51) in (3.50), we obtain:

$$|u - u_{Hh}|_{H^1(\Omega)}^2 \leq C \left(H^2 |w|_{H^2(\Omega)}^2 + H^2/\epsilon + h^{2(2-2/p)} |\varphi|_{W^{2,p}(\Lambda_\epsilon)}^p + \frac{H^4}{\epsilon^2 h} \right), \quad \forall h \leq h_0, \quad (3.62)$$

i.e., with the Poincaré inequality,

$$\|u - u_{Hh}\|_{H^1(\Omega)} \leq C \left(H/\sqrt{\epsilon} + h^{2-2/p} + \frac{H^2}{\epsilon h^{1/2}} \right), \quad \forall h \leq h_0. \quad (3.63)$$

Note that C and h_0 depend on u and p . \square

It is of interest to write out in Proposition 3.8 the case where $p \rightarrow 4/3$ and $\epsilon = \alpha H^\beta$, where α and β denote constants, $\beta < 1$. Then the relation (3.48) yields the following (for h small enough):

$$\|u - u_{Hh}\|_{H^1(\Omega)} \leq C (H^{1-\beta/2} + h^{1/2} + H^{2-\beta}/h^{1/2}). \quad (3.64)$$

Hence when ϵ is proportional to H^β , we choose h proportional to $H^{2-\beta}$ to optimize the estimate and obtain a convergence of order $1 - \beta/2$ in H .

In the case where $\beta = 0$, i.e. the patch is fixed ($\epsilon = \alpha$), we note that convergence of order one in H can be obtained (in the limit $p \rightarrow 4/3$) when choosing h proportional to H^2 . It is adequate to note here that Grisvard proves, under certain conditions [41, Theorem 8.4.1.6] on the local refinement of a family of triangulations, that optimal convergence order, i.e. order one in $H^1(\Omega)$ -norm, can be reached despite the singularity [41, Corollary 8.4.1.7]. In the present situation of a solution in $W^{2,p}(\Omega)$ with $p \in [1; \frac{4}{3})$, the conditions of Grisvard's Theorem 8.4.1.6 [41] aim that, as $H \rightarrow 0$, there exists a constant σ such that

- (i) $\max_{K \in \mathcal{T}_H} H_K / \rho_K \leq \sigma$ where H_K is the diameter and ρ_K the interior diameter of K , i.e., the family of triangulations is regular;
- (ii) $H_K \leq \sigma H^2$, for any triangle K with one corner at the origin;
- (iii) $H_K \leq \sigma H \inf_K r^{1/2}$, for any triangle K without corner at the origin.

Condition (ii) of the above result on the refined families requires that the diameter of triangles around the origin is of order H^2 . The above discussion of Proposition 3.8 also implies this crucial condition which can also be found in the work [55] by Raugel.

As we see in the sequel, it is of practical interest to choose patches of variable size when refining, in particular with respect to memory usage and computation time. In Table 3.1 we give an overview of different situations to be studied and the extremal convergence order that we can expect to obtain.

Furthermore, using the convincing numerical results that we present in the next paragraph, we conjecture that Proposition 3.8 is true in a more general framework of unstructured triangulations and other forms of patches.

Numerical results.

In order to assess the given error estimate, we consider the problem of approximating $u = w + \varphi$ with $w = 0$ and φ given by (3.30). We consider the geometry illustrated

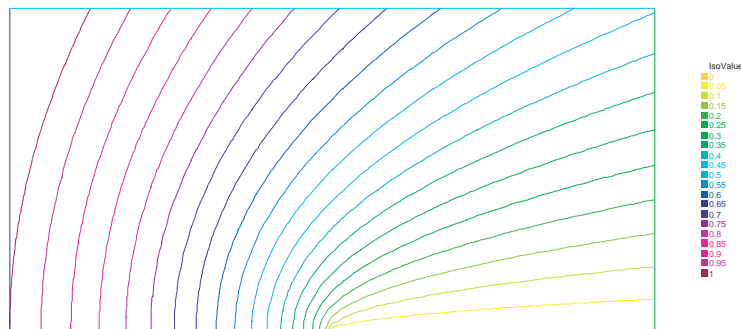


Figure 3.5: Isolines of the solution $u = \varphi$ of problem (3.65).

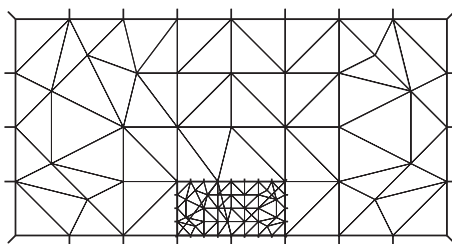


Figure 3.6: Grid constellation showing a patch with $\epsilon = 0.25$ and $N = M = 8$.

in Figure 3.2. Numerical tests when solving the problem of finding $u \in H^1(\Omega)$ such that

$$\begin{cases} -\Delta u = 0 & \text{in } \Omega, \\ u = 0 & \text{on } \Gamma_1, \\ \frac{\partial u}{\partial n} = 0 & \text{on } \Gamma_2, \\ u = \varphi|_{\Gamma_3} & \text{on } \Gamma_3, \end{cases} \quad (3.65)$$

yielding obviously $u = \varphi$, are reported in the following. An illustration of the isolines of the solution $u = \varphi$ is given in Figure 3.5.

Over $\bar{\Omega} = [-L; L] \times [0; l]$ we consider an unstructured regular triangulation \mathcal{T}_H using a discretization of $(-L; L)$ with N intervals and $(0; l)$ with $N/2$ intervals. We choose $L = l = 1$. We choose $\Lambda_\epsilon = (-\epsilon; \epsilon) \times (0; \epsilon)$ and consider \mathcal{T}_h regular and unstructured based on a discretization of $(-\epsilon; \epsilon)$ with M intervals. In Figure 3.6 we illustrate the grid constellation with $\epsilon = 0.25$ and $N = M = 8$.

Since the discussion at the end of the last paragraph, we choose $\epsilon = \alpha H^\beta$, $\beta < 1$, and accordingly $h = H^{2-\beta}$. With the above notation this is obtained when choosing $M = \alpha^2 2^{2\beta-1} N^{2-2\beta}$. (If this formula yields a non integer number, we take its integer part.) Conjecturing the validity of Proposition 3.8 in the present setting, we expect that the extremal ($p \rightarrow 4/3$) *a priori* order of convergence in H for the $H^1(\Omega)$ -norm error is given by $1 - \beta/2$. An overview of different cases applied on the solution of problem (3.65) is reported in Table 3.1. In the last column of the latter we report the numerically obtained order in H . This value corresponds to the slope at the last level of refinement of

ϵ	h	M	CO1	CO2
0.25	H^2	$N^2/32$	1	1.06
$H^{1/4}$	$H^{7/4}$	$N^{3/2}/\sqrt{2}$	$7/8 = 0.875$	0.85
\sqrt{H}	$H^{3/2}$	N	$3/4 = 0.750$	0.72
$H^{3/4}$	$H^{5/4}$	$N^{1/2}\sqrt{2}$	$5/8 = 0.625$	0.61
no patch			0.5	0.50

Table 3.1: Synoptic table of chosen patches and $H^1(\Omega)$ -norm convergence orders (CO1 = extremal *a priori* order in H , CO2 = order obtained by numerical experience).

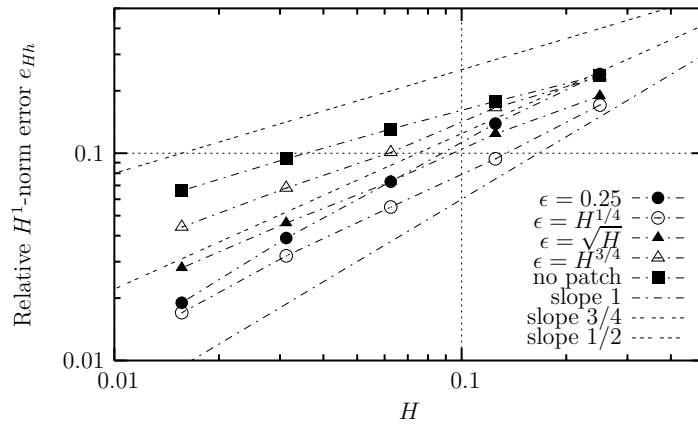
the decreasing relative $H^1(\Omega)$ -norm error in the mesh size H . This is readily seen from Figure 3.7(a) where we illustrate the concluding convergence behavior in the mesh-size graphically.

Note that the option with $\epsilon = 0.25$ makes the number of nodes of the problem of the correction level grow like M^2 proportional to N^4 as H decreases. In Figure 3.7(b) we assess the error reduction with respect to the number of discretization points used. With comparison to solving the problem without patch, the method using a fixed patch with number of discretization points increasing as N^4 for decreasing H is memory consuming and not satisfactory. A fixed patch is uninteresting in terms of memory usage. However the correction method using a variable patch is economic with respect to memory usage. Well applied patches decrease the error efficiently. In particular when high precision is needed the variable patch is most interesting.

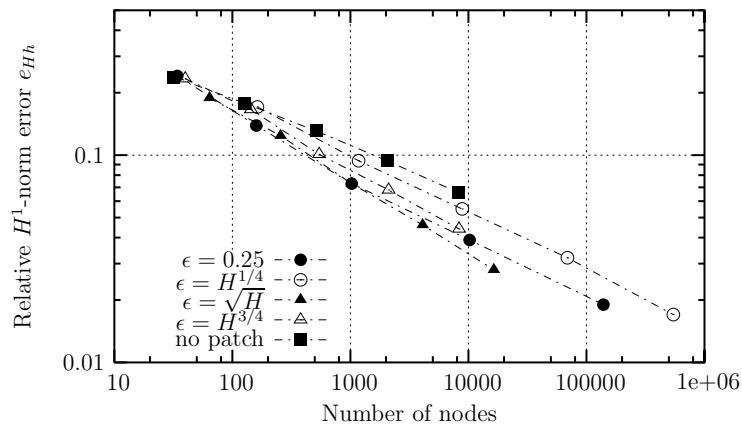
The results of this model problem with a singularity due to the change in the boundary conditions are transferable to singularities whose origin lies in a computational domain with entrant corner. This study is presented in the next section.

3.3 Problem in a domain with entrant corner

In the previous section we have discussed the Poisson problem in a rectangular domain with changing Dirichlet-Neumann boundary conditions. Through numerical results we have successfully shown that applying a patch in the region where the solution is less regular, we can improve the convergence order and the accuracy of the solution with only slight increase of memory usage. In the sequel, the objective is to examine a Poisson-Dirichlet problem in domains with entrant corners. This will be done similarly to the foregoing analysis. At first we proceed with a regularity analysis for a domain with an entrant corner on its boundary. We particularize the result to a situation with an L-shaped domain and give numerical results.



(a) Convergence order in the mesh size.



(b) Error reduction vs. number of nodes.

Figure 3.7: Convergence of u_H resp. u_{Hh} to u with respect to the mesh size H and the number of nodes.

Regularity result.

A brief analysis of this situation can be found in Grisvard [41, Sections 4.4 and 8.4] and [42, Pages 49–51 and Section 2.4]. It is useful to explicit the details of the analysis here.

Consider the domain $\Omega_\infty \subset \mathbb{R}^2$ as depicted in Figure 3.8. We consider for Ω_∞ the part of \mathbb{R}^2 where the internal angle β at the origin on the boundary Γ_∞ is such that $\pi < \beta < 2\pi$. We consider the problem of finding the functions $v \in H_{\text{loc}}^1(\overline{\Omega}_\infty)$ verifying $\Delta v = 0$ in $L_{\text{loc}}^2(\overline{\Omega}_\infty)$ with homogeneous Dirichlet boundary conditions on Γ_∞ . We consider polar coordinates centered at the corner O of Γ_∞ . The situation and the notation are illustrated in Figure 3.8.

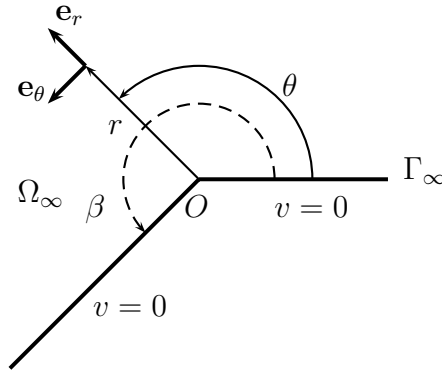


Figure 3.8: Illustration of the situation and notation.

We follow the same analysis of the solution as in the previous section. Taking into account the shape of the domain and the boundary conditions, we consider the function $v = v(r, \theta)$ of the form

$$v(r, \theta) = \sum_{m \geq 1} \rho_m(r) \sin \left(m \frac{\pi}{\beta} \theta \right). \quad (3.66)$$

Its gradient and Laplacian are given by the following expressions:

$$\nabla v = \sum_{m \geq 1} \rho'_m(r) \sin \left(m \frac{\pi}{\beta} \theta \right) \mathbf{e}_r + \sum_{m \geq 1} \frac{1}{r} m \frac{\pi}{\beta} \rho_m(r) \cos \left(m \frac{\pi}{\beta} \theta \right) \mathbf{e}_\theta, \quad (3.67)$$

and

$$\Delta v = \frac{1}{r} \left[\sum_{m \geq 1} \left(\partial_r (r \rho'_m(r)) - \left(m \frac{\pi}{\beta} \right)^2 \frac{1}{r} \rho_m(r) \right) \sin \left(m \frac{\pi}{\beta} \theta \right) \right]. \quad (3.68)$$

The condition $\Delta v = 0$ implies

$$r \rho_m''(r) + \rho'_m(r) - \left(m \frac{\pi}{\beta} \right)^2 \frac{1}{r} \rho_m(r) = 0, \quad m = 1, 2, 3, \dots \quad (3.69)$$

If $\rho_m(r) = r^\gamma$, solving (3.69) for γ yields

$$\gamma = \pm m \frac{\pi}{\beta}, \quad m = 1, 2, 3, \dots \quad (3.70)$$

The harmonic functions in Ω_∞ with homogeneous Dirichlet boundary condition on Γ_∞ are expressed in polar coordinates by

$$v(r, \theta) = \sum_{m \geq 1} (c_m r^{m\pi/\beta} + c_{-m} r^{-m\pi/\beta}) \sin\left(m \frac{\pi}{\beta} \theta\right), \quad (3.71)$$

where c_m, c_{-m} are real coefficients.

