Iterative Substructuring Algorithms for the p-version Finite Element Method for Elliptic Problems

Ion Bică
Courant Institute of Mathematical Sciences
New York University

September 1997

A dissertation in the Department of Mathematics Submitted to the Faculty of the Graduate School of Arts and Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy at New York University.

Approved: ______________________________
Olof B. Widlund, Advisor

ACKNOWLEDGEMENTS

I am deeply grateful to my advisor, Olof Widlund for his guidance, support, and friendship throughout my years of graduate studies at Courant. In particular, I would like to thank him for suggesting me this problem, for his help and directions.

I also wish to thank Professors Charles Peskin and Peter Lax for their friendly advice and support and Luca Pavarino for the helpful discussions about his and my work.

I also thank like to thank all the faculty, staff and students at Courant for creating a warm atmosphere, in particular to Tamar Arnon who has been very supportive all these years.

ABSTRACT

In this thesis, we study iterative substructuring methods for linear elliptic problems approximated by the p-version finite element method. They form a class of nonoverlapping domain decomposition methods, for which the information exchange between neighboring subdomains is limited to the variables directly associated with the interface, i.e. those common to more than one subregion. Our objective is to design algorithms in $3D$ for which we can find an upper bound for the *condition number* κ of the preconditioned linear system, which is independent of the number of subdomains and grows slowly with p.

Iterative substructuring methods for the $h-$version finite element, and spectral elements have previously been developed and analysed by several authors. However, some very real difficulties remained when the extension of these methods and their analysis to the $p-$version finite element method were attempted, such as a lack extension theorems for polynomials. The corresponding results are well known for Sobolev spaces, but their extension to finite element spaces is quite intricate. In our technical work, we use and further develop extension theorems for polynomials in order to prove bounds on the condition numbers of several algorithms.

We have also made many numerical tests. We can use our programs for several purposes. Not only can we compute the condition numbers and study the rate of convergence for a variety of the algorithms that we have developed, but we can also compute the bounds on these condition numbers, as given by the theory. This is useful because the theory predicts the order of magnitude and the asymptotic growth of these bounds, not of the actual condition numbers.

Contents

List of Figures

Chapter 1

Introduction

1.1 Overview

Domain decomposition methods are iterative methods for solving the linear systems of equations that arise in the discretization of elliptic partial differential equations by, e.g., finite elements. There are almost yearly conferences and recently a first monograph by Smith, Bjorstad, and Gropp [62] has appeared. There is a general abstract theory, known as Schwarz theory; see, e.g., [25, 62].

The initial problem is divided into smaller subproblems that correspond to the decomposition of the initial domain into overlapping or nonoverlapping domains or, more generally, to the decomposition of the finite element space into a sum of subspaces. The subproblems are solved directly or iteratively, and the exchange of information between subspaces is handled by an iterative method. These subproblems can be viewed as building blocks of a preconditioner. In addition, a coarse space that handles the global exchange of information is needed. The choice of this subspace is a nontrivial issue.

The iterative substructuring methods form a class of nonoverlapping domain decomposition methods for which the information exchange between neighboring subdomains is limited to the variables directly associated with the interface, i.e. those common to more than one subregion.

An analysis of the iterative substructuring methods for the $h-$version in three dimensions can be found in Dryja, Smith, and Widlund [25]; see also Bramble, Pasciak, and Schatz [13, 14] for important earlier work.

Iterative substructuring methods for the $2D$ $p-$version are analyzed in Babuška, Craig, Mandel, and Pitkäranta [9]. For $3D$ spectral elements, see Pavarino and Widlund [55, 56], Mandel [42], Casarin [18, 19], and Korneev and Jensen [35]. The $hp-$version is analyzed in Oden, Patra, and Feng [50] and Guo and Cao [31, 33, 32]. For a different approach to the $3D$ $p-$version, see Sherwin and Karniadakis [59, 60, 61] and Pathria and Karniadakis [51].

In Chapter 1 of this thesis, we review some basic definitions and results about Sobolev spaces and the $p-$version finite element method.

In Chapter 2, we present the abstract Schwarz framework.

In Chapter 3, we introduce several algorithms, which are analogous to the algorithms defined by several authors for the h-version and spectral elements.

In Chapter 4, we prove two extension theorems for polynomials which we use to construct our *low energy* basis functions. Using these, we find that the condition number of the resulting algorithms grows only polylogarithmically in p. We also collect a series of technical results about polynomial subspaces of Sobolev spaces, which have been proven by several authors. Finally, we obtain our asymptotic bounds on the condition numbers.