We consider now the gradient of v . We calculate

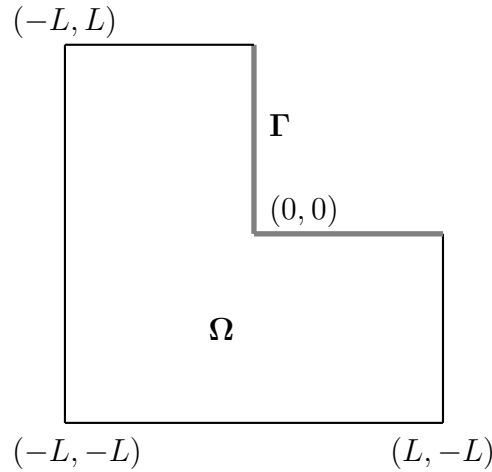
$$\begin{aligned} |\nabla v|^2 &= \left[\sum_{m \geq 1} \left(m \frac{\pi}{\beta} c_m r^{m\pi/\beta-1} - m \frac{\pi}{\beta} c_{-m} r^{-m\pi/\beta-1} \right) \sin\left(m \frac{\pi}{\beta} \theta\right) \right]^2 \\ &\quad + \left[\sum_{m \geq 1} \frac{1}{r} m \frac{\pi}{\beta} (c_m r^{m\pi/\beta} + c_{-m} r^{-m\pi/\beta}) \cos\left(m \frac{\pi}{\beta} \theta\right) \right]^2. \end{aligned} \quad (3.72)$$

For $|\nabla v|^2$ to be locally integrable in $\overline{\Omega}_\infty$ a priori, we need to impose that, if $c_m \neq 0$, $\left(m \frac{\pi}{\beta} - 1\right) 2 + 1 > -1$, and if $c_{-m} \neq 0$, $\left(-m \frac{\pi}{\beta} - 1\right) 2 + 1 > -1$. The first condition is always verified for $m \geq 1$. The second implies $m < 0$, and hence $c_{-1} = c_{-2} = c_{-3} = \dots = 0$. Thus, the functions of the form (3.71) that are $H_{\text{loc}}^1(\overline{\Omega}_\infty)$ are expressed by

$$v(r, \theta) = \sum_{m \geq 1} c_m r^{m\pi/\beta} \sin\left(m \frac{\pi}{\beta} \theta\right). \quad (3.73)$$

Considering the second derivatives of v , we note that, if $c_1 \neq 0$, then v does not belong to $H_{\text{loc}}^2(\overline{\Omega}_\infty)$. Thus we are interested in calculating p such that v of the form (3.73) with $c_1 \neq 0$ is in $W_{\text{loc}}^{2,p}(\overline{\Omega}_\infty)$. For finding such p , we evaluate the second derivative of $r^{\pi/\beta}$ ($c_1 \neq 0$) and require it to be p -integrable. We obtain the relation $\left(\frac{\pi}{\beta} - 2\right) p + 1 > -1$, and hence $p < \frac{2}{2-\pi/\beta}$.

In conclusion, if $v \in H_{\text{loc}}^1(\overline{\Omega}_\infty)$ is an harmonic function in Ω_∞ verifying the homogeneous Dirichlet boundary condition on Γ_∞ , then $v \in W_{\text{loc}}^{2,p}(\overline{\Omega}_\infty)$ with $p \in [1; \frac{2}{2-\pi/\beta})$. Note that in the extremal case where $\beta \rightarrow 2\pi$ we are left with $W_{\text{loc}}^{2,p}$ -regularity, $p \in [1; \frac{4}{3})$. Thus the regularity of this extreme situation in the current problem of a domain with entrant corner corresponds to the regularity of the problem with changing Dirichlet-Neumann boundary conditions on a straight boundary as studied in Section 3.2.


 Figure 3.9: Illustration of the domain Ω .

Model problem and *a priori* error estimate.

Let the domain $\Omega \subset (-L; L)^2 \subset \mathbb{R}^2$ be the L-shaped domain as depicted in Figure 3.9. We consider the following Poisson problem with homogeneous Dirichlet boundary conditions:

For given $f \in L^2(\Omega)$, find $u \in H^1(\Omega)$ such that

$$\begin{cases} -\Delta u = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \quad (3.74)$$

Since Grisvard and the above analysis, particularized with $\beta = 3\pi/2$, we know that the unique solution u of (3.74) can be written as $u = w + \varphi$ where $w \in H^2(\Omega)$ and

$$\varphi(r, \theta) = c_1 r^{2/3} \sin(2\theta/3), \quad (3.75)$$

where $\varphi \in W^{2,p}(\Omega)$ with $p \in [1; \frac{3}{2})$.

Recall that \mathcal{T}_H denotes a regular triangulation over $\bar{\Omega}$ with triangles K . We call $H = \max_{K \in \mathcal{T}_H} \text{diam}(K)$.

Since $u \in W^{2,p}(\Omega)$, $p \in [1; \frac{3}{2})$, and since Proposition 3.5 we have the following *a priori* error estimate: For $p \in (1; 3/2)$, there exists $H_0 = H_0(p, u)$ such that the approximation u_H to u satisfies

$$\|u - u_H\|_{H^1(\Omega)} \leq C H^{2-2/p} |u|_{W^{2,p}(\Omega)}^{p/2}, \quad \forall H \leq H_0, \quad (3.76)$$

where C is a constant independent of H and u but depending on p .

Hence, the *a priori* convergence order in H in the H^1 -norm is smaller than $2 - \frac{2}{3/2} = 2/3$.

Improving the *a priori* convergence order through using patches.

In this section we keep considering the model problem (3.74) introduced above. The aim is to use the correction algorithm to obtain a better *a priori* convergence order.

Complete analysis alike in Section 3.2 can be performed and lead to a particular result similar to Proposition 3.8 adapted to the current problem. We will not reiterate such a reasoning here but develop a conjecture based on the results of Section 3.2.

Consider two families of regular triangulations \mathcal{T}_H over $\bar{\Omega}$ (as above) and \mathcal{T}_h over $\bar{\Lambda}_\epsilon$, $\Lambda_\epsilon = (-\epsilon; \epsilon)^2 \cap \Omega$, with $H/\epsilon \rightarrow 0$. Recall that $H = \max_{K \in \mathcal{T}_H} \text{diam}(K)$ and call $h = \max_{K \in \mathcal{T}_h} \text{diam}(K)$ the diameter of the triangles K . We use linear elements and approximate the solution u of (3.74) with $u_{Hh} = u_H + u_h$, u_H and u_h defined on \mathcal{T}_H and \mathcal{T}_h by Algorithm 1.3.

We conjecture the following: If u is the solution of (3.74) and $p \in (1; 3/2)$, then there exists C and h_0 such that the approximation u_{Hh} to u satisfies the *a priori* error estimate

$$\|u - u_{Hh}\|_{H^1(\Omega)} \leq C \left(\frac{H}{\epsilon^{1/3}} + h^{2-2/p} + \frac{H^2}{\epsilon^{5/6} h^{1/2}} \right), \quad \forall h \leq h_0, H/\epsilon \rightarrow 0, \quad (3.77)$$

where C and h_0 are constants independent of H , h and ϵ but depending on p and u .

It is of interest to write out (3.77) when $p \rightarrow 3/2$ and $\epsilon = \alpha H^\beta$, where α and β denote constants, $\beta < 1$. When h is small enough, we have:

$$\|u - u_{Hh}\|_{H^1(\Omega)} \leq C \left(H^{1-\beta/3} + h^{2/3} + H^{2-5\beta/6}/h^{1/2} \right). \quad (3.78)$$

Hence when ϵ is proportional to H^β , we choose h proportional to $H^{3/2-\beta/2}$ to optimize the estimate and obtain a convergence of order $1 - \beta/3$ in H .

Numerical results.

In order to assess the given error estimates, we consider the problem of approximating $u = w + \varphi$ with $w = 0$ and φ given by 3.75. Numerical tests when solving the problem of finding $u \in H^1(\Omega)$ such that

$$\begin{cases} -\Delta u = 0 & \text{in } \Omega, \\ u = \varphi|_{\partial\Omega} & \text{on } \partial\Omega, \end{cases} \quad (3.79)$$

yielding obviously $u = \varphi$, are reported in the following. An illustration of the isolines of the solution $u = \varphi$ is given in Figure 3.10.

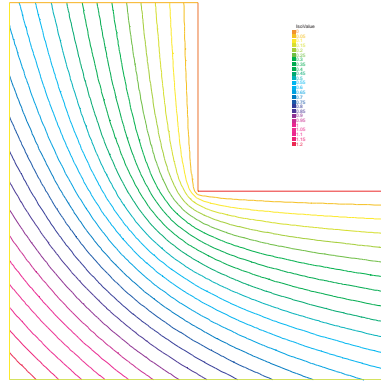


Figure 3.10: Isolines of the solution $u = \varphi$ of problem (3.79).

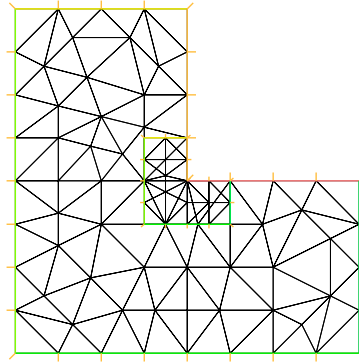


Figure 3.11: Illustration of the used grid constellation with $\epsilon = 0.25$ and $N = 8$, $M = 4$.

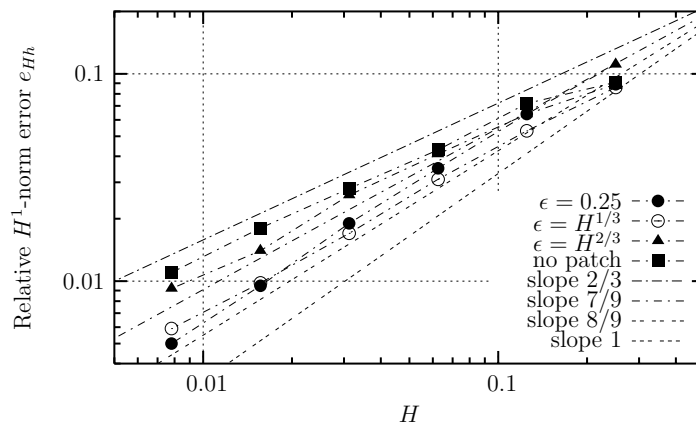
We recall the triangulations \mathcal{T}_H over $\bar{\Omega}$ and \mathcal{T}_h over $\bar{\Lambda}_\epsilon$. We consider a discretization of $(-L; L)$ with N intervals and set $L = 1$. For $(-\epsilon; \epsilon)$ we use M intervals. In Figure 3.11 we illustrate the situation by the triangulations used for $\epsilon = 0.25$ and $N = 8$, $M = 4$.

Since the above discussion, we choose $\epsilon = \alpha H^\beta$, $\beta < 1$, and accordingly $h = H^{3/2-\beta/2}$. With the above notation this is obtained when choosing $M = \alpha^{3/2} 2^{3\beta/2-1/2} N^{3/2-3\beta/2}$. Since the conjecture we expect that the extremal ($p \rightarrow 3/2$) *a priori* order of convergence in H for the $H^1(\Omega)$ -norm error is given by $1 - \beta/3$. An overview of different cases applied on the solution of problem (3.79) is reported in Table 3.2. In the last column of the latter we report the numerically obtained order in H . In Figure 3.7(a) we illustrate the concluding convergence behavior in the mesh-size graphically.

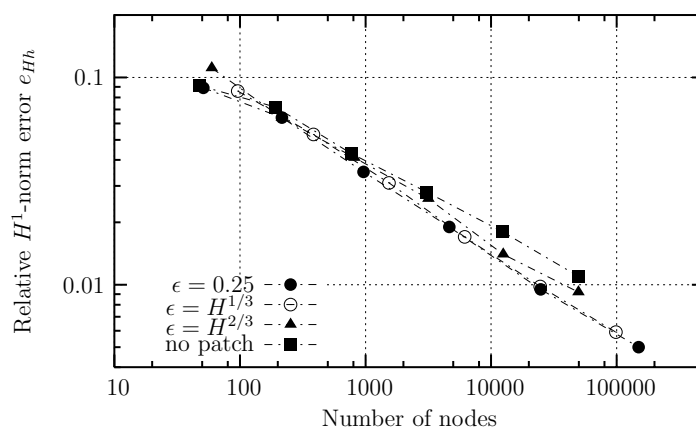
In Figure 3.7(b) we assess the error reduction with respect to the total number of nodes used.

ϵ	h	M	CO1	CO2
0.25	$H^{3/2}$	$N^{3/2}/2^{7/2}$	1	0.93
$H^{1/3}$	$H^{4/3}$	N	$8/9 \approx 0.89$	0.79
$H^{2/3}$	$H^{7/6}$	$N^{1/2}\sqrt{2}$	$7/9 \approx 0.78$	0.74
no patch			$2/3 \approx 0.67$	0.63

Table 3.2: Synoptic table of chosen patches and $H^1(\Omega)$ -norm convergence orders (CO1 = extremal *a priori* order in H , CO2 = order obtained by numerical experience).



(a) Convergence order in the mesh size.



(b) Error reduction vs. number of nodes.

Figure 3.12: Convergence of u_H resp. u_{Hh} to u with respect to the mesh size H and the number of nodes.