In Chapter 5, we start by numerically computing the local condition numbers of our algorithms. It is important to note that for one of the algorithms, the *wire basket based algorithm* the global condition number is the largest of these local condition numbers. We then compute the bounds on these condition numbers, as given by the theory. This is useful because the theory predicts the order of magnitude and the asymptotic growth of these bounds, not of their actual values. To this end, we compute all the factors that give the upper bounds on the condition numbers, such as the norm of the low energy extension and a best discrete Sobolev constant. We conclude that we have made a good choice since the norm of the extension, which depends on the low energy extensions is much smaller than the best discrete Sobolev constant, which depends solely on the geometry. Finally, we compute the global condition numbers and study the rate of convergence of the additive, hybrid, and multiplicative variants of our algorithms. We also compare the condition numbers resulting from the use of the low energy and standard basis functions and for algorithms with coarse spaces that contain and do not contain the constants. We report on experiments for several choices of the vertex, edge, and coarse spaces, to be able to make specific recommendations on the best choice of preconditioners.

We have computed the local stiffness matrices symbolically, which is a relatively major task. We have used Maple V to do so. Most of the other programs have been written in Matlab.

1.2 Sobolev spaces

1.2.1 Definitions

We assume that Ω is a bounded Lipschitz domain in R^n.

We denote by $H^m(\Omega)$ the subspace of functions $u \in L^2(\Omega)$ such that $D^\alpha u \in L^2(\Omega)$ for $|\alpha| \le m$. Here, m is a nonnegative integer, $\alpha = (\alpha_1, \dots, \alpha_n)$ a multiindex, where the α_i are nonnegative integers, $|\alpha| = \alpha_1 + \cdots + \alpha_n$, and $D^\alpha = \partial^\alpha / \partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}$. $H^m(\Omega)$ is the Hilbert space with the inner product

$$(u, v)_{H^m(\Omega)} = \sum_{|\alpha| \le m} (D^\alpha u, D^\alpha v)_{L^2(\Omega)}. \tag{1.1}$$

We denote by $|\cdot|_{H^m(\Omega)}$ the seminorm

$$|u|^2_{H^m(\Omega)} = \sum_{|\alpha| = m} ||D^\alpha u||^2_{L^2(\Omega)}. \tag{1.2}$$

It is known, see [1], that $H^m(\Omega)$ can be defined equivalently as the completion of the space of infinitely differentiable functions, under the inner product (1.1). In our work, we consider only the case $m = 1$.

Some Sobolev spaces of fractional order are directly related to the problem of traces. If $0 < s < 1$, we denote by $H^s(\Omega)$ the space of functions $u \in L^2(\Omega)$ such that

$$\frac{|u(x) - u(y)|}{|x - y|^{s+n/2}} \in L^2(\Omega) \times L^2(\Omega).$$

It is a Hilbert space with the norm

$$||u||^2_{H^s(\Omega)} = ||u||^2_{L^2(\Omega)} + \int_\Omega \int_\Omega \frac{|u(x) - u(y)|^2}{|x - y|^{2s+n}} dx dy. \tag{1.3}$$

$H^s(\Omega)$ can be also defined by interpolation between H^1 and L^2; see [39]. Only the case of C^∞ boundary is treated there, but many results carry over to the case of polyhedral domains.

In all of R^n, the Sobolev spaces, of integer or fractional order, can also be defined via the Fourier transform; see [47, 39]. Extension theorems for Sobolev spaces are used to extend this equivalent definition to bounded regions; see [47, 39].

Any function that belongs to a Sobolev space $H^m(\Omega)$, with m large enough, is continuous; the higher the dimension n of R^n, the larger m is required. This well-known result follows from Sobolev's lemma.

Theorem 1.1 *If $2m > n$, then, for any $u \in H^m(\Omega)$*

$$\max_{x\in\Omega} |u(x)| \leq C||u||_{H^m(\Omega)}.$$

Consequently, u is continuous on $\bar{\Omega}$.

If Ω is a bounded interval in R^1, all functions in $H^1(\Omega)$ are continuous. This can be obtained immediately, without Sobolev's lemma, by elementary calculations.

1.2.2 Traces and extensions

Of particular importance to boundary value problems is whether functions in Sobolev spaces have traces on submanifolds of their domain of definition or not. We also need to characterize the trace classes.