Chapter 4

Application to glacier modeling

The objective of this chapter is to apply the presented method to the modeling of glaciers. Considering a 2D vertical cut in the direction of the motion of the glacier and horizontal invariance, as studied by Reist [56], we present a model to simulate the velocity field of the ice mass and the effective stress. The complexity of the governing equations is reduced through approximations, and the changing boundary conditions are analyzed. With regard to the results of Section 3.2, we apply patches in certain regions on the glacier domain and show an improvement in the precision of the stress field.

The outline of this chapter is the following:

4.1	Introduction	80
4.2	Governing field equations and numerical model	82
4.3	Patches and precision of the glacier stress field	94

In Section 4.1 we give a short introduction and survey of works and models used in the study of glaciers. Basal boundary conditions are a crucial issue for accurate simulations. Section 3.2 where we studied the application of patches and the convergence order on a model problem with changing Dirichlet-Neumann boundary conditions will be used. In Section 4.2, we establish the equations that define the velocity and stress field of the glacier ice mass and present a sensible approximation (Blatter [18]). This yields a mathematical model of the glacier. Numerical issues are addressed and results on the Gries glacier (Swiss Alps) are presented. In Section 4.3 we apply patches on the problem and prove the efficiency of our method on the stress field of the Gries glacier.

4.1 Introduction

Over the last 150 years, the Swiss Alps have lost about 40% in surface and 50% in volume of their glacier ice. Today, most of alpine glaciers are retreating, leaving visible marks on the landscape. The future evolution of glaciers is not only of concern for the tourist industry, but also for agriculture and hydro-power production for which glacier ice serves as a water reservoir. Moreover, the retreat and advance of glaciers can cause natural hazards endangering humans.

The main interest for the study of glaciers, however, lies in their important active and passive role in the climate system. On one hand, the variation of the area of the global snow and ice cover changes the radiation budget and hydrological cycle of the earth. On the other hand, polar ice sheets and cold alpine glaciers represent unique archives of climatic change.

Numerical glacier modeling has become an important tool for glaciologists, which can aid in the reconstruction of the shape of past ice sheets and the mass balance of glaciers, in the simulation of the interaction between glacier and climate, in the interpretation of ice cores or to advance the understanding of the mechanical behavior of glaciers (for instance in the study of special phenomena such as calving or surging).

Glaciers are extended ice masses resting on solid land, formed through accumulation of snow over the course of millenia. Glaciers are not static object. There are two main processes determining the size and shape of the glacier over the course of time. First, climatic influences causing loss or gain of ice mass; second, the gravitational force that causes the glacier to deform under the pressure of its own weight causing it to flow down-valley.

Accumulation is called the total of all processes in which a glacier gains in mass. This usually occurs in the form of snowfall, but also by wind drifted snow and avalanches. Snow accumulated on top of the glacier is compacted to firn as new layers of snow build up on its top. Then, under increasing pressure from the above layers and by chemical process, the firn is fused into solid glacier ice. The total of all processes in which the glacier loses mass is called ablation. Ablation usually occurs in lower (warmer) elevations in the form of melting and evaporation, but can also occur in the form of calving or removal of snow by wind drift (high elevations).

Would accumulation and ablation be the only processes in glacier dynamics, the ice mass would steadily grow in the accumulation zone and shrink, and finally disappear completely, in the ablation zone. The gravitation counterbalances this effect. Ice, accumulated to sufficient depth, exert a downward force on the lower glacier layers. Under this pressure the glacier deforms viscously, allowing the glacier to flow over the glacier bed.

The glacier advances when more ice is being transported downstream than is being ablated at its terminus and retreats in the reverse case.

Ice is usually treated as incompressible viscous (in the case of cold ice) heat conducting fluid, with a Glen-type rheology [34]. The basic continuum mechanics equation for mass conservation, momentum conservation and constitutive relation for such an ice flow have been studied by Fowler and Larson [33], Hutter [44] and Morland [52].

The first numerical models have been developed for the modeling of ice sheets (glaciers covering whole land masses, e.g. Greenland or Antarctica), using simplifications that take advantage of the “shallowness” (small aspect ratio) of this ice masses. The complexity of many models require that the basic flow models be simplified for efficient numerical computations. Most ice sheet models are therefore based on the so called “shallow ice approximation” rigorously established in [33, 44, 52]. It is asymptotically valid for large parts of ice sheets, but not so in places such as ice divide, near the ice margin or close to the ice surface (see Baral et al. [16]), where the results differ largely from results obtained from higher order (in the aspect ratio of the ice mass) models. This is even more true for glaciers (e.g. small alpine glaciers) that have a relatively large aspect ratio. Therefore a number of higher order models have been proposed for inclusion of deviatoric stress gradients. They capture better some important characteristics of ice flow. Many models exist and they differ mainly on the deviatoric stress gradients that are included. We are going to use in this work a model by Blatter [18], which, as shown by Baral et al. [16] corresponds to an incomplete second order model.

Changes in the basal surface, the interface between the ice mass and the rock bed, imply different conditions for the flow. Furthermore, it is well established (refer to the discussion by Blatter [19, §3.8] for references) that the glacier base can move. Whether it is a true sliding of the glacier sole over the glacier bed or a movement of the sole on a deforming sub-glacial layer of some other material, or a combination of both, depends on local conditions at the glacier bed. Variations in basal motion are reflected in variation of the ice velocity at the glacier surface. We suppose that boundary conditions can change from free sliding to prescribed velocity, and reverse, along the basis. This directly relates to the solution for the velocity and thus for the stress field which is more or less regular. A good precision is necessary as the change from Neumann to Dirichlet conditions reflects non-locally in variations all over Ω . This is why, in Section 3.2, we have studied on a Poisson model problem our correction method and the application of patches to obtain a better precision in certain regions.

This chapter is organized as follows: The objective of Section 4.2 is to introduce the equations governing glaciers and to develop a mathematical model for calculating the velocity and stress fields. We start with a brief introduction to the model quoting the usual conservation equations for mass, momentum and a specific rheology law for ice. We describe the numerical method used and give a short review of theoretical analysis of the model. The problem consists in solving numerically a non-linear elliptic equation

with mixed Dirichlet-Neumann boundary conditions (4.33). We discretize the equation, applying a Galerkin method, using continuous, piecewise linear finite elements. We give a simple algorithm to linearize the discrete equations, using a frozen coefficient method. We recall results that establish the existence and uniqueness of the weak solution of the system and that prove the solution of the discrete linearized problem to be convergent to the exact solution. We give numerical convergence order estimates (Table 4.1). Finally, in Section 4.3, we apply our correction method to improve the precision of the effective stress field of the Gries glacier.

4.2 Governing field equations and numerical model

Ice is treated as an incompressible fluid. Its mechanical properties depend on physical quantities such as temperature and stress. At a given moment, the geometry of the ice mass is defined by the upper free surface S given by $z = S(x, y)$ which is supposed regular (i.e. no overhanging), and the basal surface B given by $z = B(x, y)$ in cartesian coordinates (x, y, z) with the z -axis pointing opposite to the direction of gravity.

In the sequel we suppose that the glacier is very large and consider a 2D vertical cut in the direction of the motion of the glacier (x -direction). We assume that none of the physical variables depend on the y -direction. In this case we say that we treat a “two”-dimensional glacier. In the sequel we restrict ourselves to the analysis of a two-dimensional glacier.

Let $[x_L; x_R]$ be the projection of the mountain base onto the x -axis. We call Γ_S the upper surface of the glacier given by the points (x, z) such that $z = S(x)$, $x \in [x_L; x_R]$, and Γ_B the mountain base, the points (x, z) such that $z = B(x)$, $x \in [x_L; x_R]$. We suppose that $B(x) < S(x)$, $\forall x \in (x_L; x_R)$. The glacier domain occupied by ice is denoted by Ω and is the set of points $\Omega = \{(x, z) \text{ such that } B(x) \leq z \leq S(x), x \in [x_L; x_R]\}$. An illustration of the introduced notation is given in Figure 4.1.

All considerations in this section, if not otherwise stated, are based on the paper from Blatter [18], the book by Hutter [44, Chapter 2] and the work [56] by Reist.

Mass conservation.

We assume that the ice is incompressible. Thus we have the usual continuity mechanics equation for mass conservation within the ice mass. If we write $\mathbf{u} = (u, w)$ for the velocity of an ice particle where u and w are the velocity components in the x and z directions respectively, the mass continuity equation $\nabla \cdot \mathbf{u} = 0$ becomes

$$\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} = 0. \quad (4.1)$$

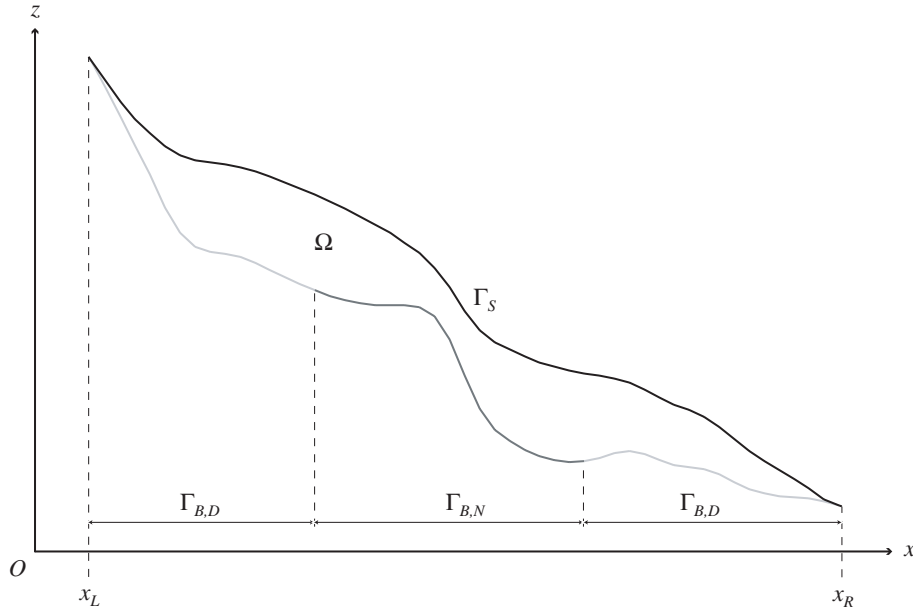


Figure 4.1: Illustration of the notation for the “two”-dimensional glacier.

Momentum conservation.

If τ is the Cauchy stress tensor, the equation for momentum conservation is $\nabla \cdot \tau + \rho \mathbf{g} = 0$, where ρ is the density of ice and $\mathbf{g} = (0, -g)$ is the acceleration of gravity. Since angular momentum conservation, we know that τ is symmetric. Thus we have the equations for linear momentum

$$\frac{\partial \tau_{xx}}{\partial x} + \frac{\partial \tau_{xz}}{\partial z} = 0, \quad (4.2)$$

$$\frac{\partial \tau_{xz}}{\partial x} + \frac{\partial \tau_{zz}}{\partial z} = \rho g, \quad (4.3)$$

where τ_{ij} are the components of the stress tensor τ .

Stress–strain relation.

The constitutive response of glacier ice to external forces depends on the physical nature of the applied force and characteristic times of the process. Under a slowly-varying state of stress applied over a very long period of time, as it is typical for glaciers, ice can be considered an incompressible viscous fluid. Glacier ice is treated as a non-Newtonian fluid. The stress–strain-rate is expressed with the stress tensor τ split into an isotropic and a deviatoric part, $\tau = -pI + \sigma$, where $p = -\frac{1}{2}(\tau_{xx} + \tau_{zz})$ is the pressure and σ the deviatoric stress tensor $\sigma = \mu(\nabla \mathbf{u}^T + \nabla \mathbf{u})$, where μ is the viscosity. The deviatoric stress

tensor is related to the velocity field through the following forms:

$$\frac{\partial u}{\partial x} = AF(\sigma)\sigma_{xx}, \quad (4.4)$$

$$\frac{1}{2} \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \right) = AF(\sigma)\tau_{xz}. \quad (4.5)$$

The term $AF(\sigma)$, with A a rate factor and $F = F(\sigma)$ a creep response function, represents a flow law.

Flow law.

Referring to Blatter [19, eqn. 156], previous work [18] and references contained therein, we use a flow law of the form

$$AF(\sigma) = A \left(\sigma_0^{n-1} + \sigma_{(II)}^{n-1} \right), \quad (4.6)$$

where $\sigma_{(II)}$ is the effective stress, i.e. the second invariant of the deviatoric tensor,

$$\sigma_{(II)}^2 = \frac{1}{2} \text{tr}(\sigma^T \sigma) = \sigma_{xx}^2 + \tau_{xz}^2, \quad (4.7)$$

where we used $\sigma_{xx} = -\sigma_{zz}$, since (4.1), (4.4) and $\frac{\partial w}{\partial z} = AF(\sigma)\sigma_{zz}$. Here σ_0 is a constant and n an exponent: Measurements for the exponent n vary between 2 and 4. For very low stresses, lower than 1 or 2 bars, the observed stress–strain-rate relation is linear, which corresponds to an exponent $n = 1$. The rate factor A is assumed to depend on temperature only for cold ice, i.e. below freezing point, and on the fraction of water (moisture content) in the water–ice mix for temperate ice at local pressure melting point. The variation of A in a typical temperate glacier, however, is small and will therefore be considered a constant.

Since $\sigma = \mu(\nabla \mathbf{u}^T + \nabla \mathbf{u})$, with (4.4) and (4.5), we have

$$AF(\sigma) = \frac{1}{2\mu}. \quad (4.8)$$

It is the flow law (4.6) that determines F , and hence the viscosity μ .

Viscosity and stress field.

Setting $\dot{\varepsilon} = \frac{1}{2}(\nabla \mathbf{u}^T + \nabla \mathbf{u})$, the constitutive relation $\sigma = \mu(\nabla \mathbf{u}^T + \nabla \mathbf{u})$ writes

$$\sigma = 2\mu\dot{\varepsilon}. \quad (4.9)$$

In order to eliminate the stress field in the flow law, we use (4.9) in (4.7) and obtain $\sigma_{(II)} = 2\mu \left(\frac{1}{2} \text{tr}(\dot{\varepsilon}^T \dot{\varepsilon}) \right)^{1/2}$. Next we set $s = \left(\frac{1}{2} \text{tr}(\dot{\varepsilon}^T \dot{\varepsilon}) \right)^{1/2}$ and consequently $\sigma_{(II)} = 2\mu s$. Hence we obtain from (4.6) and (4.8) the following implicit equation for $\mu = \mu(s)$,

$$A \left((2\mu s)^{n-1} + \sigma_0^{n-1} \right) = \frac{1}{2\mu}. \quad (4.10)$$

Since Glowinski and Rappaz [38, Lemma 1], for all $s \in \mathbb{R}_+$ and $n \geq 1$, there exists a unique $\mu \in \mathbb{R}_+$ satisfying (4.10).

Boundary conditions.

The boundary conditions for stress at the upper free surface Γ_S of the glacier are given by

$$\boldsymbol{\tau} \cdot \mathbf{n} = \mathbf{P}, \quad (4.11)$$

where $\mathbf{P} = -p\mathbf{n}$ with p denoting the atmospheric pressure and $\mathbf{n} = (n_x, n_z) = (-\partial S/\partial x, 1)$ a normal vector pointing outward from the ice domain Ω . This yields

$$n_x \tau_{xx} + n_z \tau_{xz} = -n_x p, \quad (4.12)$$

$$n_x \tau_{xz} + n_z \tau_{zz} = -n_z p. \quad (4.13)$$

Using $\sigma_{xx} = \tau_{xx} + p$, (4.12) and (4.13) write out

$$-\frac{\partial S}{\partial x} \sigma_{xx} + \tau_{xz} = 0, \quad (4.14)$$

$$-\frac{\partial S}{\partial x} \tau_{xz} + \sigma_{zz} = 0. \quad (4.15)$$

It will appear in the sequel that (4.14) corresponds to a Neumann boundary condition.

At the glacier bed Γ_B we impose the homogeneous boundary condition

$$\boldsymbol{\tau} \cdot \mathbf{n} = 0, \quad (4.16)$$

where here $\mathbf{n} = (\partial B/\partial x, -1)$ is a normal outward vector. Explicitly, since $p = 0$ this yields

$$\frac{\partial B}{\partial x} \sigma_{xx} - \tau_{xz} = 0, \quad (4.17)$$

$$\frac{\partial B}{\partial x} \tau_{xz} - \sigma_{zz} = 0. \quad (4.18)$$

Alternatively we can use Dirichlet boundary conditions, prescribing a velocity \mathbf{u} on the mountain base. Imposing $\mathbf{u} = \mathbf{0}$ means that the glacier is fixed on the mountain.

In the sequel, we adopt the following view: we denote by $\Gamma_{B,N}$ the part of the glacier bed with Neumann boundary conditions, i.e. of the first type presented (the fact that it is ‘‘Neumann’’-type will be seen below), and by $\Gamma_{B,D} = \Gamma_B \setminus \Gamma_{B,N}$ the part of the mountain base where we assume Dirichlet boundary conditions (see Figure 4.1).

First order approximation.

We aim to simplify the system of equations introduced above. To arrive at a consistent simplified set of equations, we need to estimate the order of magnitude of the various terms in the equations. Then we eliminate those that are small compared to others.

Let $\{L\}$ and $\{H\}$ denote the magnitude of the characteristic horizontal and vertical extents of the glacier. We use the fact that the aspect ratio $\epsilon = \{H\}/\{L\}$ of glaciers is

small. For glaciers ϵ is of order 10^{-2} and for ice sheets 10^{-3} . We introduce a scaling for the spatial variables as in Blatter [18] and rewrite our equations (see, e.g., [56]).

However there exist several approximations. The shallow-ice approximation consists in eliminating all terms from the equations that are of order $O(\epsilon)$ or smaller in the scaled equations. This means that the ice deforms only by shearing in horizontal planes and that longitudinal stress deviators are neglected in the force balance. In order to retain some of the deviatoric stress gradient terms, eliminated by the shallow-ice approximation, Blatter [18] proposes a slightly different approach for scaling the equations of the glacier problem. For the first order approximation we eliminate terms of order $O(\epsilon^2)$ and smaller. It is this approximation that we retain in our work. We will not present here the detailed list of the order in ϵ of all variables occurring. The latter are developed in the mentioned reference. For the understanding of the reader we point out the simplifications that are implied at each step.

Applying the first order approximation yields the following set of equations. The equation of mass conservation is unchanged from (4.1),

$$\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} = 0, \quad (4.19)$$

and momentum conservation yields

$$2 \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \tau_{xz}}{\partial z} = \rho g \frac{\partial S}{\partial x}, \quad (4.20)$$

which we obtain from (4.2) and (4.3) in the following way: We solve (4.3) for τ_{zz} by integration on z using the bounds $S(x)$ and z . Then we introduce this expression for τ_{zz} in (4.2), neglecting $\partial^2 \tau_{xz} / \partial x^2$ which is $O(\epsilon^2)$.

The constitutive relations (4.4) and (4.5) become

$$\frac{\partial u}{\partial x} = AF(\sigma_{(II)}^2) \sigma_{xx}, \quad (4.21)$$

$$\frac{\partial u}{\partial z} = 2AF(\sigma_{(II)}^2) \tau_{xz}, \quad (4.22)$$

with $\sigma_{(II)}^2 = \sigma_{xx}^2 + \tau_{xz}^2$, where we have neglected the term $\partial w / \partial x$ of order ϵ^2 .

The boundary condition for stress (4.14) on the upper surface Γ_S is

$$-2 \frac{\partial S}{\partial x} \frac{\partial u}{\partial x} + \frac{\partial u}{\partial z} = 0, \quad (4.23)$$

where we have neglected $\partial w / \partial x$ and used the relations (4.21) and (4.22). The boundary condition (4.17) on $\Gamma_{B,N}$ is

$$2 \frac{\partial B}{\partial x} \frac{\partial u}{\partial x} - \frac{\partial u}{\partial z} = 0. \quad (4.24)$$

A mathematical model of a glacier.

Following the framework of [56], using the resulting equations from the above report of the first order approximation in two dimensions we solve the equation for momentum conservation (4.20) in the given domain Ω ,

$$2\frac{\partial\sigma_{xx}}{\partial x} + \frac{\partial\tau_{xz}}{\partial z} = \rho g \frac{\partial S}{\partial x}, \quad (4.25)$$

with the constitutive relations, from (4.21) and (4.22) with (4.8),

$$\sigma_{xx} = 2\mu \frac{\partial u}{\partial x}, \quad (4.26)$$

$$\tau_{xz} = \mu \frac{\partial u}{\partial z}. \quad (4.27)$$

We point our interest only to the horizontal component of the velocity field. In fact it is the only component of use to solve, e.g., the glacier transport problem (see [56]). Hence we omit here the additional equation for the vertical component of the velocity field. Once the stress tensor known, this component can be evaluated straightforwardly.

In view to eliminate the stress field, we substitute (4.26) and (4.27) into $\sigma_{(II)}^2 = \sigma_{xx}^2 + \tau_{xz}^2$ to obtain

$$\sigma_{(II)} = 2\mu \left[\left(\frac{\partial u}{\partial x} \right)^2 + \frac{1}{4} \left(\frac{\partial u}{\partial z} \right)^2 \right]^{1/2}. \quad (4.28)$$

Setting

$$s = \left[\left(\frac{\partial u}{\partial x} \right)^2 + \frac{1}{4} \left(\frac{\partial u}{\partial z} \right)^2 \right]^{1/2}, \quad (4.29)$$

we remark that $\sigma_{(II)} = 2\mu s$, and using (4.10) yields the equation for μ ,

$$A \left((2\mu s)^{n-1} + \sigma_0^{n-1} \right) = \frac{1}{2\mu}, \quad (4.30)$$

which defines μ implicitly as a function of s , which in turn is a function of the velocity gradient. To close the system we introduce (4.26) and (4.27) into (4.25) to obtain

$$4\frac{\partial}{\partial x} \left(\mu \frac{\partial u}{\partial x} \right) + \frac{\partial}{\partial z} \left(\mu \frac{\partial u}{\partial z} \right) = \rho g \frac{\partial S}{\partial x}. \quad (4.31)$$

The boundary conditions on Γ_S and $\Gamma_{B,N}$ are given by (4.23) and (4.24) respectively. On $\Gamma_{B,D}$ we impose homogeneous Dirichlet boundary conditions $u = 0$.

We thus obtain the first order approximation of the velocity field u for a glacier Ω by solving the problem

$$\begin{cases} 4\frac{\partial}{\partial x}\left(\mu\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial z}\left(\mu\frac{\partial u}{\partial z}\right) = \rho g\frac{\partial S}{\partial x} & \text{in } \Omega, \\ -2\frac{\partial S}{\partial x}\frac{\partial u}{\partial x} + \frac{\partial u}{\partial z} = 0 & \text{on } \Gamma_S, \\ 2\frac{\partial B}{\partial x}\frac{\partial u}{\partial x} - \frac{\partial u}{\partial z} = 0 & \text{on } \Gamma_{B,N}, \\ u = 0 & \text{on } \Gamma_{B,D}, \end{cases} \quad (4.32)$$

where $\mu = \mu(s(u))$, with s given by (4.29) and μ verifying (4.30).

Finally, a rescaling of the spatial variable $\tilde{z} = 2z$ enables us to simplify the formulation of the problem. Furthermore, recalling the expression of the normal vector $\mathbf{n} = (-\partial S/\partial x, 1)$ on Γ_S , we have $-\frac{\partial S}{\partial x}\frac{\partial u}{\partial x} + \frac{\partial u}{\partial \tilde{z}} = \nabla u \cdot \mathbf{n}$. With the normal vector $\mathbf{n} = (\partial B/\partial x, -1)$ on Γ_B , we also get $\frac{\partial B}{\partial x}\frac{\partial u}{\partial x} - \frac{\partial u}{\partial \tilde{z}} = \nabla u \cdot \mathbf{n}$. Hence the problem writes out as follows:

Find u defined in Ω such that

$$\begin{cases} -\operatorname{div}(\mu(|\nabla u|)\nabla u) = f & \text{in } \Omega, \\ \frac{\partial u}{\partial n} = 0 & \text{on } \Gamma_S, \\ \frac{\partial u}{\partial n} = 0 & \text{on } \Gamma_{B,N}, \\ u = 0 & \text{on } \Gamma_{B,D}, \end{cases} \quad (4.33)$$

where

$$f = -\frac{1}{4}\rho g\frac{\partial S}{\partial x}, \quad (4.34)$$

and $\mu = \mu(|\nabla u|)$ is given by

$$A([\mu(|\nabla u|)]^{n-1}|\nabla u|^{n-1} + \sigma_0^{n-1}) = \frac{1}{2\mu(|\nabla u|)}, \quad (4.35)$$

with A and σ_0 constants, and $n \geq 1$ an exponent. Recall that with Glowinski and Rappaz [38, Lemma 1] μ is unique. We observe that for $n = 1$, $\mu(|\nabla u|) = 1/4A$, and for $n = 2$, μ is readily explicit, $\mu(|\nabla u|) = \left(-A\sigma_0 + \sqrt{A^2\sigma_0^2 + 2A|\nabla u|}\right)/2A|\nabla u|$.

The horizontal velocity field u is given by the above problem (4.33). Furthermore the effective stress field $\sigma_{(II)}$ is expressed through

$$\sigma_{(II)} = \mu(|\nabla u|)|\nabla u|. \quad (4.36)$$

We compute the velocity field of the glacier by solving numerically problem (4.33). We use a finite element method, using a computational mesh adapted to the geometry of the problem. Colinge and Rappaz prove in [31, Theorem 1] the uniqueness of the solution in the case of Dirichlet boundary conditions on all $\partial\Omega$. This result can be extended to our case of problem (4.33) with mixed Dirichlet-Neumann conditions.

Weak formulation.

We establish a weak formulation of problem (4.33). By multiplying the first equation of the latter system by φ vanishing on $\Gamma_{B,D}$, by integrating by part on Ω and taking into account the natural boundary conditions (second and third equation of (4.33)) on Γ_S and $\Gamma_{B,N}$, we obtain:

$$\int_{\Omega} \mu(|\nabla u|) \nabla u \cdot \nabla \varphi \, d\Omega = \int_{\Omega} f \varphi \, d\Omega. \quad (4.37)$$

By setting

$$a_{\mu}(u, \varphi) = \int_{\Omega} \mu(|\nabla u|) \nabla u \cdot \nabla \varphi \, d\Omega, \quad (4.38)$$

and

$$\langle f | \varphi \rangle = \int_{\Omega} f \varphi \, d\Omega, \quad (4.39)$$

we see that (4.37) is equivalent to

$$a_{\mu}(u, \varphi) = \langle f | \varphi \rangle. \quad (4.40)$$

Remark that $a_{\mu}(u, \varphi)$ is linear with respect to φ , but nonlinear with respect to u because μ depends on u . Let now V be the Banach space,

$$V = \{\varphi \in W^{1,p}(\Omega) : \varphi = 0 \text{ on } \Gamma_{B,D}\}, \quad (4.41)$$

where p is defined by

$$p = \frac{n+1}{n}, \quad (4.42)$$

with n the exponent appearing in the flow law (4.35), and $W^{1,p}(\Omega)$ denotes the usual Sobolev space of functions with first derivatives in $L^p(\Omega)$. Since property [38], there exist $c_1, c_2 > 0$ such that

$$c_1(1 + |\nabla u|)^{\frac{1}{n}-1} \leq \mu(|\nabla u|) \leq c_2(1 + |\nabla u|)^{\frac{1}{n}-1}, \quad \forall |\nabla u| \in (0; +\infty), \quad (4.43)$$

and hence $\mu(|\nabla u|) \nabla u \in L^q(\Omega)$ with $\frac{1}{p} + \frac{1}{q} = 1$ ($\Rightarrow q = n+1$), when $u \in V$. Consequently, the form (4.38) makes sense when $u, \varphi \in V$.

It follows that the weak formulation of problem (4.33) with natural boundary conditions on the upper surface Γ_S and the part $\Gamma_{B,N}$ of the basal surface, and Dirichlet conditions on the part $\Gamma_{B,D}$ of the basal surface, is:

Find $u \in V$ such that

$$a_{\mu}(u, \varphi) = \langle f | \varphi \rangle, \quad \forall \varphi \in V. \quad (4.44)$$

Discretization.

We discretize problem (4.44) by using linear finite elements. First we choose the mesh size H and consider an approximation of $\bar{\Omega}$ by a polygonal domain Ω_H with sides of length H . We denote by \mathcal{T}_H a regular triangulation of $\bar{\Omega}_H$ with triangles $K \in \mathcal{T}_H$ such that $\forall K \in \mathcal{T}_H$, $\text{diam}(K) \leq H$. We assume that for each triangle $K \in \mathcal{T}_H$, $K \cap \Gamma_{B,N,H}$ is either void or a side of K or a vertex of K , where $\Gamma_{B,N,H}$ denotes the polygonal line of $\partial\Omega_H$ joining the fixed points A and B (see Figure 4.2). We introduce

$$V_H = \{\psi \in \mathcal{C}^0(\bar{\Omega}_H) : \psi|_K \in \mathbb{P}_1(K), \forall K \in \mathcal{T}_H \text{ and } \psi = 0 \text{ on } \Gamma_{B,D,H}\}, \quad (4.45)$$

where $\Gamma_{B,D,H}$ denotes the part of $\partial\Omega_H$ corresponding to the part of the basal surface with Dirichlet boundary conditions, and call $\mu_H = \mu(|\nabla u_H|)$. We consider the following discrete problem:

Find $u_H \in V_H$ such that

$$a_{\mu_H}(u_H, \varphi_H) = \langle f | \varphi_H \rangle, \quad \forall \varphi_H \in V_H. \quad (4.46)$$

Linearization.

In order to solve numerically the discrete variational problem (4.46) we use Picard's iterative method. We apply the following steps:

- Initialization: Set $u_{H,0}$ the solution of (4.46) with $n = 1$, i.e. $\mu = \mu_H = 1/4A$.
- For $k = 1, 2, 3, \dots$ solve the following linear problem: Find $u_{H,k} \in V_H$ satisfying

$$a_{\mu_{H,k-1}}(u_{H,k}, \varphi_H) = \langle f | \varphi_H \rangle, \quad \forall \varphi_H \in V_H, \quad (4.47)$$

where $\mu_{H,k-1} = \mu(|\nabla u_{H,k-1}|)$.

A proof of convergence of this algorithm is given by Reist in [56, §2.1.1]. Hence $u_{H,k} \rightarrow u_H$ for $k \rightarrow \infty$ [56, Theorem 2.1.3].

Numerical illustration.

We illustrate our numerical model for a two-dimensional glacier on a vertical section along a flow line passing through the center of Gries glacier (Wallis, Swiss Alps). As the Gries glacier is large in the transverse direction our two-dimensional computations are relevant. The shape of the glacier [56] and the basic meshing with $N = 50$ intervals in the x -direction (and accordingly discretized in the z -direction to obtain a regular triangulation) is illustrated in Figure 4.2.

We study the linear case ($n = 1$) and the nonlinear case with $n = 2$. For defining the right-hand side f , defined in (4.34), we set $\rho g = 900 \cdot 9.81 \cdot 10^{-5}$ and in the flow law

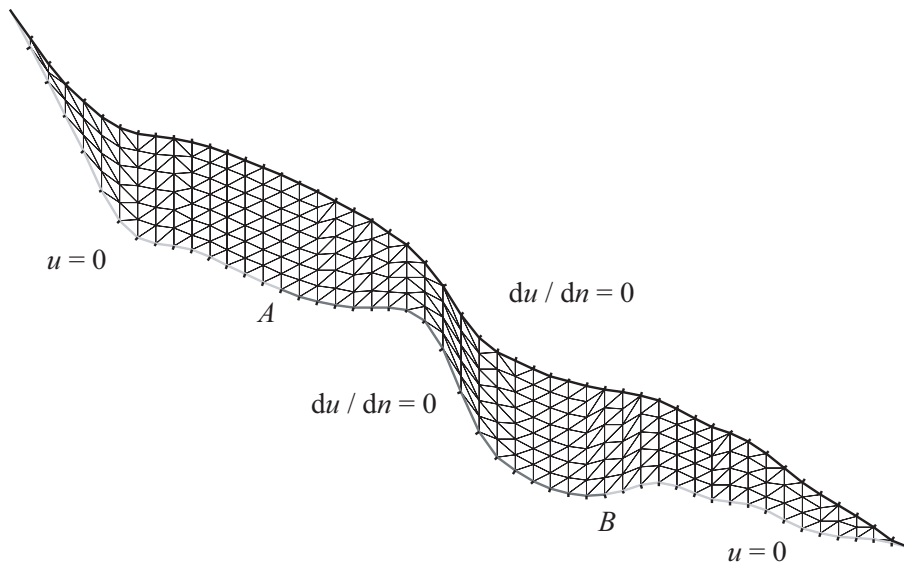


Figure 4.2: Geometry, boundary conditions and triangulation ($N = 50$) of the Gries glacier.

(4.35) we take $A = 0.08$ and $\sigma_0 = 0.1$, as suggested by [56] and converted from the S.I. unit values from Greve [39, p. 936].

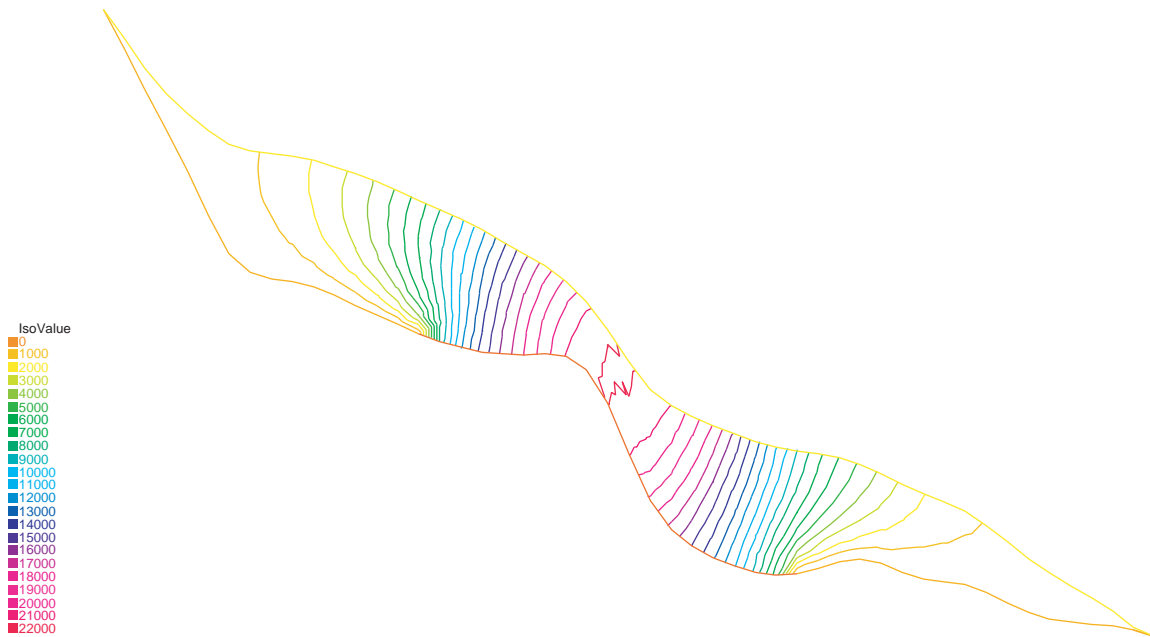
We use the relative L^2 -discrepancy of the stress field as stopping criteria for the iterations due to linearization. When running the algorithm with a tolerance 10^{-3} , we note that about 10 iterations are required to obtain convergence. We have implemented the model with the software `Freefem++` [43].

The isolines of the velocity field obtained in the nonlinear case ($n = 2$) are illustrated in Figure 4.3(a). Remark that the velocity field varies rapidly around the points where the boundary conditions change. Its gradient, directly related to the effective stress field $\sigma_{(II)}$, with (4.36), is large in the mentioned neighborhood. The latter is illustrated in Figure 4.3(b). This will give rise to a study in Section 4.3.

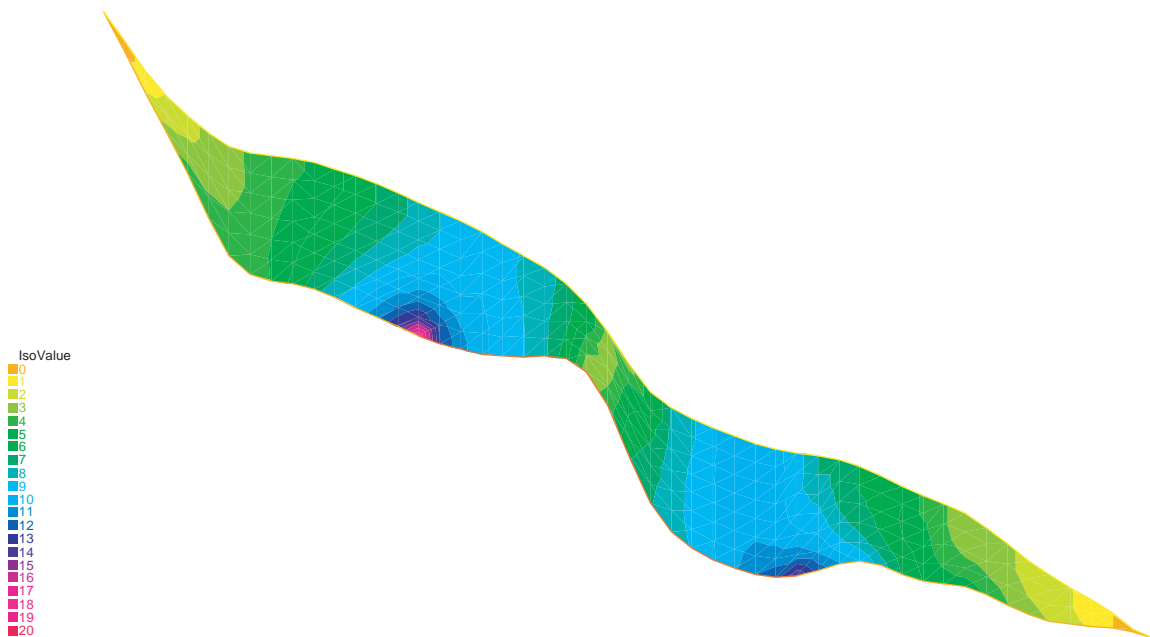
Convergence order in the mesh size.

Glowinski and Rappaz [38] supplement the analysis of the model with *a priori* error estimates. The nonlinearity of the problem introduced by μ leads us to search for a weak solution in the space $V \subset W^{1,p}(\Omega)$ where p is given by (4.42), $p = (n+1)/n$. For example in the linear case, when $n = 1$, we have $V \subset W^{1,2}(\Omega) = H^1(\Omega)$. However, when $n = 2$ we require merely $W^{1,3/2}(\Omega)$ -regularity on u .

It is of interest to analyze the convergence order in the mesh size of the grid used.



(a) Velocity field u_H .



(b) Stress field $\sigma_{(II)}(u_H)$.

Figure 4.3: Velocity and stress field of the Gries glacier in the nonlinear case ($N = 50$, $n = 2$). Numerical values are not scaled to physical units. The Figures merely illustrate the relative variations in the ice mass.

N	Linear problem ($n = 1$)	Nonlinear problem ($n = 2$)
50	0.261	0.235
100	0.204	0.154
200	0.172	0.0996
400	0.111	0.0535
800	0.0605	0.0223
1600	0.0463	0.0104
obtained order	$O(H^{0.50})$	$O(H^{0.90})$

Table 4.1: Evolution of H^1 - resp. $W^{1,3/2}$ -norm relative error on the velocity field and convergence orders.

We consider regular triangulations adapted to the glacier geometry with $N = 50$ to 3200. For evaluating the convergence behavior, we consider the approximation u_H on the grid with $N = 3200$ as “almost exact” solution u to the problem (4.44).

In Table 4.1 we report the relative errors with respect to the “almost exact” solution in the $W^{1,p}$ -norm for u_H computed on increasingly finer meshes in both linear ($n = 1$, $p = 2$) and nonlinear ($n = 2$, $p = 3/2$) cases. In the linear case we obtain as expected (since Section 3.2) a convergence of order 0.5 in H . However in the nonlinear case with $n = 2$ the convergence in H is beyond any available *a priori* results. In fact, since the Dirichlet-Neumann change in the boundary conditions we can expect $u \in W^{2,r}(\Omega)$ for any $r \in [1; 4/3)$ (case of the linear problem studied in Section 3.2). Since the Sobolev injections, we have that if $u \in W^{2,r}(\Omega)$ for any $r \in [1; 4/3)$, then $u \in W^{1,\sigma}(\Omega)$, for any $\sigma \in [1; 4)$. Hence we should have $u \in W^{1,2}(\Omega) = H^1(\Omega)$ which implies (see [38] and Proposition 3.5), $\|u - u_H\|_{1,p,\Omega} \leq C \min_{\varphi_H \in V_H} \|u - \varphi_H\|_{1,2,\Omega} \leq C\sqrt{H}$, for H small enough, where C denotes a generic constant independent of H and noting that the last inequality is not rigorous as obtained when $r \rightarrow 4/3$. Since we use the norm $\|\cdot\|_{1,p,\Omega}$ the order in H can be expected to be larger than 0.5. However the smoothing role of μ around A and B should allow u to be less regular. In conclusion, we cannot present any theoretical answer regarding the *a priori* convergence order.

Stress field on the mountain base.

Let us now turn to the stress field of the glacier (Figure 4.3(b)) and consider the nonlinear case. Before studying how to increase the precision on the approximation of the latter, we briefly outline its behavior along the mountain base, i.e. the boundary Γ_B . Our knowledge of the behavior of the velocity field in the case of the (linear) Poisson problem with changing Dirichlet-Neumann boundary conditions (Section 3.2), and the result (4.43) meaning that μ behaves like $|\nabla u|^{\frac{1}{n}-1}$ when $|\nabla u| \rightarrow \infty$, leads us to conjecture that the

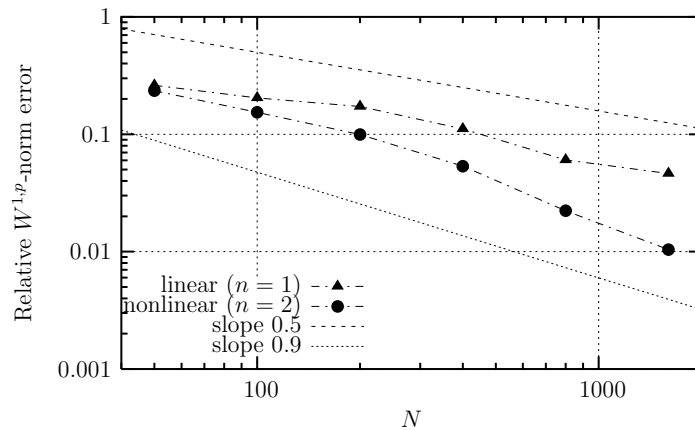


Figure 4.4: Convergence orders for the velocity field on the models with $n = 1, 2$ of the Gries glacier.

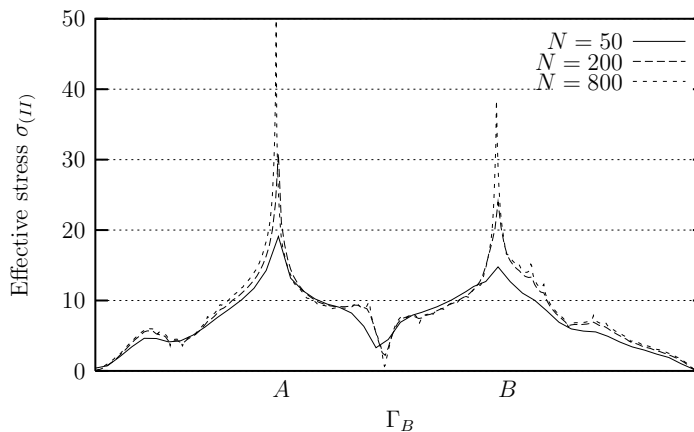


Figure 4.5: Behavior of the stress field $\sigma_{(II)}(u_H)$ on Γ_B in the nonlinear case $n = 2$. Numerical values are not scaled to physical units. The Figure merely illustrates the relative behavior for different N .

effective stress field $\sigma_{(II)}$ blows up at the points A and B . This is confirmed by the illustration in Figure 4.5 where the values of $\sigma_{(II)}$ are plotted along Γ_B in the case $n = 2$. A good precision is needed around A and B where the stress field blows up. In the next section, we consider applying patches in these small regions.

4.3 Patches and precision of the glacier stress field

We consider in this section the stress field of the glacier (Figure 4.3(b)) and reconsider the application of patches. We show on the example of the Gries glacier and the nonlinear model with $n = 2$ that the application of patches is efficient for reducing the error around the points of changing boundary conditions on the basal surface.

The mathematical model and the numerical approach are discussed in the previous section. The end of the last section concluded with the computation of the velocity and stress fields of the Gries glacier. The stress field is directly related to the velocity. The isolines of the effective stress field $\sigma_{(II)}$ obtained in the nonlinear case ($n = 2$), given by (4.36), are illustrated in Figure 4.3(b). Since the previous study of the effective stress field (Figure 4.5), we know that the stress field blows up around the points of the boundary where the boundary conditions change.

With (4.36), $\sigma_{(II)} = \mu(|\nabla u|)|\nabla u|$, and (4.43), the minimal regularity requirement on the velocity field $u \in V$ a subset of $W^{1,p}(\Omega)$ with $p = (n + 1)/n$, suffices to have $\sigma_{(II)} \in L^{1+n}(\Omega)$, i.e. at least $L^2(\Omega)$, since $n \geq 1$. Hence the stress field is L^2 -integrable, and the L^2 -norm is adapted to measure it.

In our model, the boundary of the glacier domain Ω has two points on the basal surface where it presents a Dirichlet-Neumann change in the boundary conditions (see Figure 4.1). We call A and B these points. Furthermore, we apply different patches with $(M + 1)(M/2 + 1)$ discretization points in the regions around A and B in order to see how to sharpen the results.

We consider two types of patches as in Section 3.2. A first type, called fixed patch Λ_A, Λ_B , which in the original mesh with $N = 50$ points covers exactly one triangle in all directions around A and B , and a second type which covers, at any refinement level, one triangle in all directions: it is what we call a variable patch. An illustration of fixed patches and their size for the refined mesh with $N = 50$ and $M = 4$, i.e. $H/h = 4$, is depicted in Figure 4.6. For the triangulations of Λ_A and Λ_B , we consider refinements such that the diameter h of the fine triangulation is $h = H/2, H/4$ or $H/8$.

As mentioned previously, about 10 iterations are necessary to cope with the iterative method for the linearized problem. Currently, we have an inside loop corresponding to the iterations of the correction algorithm. Requiring the relative discrepancy of the velocity field in the L^2 -norm to be below 10^{-2} (see Section 2.3), the inner-loop convergence needs 2 to 3 full iterations.

We measure the $W^{1,3/2}$ - and L^2 -norm relative error ($n = 2$) respectively of the velocity u_{Hh} and the stress field $\sigma_{Hh} = \sigma_{(II)}(u_{Hh})$ with respect to the ‘‘almost exact’’ solution u given by u_H with $N = 3200$ and $\sigma = \sigma_{(II)}(u)$. We evaluate the norms in domains Λ_A and Λ_B for $N = 50$ and 100, and for situations without patch and with patch. For the situation where we apply a patch we consider either two fixed patches or two variable patches (smaller for $N = 100$) around A and B . In each case, the error is evaluated in the region covered by the patch. The respective values are reported in Table 4.2.

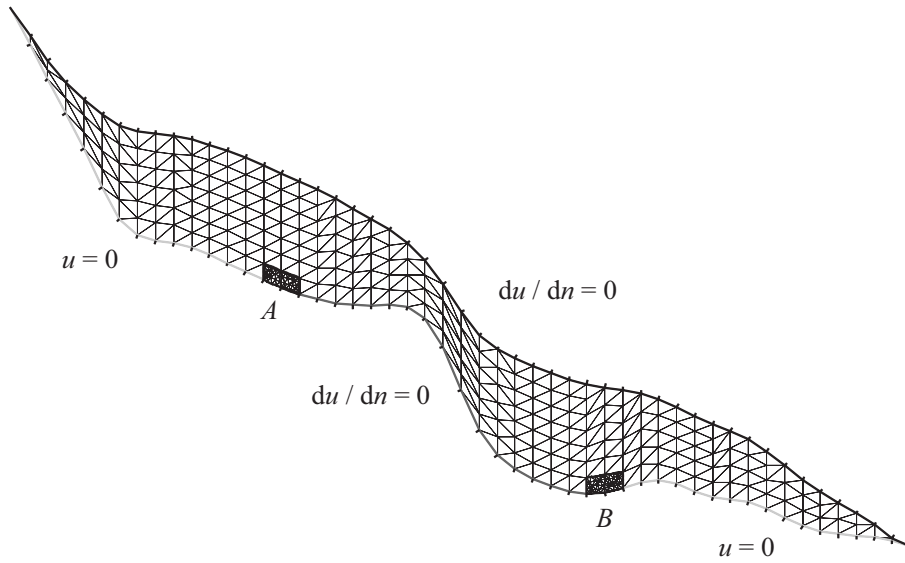


Figure 4.6: Grid constellation for the Gries glacier ($N = 50$) and illustration of two patches ($M = 4$, $H/h = 4$) around the points A and B .

N	M		$\frac{\ u - u_{Hh}\ _{W^{1,3/2}(\Lambda_A)}}{\ u\ _{W^{1,3/2}(\Lambda_A)}}$	$\frac{\ \sigma - \sigma_{Hh}\ _{L^2(\Lambda_A)}}{\ \sigma\ _{L^2(\Lambda_A)}}$	$\frac{\ u - u_{Hh}\ _{W^{1,3/2}(\Lambda_B)}}{\ u\ _{W^{1,3/2}(\Lambda_B)}}$	$\frac{\ \sigma - \sigma_{Hh}\ _{L^2(\Lambda_B)}}{\ \sigma\ _{L^2(\Lambda_B)}}$
50	–	no patch	0.734	0.264	0.415	0.310
	2	$H/h = 2$	0.681	0.194	0.341	0.237
	4	$H/h = 4$	0.662	0.168	0.220	0.170
	8	$H/h = 8$	0.648	0.184	0.256	0.155
<i>Fixed patch, large region</i>						
100	–	no patch	0.643	0.201	0.272	0.215
	4	$H/h = 2$	0.620	0.171	0.200	0.134
	8	$H/h = 4$	0.613	0.181	0.265	0.120
	16	$H/h = 8$	0.636	0.164	0.365	0.111
<i>Variable patch, small region</i>						
100	–	no patch	0.818	0.265	0.413	0.290
	2	$H/h = 2$	0.780	0.225	0.404	0.220
	4	$H/h = 4$	0.773	0.239	0.405	0.155
	8	$H/h = 8$	0.791	0.232	0.592	0.193

Table 4.2: $W^{1,3/2}$ - and L^2 -norm relative error of the velocity and the stress fields in the patches (with $n = 2$).

Considering the data reported in Table 4.2, we note that the reduction of the error, particularly for the stress field, is efficient. If we evaluate the errors in *variable* regions, though refining the grids by a factor 2, from $N = 50$ to 100 with no patch for example, we note no considerable change in the value of the error as the domain for evaluation has in the same time tightened around the point of singular behavior.

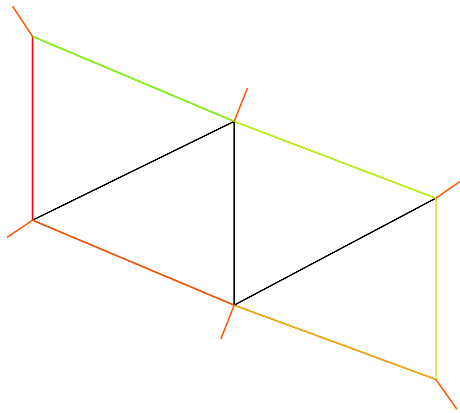
We conclude with a precision gain for the stress field due to the application of patches. Taking for example the case $N = 50$ with no patch and a patch with $M = 8$ around B , we report that the error on the stress field is divided by two! Furthermore, we can also compare the convincing pair of situations of $N = 50$ with a patch ($M = 4$ or 8) and of $N = 100$ with no patch. While using much less discretization points, applying a small, well chosen patch yields better results than a global refinement.

In Figure 4.7 we illustrate graphically the improvement of the solution in the patches applied as in Figure 4.6. Considering the region around the point A , we show in Figure 4.7(a) the actual meshing of the global mesh with $N = 50$. In (b) we show the patch that we apply over the whole region ($M = 4$). For both situations (a) without patch, and (b) with the patch and underlying triangulation as in (a), we illustrate the obtained results for the velocity and effective stress field in the region. In Figures (c) and (d) the improvement of the first derivative of the velocity field around the point of changing boundary conditions can be observed qualitatively. A relative comparison of the results (e) and (f) shows how the patch improves the quality of the solution around the change of boundary conditions where the stress field explodes. The improvement goes continuously beyond the region of the patch as it can be seen from the values of the stress field at the boundary of the region.

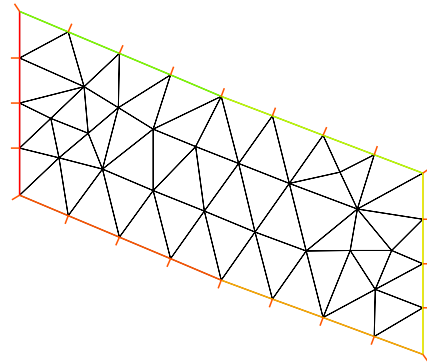
Finally, the question of the optimal size of the patch arises. To give an answer to this point, we consider the coarse discretization with $N = 50$ and consider the relative $L^2(\Omega)$ -norm error of the effective stress field with respect to the above introduced “almost exact” solution.

Without any patch, this error yields 0.153. Applying patches around A and B with the size as shown in Figure 4.6 (i.e. covering one coarse triangle in each direction around the point A or B , the patch is large by $2H$) and a ratio $H/h = 2$, we reduce the error and get 0.121.

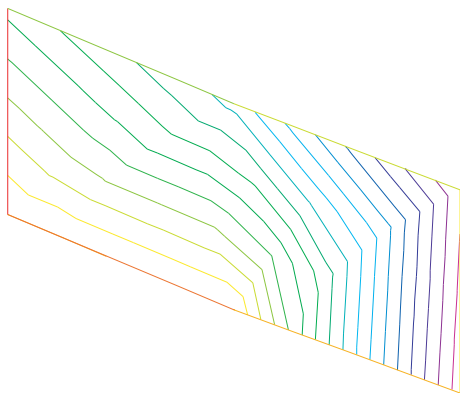
If we want to further improve the solution, we can either enlarge the regions over which we consider the patches, or refine the chosen patches. Refining the patches such that $H/h = 8$, reduces the error to the value 0.110. However applying patches around A and B that are $8H$ large (four times larger, four times higher), keeping the ratio $H/h = 2$ constant, the relative error does not improve: 0.126. This means that in this particular situation, the size of the patch is accurately chosen small, covering the region where the solution varies most.



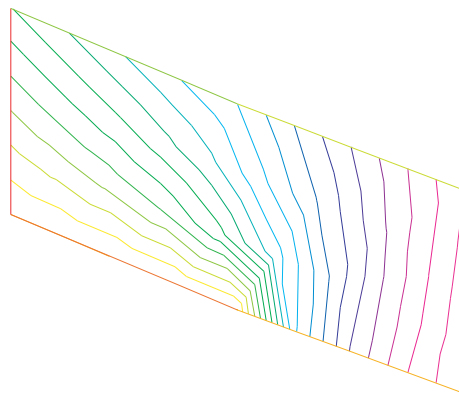
(a) Coarse mesh around A ($N = 50$).



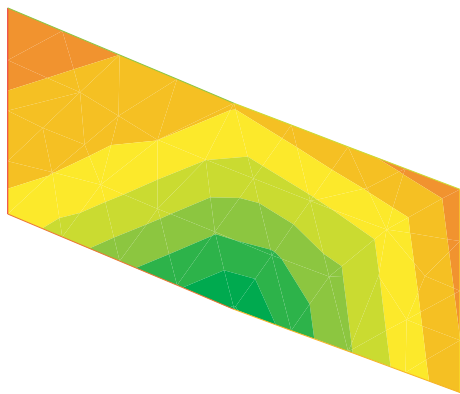
(b) Patch with $H/h = 4$ ($M = 4$).



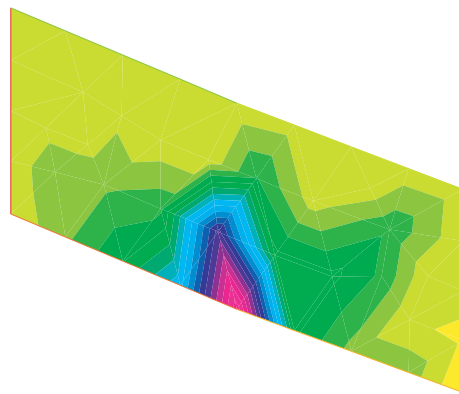
(c) Velocity field u_H .



(d) Velocity field u_{Hh} .



(e) Stress field σ_H .



(f) Stress field σ_{Hh} .

Figure 4.7: Mesh (a,b), velocity field (c,d) and stress field (e,f) in the patched region around A without and with patch $H/h = 4$.

As a conclusion of this chapter, we retain that the application of patches improves the local precision. Choosing the patch and its refinement is finally dictated by adequation depending on what we are seeking and the allowed size for problems.

Conclusion

Multi-scale problems have been investigated from a theoretical and numerical point of view.

On one hand, we have introduced a new method for numerically solving multi-scale problems. The introduced algorithm uses multiple levels of not necessary conforming grids. We calculate successive corrections to the solution in sub-domains where more precision is provided when applying patches of finite elements. We have compared the method to existing iterative methods and have concluded that our relaxed method is more flexible. We have mathematically analyzed its convergence through a spectral analysis of the iteration operator. We have carefully analyzed the parameters to optimize the convergence of the method. We have discussed implementation issues and illustrated the convergence behavior on a model situation. We have provided concluding results with regard to the usage of memory and computation time.

On the other hand, we have applied the correction method to various problems. We have studied the efficiency with respect to the improvement of the the precision and of the convergence order in the mesh size. Results are supplied for Laplace model problems: we have detailed a problem with changing Dirichlet-Neumann and a problem in a polygonal domain with entrant corner. We have also considered the modeling of glaciers. On a model simulating the stress field in the ice mass we have improved the numerical results by applying patches in the regions where changes in the basal boundary conditions are involved.

We conclude that the method presented in this work is particularly efficient when applying adequately sized and refined patches. An *a posteriori* error estimator can automatize this choice. Note that we have developed a general method in d dimensions, $d = 2, 3$. Although all examples and applications in this thesis are in two dimensions, applications in three dimensions of our algorithm can be developed straightforwardly. For this we refer to ongoing work by Rezzonico [57]. The correction method is efficient and due to its flexibility it is, particularly in three dimensions, a clear response to the issue of re-meshing.

Bibliography

- [1] B. Achchab, S. Achchab, O. Axelsson, and A. Souissi. Upper bound of the constant in strengthened C.B.S. inequality for systems of linear partial differential equations. *Numer. Algorithms*, 32:185–191, 2003.
- [2] B. Achchab, O. Axelsson, L. Laayouni, and A. Souissi. Strengthened Cauchy-Bunyakowski-Schwarz inequality for a three-dimensional elasticity system. *Numer. Linear Algebra Appl.*, 8:191–205, 2001.
- [3] B. Achchab and J.-F. Maître. Estimate of the constant in two strengthened C.B.S. inequalities for F.E.M. systems of 2D elasticity: Application of multilevel methods and a posteriori error estimators. *Numer. Linear Algebra Appl.*, 3(2):147–159, 1996.
- [4] Y. Achdou and Y. Maday. The mortar element method with overlapping subdomains. *SIAM J. Numer. Anal.*, 40(2):601–628, 2002.
- [5] R.A. Adams and J.J.F. Fournier. *Sobolev Spaces*, volume 140 of *Pure and Applied Mathematics*. Elsevier Science, Oxford, 2nd edition, 2003.
- [6] O. Axelsson. On multigrid methods of the two-level type. In W. Hackbusch and U. Trottenberg, editors, *Multigrid Methods*, volume 960 of *Lecture Notes in Mathematics*, pages 352–367. Springer-Verlag, Berlin, 1981.
- [7] O. Axelsson. Stabilization of algebraic multilevel iteration methods; additive methods. *Numer. Algorithms*, 21:23–47, 1999.
- [8] O. Axelsson and R. Blaheta. Two simple derivations of universal bounds for the C.B.S. inequality constant. Report No. 0133, Dept. of Mathematics, University of Nijmegen, The Netherlands, 2001.
- [9] O. Axelsson and I. Gustafsson. Preconditioning and two-level multigrid methods of arbitrary degree of approximation. *Math. Comp.*, 40(161):219–242, 1983.
- [10] O. Axelsson and P.S. Vassilevski. Algebraic multilevel preconditioning methods. I. *Numer. Math.*, 56:157–177, 1989.
- [11] O. Axelsson and P.S. Vassilevski. Algebraic multilevel preconditioning methods. II. *SIAM J. Numer. Anal.*, 27(6):1569–1590, 1990.

- [12] B. Philip D.J. Quinlan B. Lee, S.F. McCormick. Asynchronous Fast Adaptive Composite-grid methods for elliptic problems: Theoretical foundations. *SIAM J. Numer. Anal.*, 42:130–152, 2003.
- [13] I. Babuška and W.C. Rheinboldt. Error estimates for adaptive finite element computations. *SIAM J. Math. Anal.*, 21:736–754, 1978.
- [14] R.E. Bank, T.F. Dupont, and H. Yserentant. The hierarchical basis multigrid method. *Numer. Math.*, 52:427–458, 1988.
- [15] R.E. Bank and R.K. Smith. A posteriori error estimates based on hierarchical bases. *SIAM J. Numer. Anal.*, 30(4):921–935, 1993.
- [16] D.R. Baral, K. Hutter, and R. Greve. Asymptotic theories of large-scale motion, temperature, and moisture distribution in land-based polythermal ice sheets: a critical review and new developments. *Appl. Mech. Rev.*, 54:215–256, 2001.
- [17] R. Blaheta. Space decomposition preconditioners and parallel solvers. In M. Feistauer and al., editors, *Numerical Mathematics and Advanced Applications*, pages 20–38. Springer-Verlag, 2004.
- [18] H. Blatter. Velocity and Stress Fields in Grounded Glaciers: A Simple Algorithm for Including Deviatoric Stress Gradients. *J. Glac.*, 41(138):333–344, 1995.
- [19] H. Blatter. Physical and mathematical basis for glacier dynamics. Technical report, Institute for Atmospheric and Climate Science, ETHZ, Zurich, Switzerland, 2003.
- [20] D. Braess. The contraction number of a multigrid method for solving the Poisson equation. *Numer. Math.*, 37:387–404, 1981.
- [21] D. Braess. The convergence rate of a multigrid method with Gauss-Seidel relaxation for the Poisson equation. In W. Hackbusch and U. Trottenberg, editors, *Multigrid Methods*, volume 960 of *Lecture Notes in Mathematics*, pages 368–386. Springer-Verlag, Berlin, 1981.
- [22] D. Braess. *Finite elements: Theory, fast solvers, and applications in solid mechanics*. Cambridge University Press, Cambridge, 2nd edition, 2001.
- [23] J.H. Bramble, R.E. Ewing, J.E. Pasciak, and A.H. Schatz. A preconditioning technique for the efficient solution of problems with local grid refinement. *Comput. Methods Appl. Mech. Engrg.*, 67:149–159, 1988.
- [24] J.H. Bramble, J.E. Pasciak, J. Wang, and J. Xu. Convergence estimates for product iterative methods with applications to domain decomposition. *Math. Comp.*, 57(195):1–21, 1991.

-
- [25] F. Brezzi, J.-L. Lions, and O. Pironneau. Analysis of a chimera method. *C. R. Acad. Sci. Paris, Ser. I*, 332:655–660, 2003.
- [26] X.-C. Cai and O.B. Widlund. Multiplicative Schwarz algorithms for some nonsymmetric and indefinite problems. *SIAM J. Numer. Anal.*, 30(4):936–952, 1993.
- [27] T.F. Chan, B.F. Smith, and J. Zou. Overlapping Schwarz methods on unstructured meshes using non-matching coarse grids. *Numer. Math.*, 73:149–167, 1996.
- [28] P.G. Ciarlet. *Numerical Analysis of the Finite Element Method*. Séminaire de Mathématiques Supérieures. Les Presses de l’Université de Montréal, Canada, 1976.
- [29] P.G. Ciarlet. Basic error estimates for elliptic problems. In P.G. Ciarlet and J.L. Lions, editors, *Finite Element Methods (Part 1)*, volume II of *Handbook of Numerical Analysis*, pages 17–351. North-Holland, 1991.
- [30] Ph. Clément. Approximation by finite element functions using local regularization. *RAIRO Anal. Numér.*, 9:77–84, 1978.
- [31] J. Colinge and J. Rappaz. A Strongly Nonlinear Problem Arising in Glaciology. *M2AN*, 33(2):395–406, 1999.
- [32] V. Eijkhout and P. Vassilevski. The role of the strengthened Cauchy-Buniakowskii-Schwarz inequality in multilevel methods. *SIAM Rev.*, 33(3):405–419, 1991.
- [33] A.C. Fowler and D.A. Larson. On the flow of polythermal glaciers. I. Model and preliminary analysis. *Proc. R. Soc. London, Ser. A*, 363:217–242, 1978.
- [34] J. Glen. The creep of polycrystalline ice. *Proc. R. Soc. London, Ser. A*, 228(1175):519–538, 1955.
- [35] R. Glowinski, J. He, A. Lozinski, M. Picasso, J. Rappaz, V. Rezzonico, and J. Wagner. Finite element methods with patches and applications. In *Proceedings of the 16th International Conference on Domain Decomposition Methods*, Lecture Notes in Computational Science and Engineering. Springer, 2005.
- [36] R. Glowinski, J. He, A. Lozinski, J. Rappaz, and J. Wagner. Finite element approximation of multi-scale elliptic problems using patches of elements. *Numer. Math.*, 101(4):663–687, 2005.
- [37] R. Glowinski, J. He, J. Rappaz, and J. Wagner. Approximation of multi-scale elliptic problems using patches of finite elements. *C. R. Acad. Sci. Paris, Ser. I*, 337:679–684, 2003.
- [38] R. Glowinski and J. Rappaz. Approximation of a Nonlinear Elliptic Problem in a Non-Newtonian Fluid Flow Model in Glaciology. *M2AN*, 37(1):175–186, 2003.

- [39] R. Greve. A continuum-mechanical formulation for shallow polythermal ice sheets. *Phil. Trans. R. Soc. Lond., Ser. A*, 355:921–974, 1997.
- [40] M. Griebel and P. Oswald. On the abstract theory of additive and multiplicative Schwarz algorithms. *Numer. Math.*, 70:163–180, 1995.
- [41] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*, volume 24 of *Monographs and Studies in Mathematics*. Pitman, 1985.
- [42] P. Grisvard. *Singularities in Boundary Value Problems*, volume 22 of *Research Notes in Applied Mathematics*. Masson, 1992.
- [43] F. Hecht, O. Pironneau, and K. Ohtsuka. **Freefem++**. <http://www.freefem.org>.
- [44] K. Hutter. *Theoretical Glaciology: Material Science of Ice and the Mechanics of Glaciers and Ice Sheets*. Mathematical Approaches to Geophysics. D. Reidel Publishing Company / Terra Scientific Publishing Company, 1983.
- [45] M. Jung and J.-F. Maître. Some remarks on the constant in the strengthened CBS inequality: Estimate for hierarchical finite element discretizations of elasticity problems. *Numer. Methods Partial Differential Equations*, 15(4):469–487, 1999.
- [46] M.R. Laydi. Convergence d’un schéma itératif explicite. *C. R. Acad. Sci. Paris, Ser. I*, 326:511–514, 1998.
- [47] J.-F. Maître and F. Musy. The contraction number of a class of two-level methods; an exact evaluation for some finite element subspaces and model problems. In W. Hackbusch and U. Trottenberg, editors, *Multigrid Methods*, volume 960 of *Lecture Notes in Mathematics*, pages 535–544. Springer-Verlag, Berlin, 1981.
- [48] S.D. Margenov. Upper bound of the constant in the strengthened C.B.S. inequality for FEM 2D elasticity equations. *Numer. Linear Algebra Appl.*, 1(1):65–74, 1994.
- [49] S.F. McCormick. Fast Adaptive Composite grid (FAC) methods: Theory for the variational case. In K. Böhner and H.J. Stetter, editors, *Defect Correction Methods. Theory and Applications*, volume 5 of *Computing Supplementum*, pages 115–121. Springer, 1984.
- [50] S.F. McCormick and J.W. Ruge. Unigrid for multigrid simulation. *Math. Comp.*, 41(163):43–62, 1983.
- [51] S.F. McCormick and J. Thomas. The Fast Adaptive Composite grid (FAC) method for elliptic equations. *Math. Comp.*, 46(174):439–456, 1986.
- [52] L.W. Morland. Thermo-mechanical balances of ice sheet flows. *Geophys. Astrophys. Fluid Dyn.*, 29:237–266, 1984.

-
- [53] A. Quarteroni and A. Valli. *Domain Decomposition Methods for Partial Differential Equations*. Numerical Mathematics and Scientific Computation. Oxford Science Publications, Oxford, 1999.
- [54] J. Rappaz. Numerical approximation of PDEs and Clément’s interpolation. In *Proceedings of the Conference “Partial Differential Equations and Functional Analysis”*, OT Series. Birkhauser, to appear in 2006.
- [55] G. Raugel. Résolution numérique par une méthode d’éléments finis du problème de Dirichlet pour le laplacien dans un polygone. *C. R. Acad. Sci. Paris, Ser. A*, 286:791–794, 1978.
- [56] A. Reist. *Mathematical analysis and numerical simulation of the motion of a glacier*. PhD thesis, EPFL, Lausanne, Switzerland, 2005.
- [57] V. Rezzonico. PhD thesis, EPFL, Lausanne, Switzerland, to appear.
- [58] H.A. Schwarz. *Ueber einen Grenzübergang durch alternirendes Verfahren*, volume 2 of *Gesammelte Mathematische Abhandlungen*, pages 133–143. Chelsea Publishing Company, Bronx, New York, 1972.
- [59] B. Smith and W. Gropp P. Bjørstad. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, Cambridge, 1996.
- [60] J.L. Steger and J.A. Benek. On the use of composite grid schemes in computational aerodynamics. *Comput. Methods Appl. Mech. Engrg.*, 64:301–320, 1987.
- [61] J. Wang. Convergence analysis of the Schwarz algorithm and multilevel decomposition iterative methods II: Nonselfadjoint and indefinite elliptic problems. *SIAM J. Numer. Anal.*, 30(4):953–970, 1993.
- [62] O.B. Widlund. Domain decomposition methods for elliptic partial differential equations. In H. Bulgak and C. Zenger, editors, *Error control and adaptativity in scientific computing*, pages 325–354. Kluwer Academic Publishers, 1999.
- [63] B.I. Wohlmuth. *Discretization methods and iterative solvers based on domain decomposition*, volume 17 of *Lecture Notes in Computational Science and Engineering*. Springer, 1991.
- [64] J. Xu. Iterative methods by space decomposition and subspace correction. *SIAM Rev.*, 34(4):581–613, 1992.

- [65] J. Xu. An introduction to multilevel methods. In M. Ainsworth, J. Levesley, W.A. Light, and M. Marletta, editors, *Wavelets, multilevel methods and elliptic PDEs*, Numerical Mathematics and Scientific Computation, pages 213–302. Oxford University Press, 1997.
- [66] J. Xu. The method of subspace corrections. *J. Comput. Appl. Math.*, 128:335–362, 2001.
- [67] J. Xu and L. Zikatanov. The method of alternating projections and the method of subspace corrections in Hilbert space. *PennState, Department of Mathematics*, AM223, 2000.
- [68] K. Yosida. *Functional Analysis*, volume 123 of *A Series of Comprehensive Studies in Mathematics*. Springer-Verlag, Berlin, 4th edition, 1974.
- [69] H. Yserentant. On the multi-level splitting of finite element spaces. *Numer. Math.*, 49:379–412, 1986.
- [70] H. Yserentant. Hierarchical bases. In R.E. O’Malley, editor, *Proceedings of the Second International Conference on Industrial and Applied Mathematics*, pages 256–276. SIAM, 1991.
- [71] H. Yserentant. Old and new convergence proofs for multigrid methods. *Acta Numer.*, pages 285–326, 1993.

Appendix

On a class of solutions of the continuity and Euler's equations for inviscid and compressible fluids

I devote this appendix to a small review of research done in Theoretical Physics, trying to expand and apply the results of my master thesis [A]. In collaboration with Hon. Prof. Philippe Choquard, I continued during the last years very promising research based on my multidisciplinary master thesis in the domains of Numerical Analysis, Hydrodynamics and Theoretical Physics. The general objective of this research can be formulated as a description of the dynamics of conservative, very large and dense systems experiencing strongly correlated motion of their constituents which interact via long range potentials, and this, by means of canonically conjugated collective variables. This work yielded very interesting results and lead to several publications. The objective of the following is to give an introduction to the work done and published in [B,C,D].

Preamble.

The papers [B], [C] and [D] report results of computer simulations and of exact theoretical analysis concerning certain classes of implicit solutions of Bernoulli's equation for the velocity potential $S(x, t)$ and of the continuity equation for the mass density $\rho(x, t)$, two canonically conjugated collective variables, of Hamiltonian fluids in one dimension, perfect and with Newtonian and Coulombian self-interactions.

These classes are associated to a particular choice of correlated initial conditions between the velocity potentials and the mass densities which are that $\rho(y, 0) \propto S_{yy}(y, 0)$ and which have a quantum theoretical origin. The latter lies in a similitude between, on the one hand, the Bernoulli equation for the velocity potential $S(x, t)$ and the continuity equation for the mass density $\rho(x, t)$ of an inviscid and compressible perfect liquid in 1D, and, on the other hand, the Hamilton–Jacobi equation for the action function $J(x, t)$ and the continuity equation for the particle density $n(x, t)$ associated to the semi-classical approximation $\psi_{sc}(x, t)$ of a Schrödinger wave function $\psi(x, t) = (n(x, t))^{1/2} \exp(iJ(x, t)/\hbar)$

of a free particle of mass m in \mathbb{R}^1 . The key observation is that $n(x, t) \propto J_{xx}(x, t)$ is a solution of $n_t + \frac{1}{m}(nJ_x)_x = 0$ if J is a solution of $J_t + \frac{1}{2m}J_x^2 = 0$. This implies that $\rho(x, t) \propto S_{xx}(x, t)$ is also an admissible solution in the classical case, as emphasized in [B] and [C], where this relation has been called the Morette–Van Hove relation, by analogy with a particular form taken by the determinant that these authors have discovered in the framework of Pauli’s semi-classical approximation to Feynman’s propagator (W. Pauli, *Pauli Lectures on Physics*, Vol. 6, Chap. 7, MIT Press (1972); Ph. Choquard and F. Steiner, *Helv. Phys. Acta*, 69:636–654 (1996)).

Review.

In the master thesis [A], we consider a mono-atomic conservative liquid occupying a domain in \mathbb{R}^1 and characterized by its mass density $\rho(x, t)$ and velocity potential $S(x, t)$, two canonically conjugated variables, and its Hamiltonian $H(S, \rho) = \frac{1}{2} \int dx \rho(x) S_x^2(x) + \frac{1}{2} \int dx dy \rho(x) \varphi(|x-y|) \rho(y)$, where $\varphi(|x|)$ is the interaction potential for pairs of atoms and per square of the mass unit (Ph. Choquard, *Physica A*, 279:45–59 (2000)). The dynamics are governed by Hamilton’s equations, being here the continuity equation $\rho_t + (\rho S_x)_x = 0$ and the non-local Bernoulli equation $S_t + \frac{1}{2} S_x^2 + \phi(x) = 0$, where $\phi(x) = \int dy \varphi(|x-y|) \rho(y)$, isomorphic to a non-local Hamilton–Jacobi equation, the non-locality resulting from the presence of pair interactions between the particles of the liquid. Writing the latter as a function of the velocity field $u(x, t) = S_x(x, t)$ yields the corresponding non-local Euler equation. In this work we make an inventory of, simulate numerically and analyze the effect on the shocks, if present, of well-defined repulsive or attractive, of short or long range potentials $\varphi(|x|)$ on the behavior of the mass density $\rho(x, t)$ and the velocity potential $S(x, t)$ around the origin of the x -axis, in particular, and as function of the time, for initial conditions of the density $\rho(x, 0)$ and potential $S(x, 0)$ even in x and respectively concave and convex. We consider a local δ -potential, a long range potential of type $|x|$, combinations of both locally repulsive and long range attractive potentials, and the Kac and Morse potentials. We simulate the evolution of ρ and u for positive times in regions without shocks.

In paper [B], it is shown that a class of admissible solutions of the continuity and Bernoulli or Burgers’ equations of a perfect one-dimensional liquid is given by the Morette–Van Hove relation which stipulates that the mass density is proportional to the second derivative of the velocity potential as shown in the preamble of this appendix. Positivity of the density implies convexity of the potential, i.e. smooth solutions, no shock for positive times. Non-elementary and symmetric solutions of the above equations are given in analytical and numerical form. Analytically, these solutions are derived from the original Ansatz proposed by Choquard (*Foundations of Physics*, 31:623–640 (2001)) and from the ensuing operations which show that they represent a particular case of the general implicit solutions of Burgers’ equation. Numerically and with the help of an ad

hoc computer program, these solutions are simulated for a variety of initial conditions called “compatible” if they satisfy the Morette–Van Hove formula and “anti-compatible” if the sign of the initial velocity field is reversed. In the latter case, singular behavior is observed. Part of the theoretical development presented here is rephrased in the context of the Hopf–Lax formula (i.e. positive times) whose domain of applicability for the solution of the Cauchy problem of the homogeneous Hamilton–Jacobi equation has recently been enlarged.

In paper [C] we develop fruitful analogies for one-dimensional systems, partially first established by C.M. Newman, between the variables, functions and equations which describe the equilibrium properties of classical ferro- and antiferromagnets in the Mean Field Approximation (MFA) and those which describe the space-time evolution of compressible Burgers’ liquids. It is shown that the natural analogies are: magnetic field and position coordinate; ferro-/antiferromagnetic coupling constants and negative/positive times; free energy per spin and velocity potential; magnetization and velocity field; magnetic susceptibility and mass density. An unexpected consequence of these analogies is a derivation of the Morette–Van Hove relation. Another novelty is that they necessitate the investigation of weak solutions of Burgers’ equation for negative times, corresponding to the Curie–Weiss transition in ferromagnets. This is achieved by solving the “final-value” problem of the homogeneous Hamilton–Jacobi equation. Unification of the final- and initial-value problems results in an extended Hopf–Lax variational principle. It is shown that its applicability implies that the velocity potentials at time zero be Lipschitz continuous, a rather mild condition for the class of physically interesting and functionally compatible velocity potentials, compatible in the sense of satisfying the Morette–Van Hove relation.

Finally, in paper [D], we report results of computer simulations and of theoretical analysis done to investigate and interpret the space-time evolution of the mass density and the velocity field of the inviscid self-gravitating (attractive) and (repulsive) Coulomb liquids in 1D with correlated initial conditions, namely proportionality between the mass density and the divergence of the velocity field. Numerical data gathered for both models in a collisionless regime reveal an evolution with a time-dependent proportionality factor. Feeding this result in the continuity and div-Euler equations leads to the introduction of another field which is shown to satisfy a Burgers type of implicit equation. A thorough description of regular implosion followed by singular collapses in the attractive case, and of regular explosion in the repulsive case is obtained. Time-inversion symmetry is investigated, energy conservation and stability properties are shown to apply in the regular regions of smooth solutions. The velocity potential satisfies a new local and inhomogeneous PDE.

Bibliographical references.

- [A] J. Wagner. Etude par analyse numérique des équations de continuité et d'Euler associées à différents modèles uni-dimensionnels de liquides conservatifs, *Master thesis*, Swiss Federal Institute of Technology, Lausanne, 2002.
- [B] Ph. Choquard and J. Wagner. The homogeneous Hamilton–Jacobi and Bernoulli equations revisited, II. *Foundations of Physics*, 32(8):1225–1249, 2002.
- [C] Ph. Choquard and J. Wagner. On the “Mean-Field” Interpretation of Burgers’ equation, *J. Stat. Phys.*, 116(1–4):843–853, 2004.
- [D] Ph. Choquard and J. Wagner. On a class of implicit solutions of the continuity and Euler’s equations for 1D systems with long range interactions. *Physica D*, 201:230–248, 2005.

Curriculum Vitæ

I was born on October 31st, 1978 in Luxembourg. I have done my secondary education in Luxembourg and obtained my diploma (option Mathematics-Physics-Latin) with distinction in 1997 in the Athénée de Luxembourg, Luxembourg.

I was then admitted at the Swiss Federal Institute of Technology (EPFL) in Lausanne in 1997. In 2000, I was awarded the *Cousin Prize*, and in 2001, I received the Entrepreneurship certificate from the Chair of Entrepreneurship and Innovation (Create) directed by Prof. Jane Royston. I got the degree of EPF Physics Engineer in 2002, after having done my master thesis in the Chair of Numerical Analysis and Simulation under the supervision of Prof. Jacques Rappaz and Hon. Prof. Philippe Choquard. I was awarded the *Dommer Prize* and my diploma work was honored by the *Landry Prize*.

Since 2002, I have been working as a research assistant in the Chair of Numerical Analysis and Simulation of Prof. Jacques Rappaz. My research theme is finite element methods with patches. Upon the invitation of Prof. Jiwen He, I have been exchange visitor at the University of Houston (USA) for a couple of months in 2002 and 2003. All along my thesis work I have collaborated with Hon. Prof. Philippe Choquard on fluid dynamics with long range self-interactions.

I am coauthor of textbooks (*A comme Algèbre* with Prof. G. Ternes, 1996 ; *Dislocations et plasticité des cristaux* with Prof. J.-L. Martin, 1999) and numerous articles in Mathematics and Physics. Since 1997, I have been working as scientific and technical adviser for several companies. As a freelance consultant, I have accomplished major projects in the domain of Information and Communication Technologies.