Theorem 1.2 *Let $u \in H^s(\Omega)$, $1/2 < s < 3/2$. Then its trace γu on $\partial\Omega$ belongs to $H^{s-1/2}(\partial\Omega)$ and*

$$||\gamma u||_{H^{s-1/2}(\partial\Omega)} \leq C||u||_{H^s(\Omega)}.$$

Directly related to the trace theorem above is the subspace $H_0^s(\Omega)$ of $H^s(\Omega)$, defined as the closure of C_0^∞ in the H^s- norm. If $0 \leq s < 1/2$, $H_0^s(\Omega) = H^s(\Omega)$, and if $1/2 < s \leq 1$, $H_0^s(\Omega)$ is a subspace of $H^s(\Omega)$, which can also be defined as

$$H_0^s(\Omega) = \{u \in H^s(\Omega);\ \ \gamma u = 0 \quad \text{on} \quad \partial\Omega\}.$$

If $u \in H^s(\Omega),\ 0 \leq s < 1/2$, then the function $\tilde{u}$, defined as u in Ω and as zero outside Ω, belongs to $H^s(R^n)$. If $u \in H_0^s(\Omega),\ 1/2 < s \leq 1$, then $\tilde{u}$ also belongs to $H^s(R^n)$.

The case $s = 1/2$ deserves special attention. The closure of $C_0^\infty(\Omega)$ in $H^{1/2}(\Omega)$ is $H^{1/2}(\Omega)$, but not all the functions in $H^{1/2}(\Omega)$ belong to $H^{1/2}(R^n)$ when extended by zero. We denote by $H_{00}^{1/2}(\Omega)$ the maximal subspace of functions in $H^{1/2}(\Omega)$ that have this property. It can equivalently be defined as the closure of $C_0^\infty(\Omega)$ under the norm

$$||u||^2_{H_{00}^{1/2}(\Omega)} = |u|^2_{H^{1/2}(\Omega)} + \int_\Omega \frac{u(x)^2}{dist(x,\partial\Omega)}dx. \tag{1.4}$$

The inverse problem, of extending a function defined on $\partial\Omega$ to the whole domain Ω, has the following solution

Theorem 1.3 *There is a linear operator $E : H^{s-1/2} \to H^s(\Omega)$ such that, if $f \in H^{s-1/2}(\partial\Omega)$ and $s \geq 1/2$, then $Ef = f$ on $\partial\Omega$ and*

$$||Ef||_{H^s(\Omega)} \leq C||f||_{H^{s-1/2}(\Omega)}.$$

We remark that this theorem holds for $s = 1/2$, whereas Theorem 1.2 does not. For $s > 1/2$, the above theorem is a straightforward application of the open mapping theorem.

In our work, we are mainly concerned with trace and extension theorems for polynomial subspaces of Sobolev spaces. In Section 4.1, we will consider trace theorems that hold in the case $s = 1/2$ at a penalty that depends on the dimension of the subspace.

1.2.3 Equivalent norms

In the space $H_0^1(\Omega)$, the H^1-seminorm is a norm, equivalent to the H^1-norm. This fact follows immediately from Friedrichs' inequality; see [47].

Theorem 1.4 *Let Ω be a bounded domain. Then, for any $u \in H_0^1(\Omega)$,*

$$||u||_{L^2(\Omega)} \leq C|u|_{H^1(\Omega)}$$

We are also concerned with the spaces $H_0^1(\Omega) \subset H_\Gamma^1(\Omega) \subset H^1(\Omega)$, where Γ is a subset of $\partial\Omega$ of nonzero measure; see [47].

Theorem 1.5 *Let Ω be a bounded domain. Then, for any $u \in H^1(\Omega)$,*

$$||u||^2_{L^2(\Omega)} \leq C\left(|u|^2_{H^1(\Omega)} + (\int_\Gamma u dx)^2\right).$$

The H^1-seminorm and norm are also equivalent in the quotient space H^1/const. The result is Poincaré's inequality; see [47].

Theorem 1.6 *Let Ω be a bounded domain. Then, for any $u \in H^1(\Omega)$,*

$$||u||^2_{L^2(\Omega)} \leq C\left(|u|^2_{H^1(\Omega)} + (\int_\Omega u dx)^2\right).$$

We note that the constants in Theorems 1.4, 1.5, and 1.6 depend on the diameter of Ω.

Figure 1.1: The reference tetrahedron Ω_{ref}

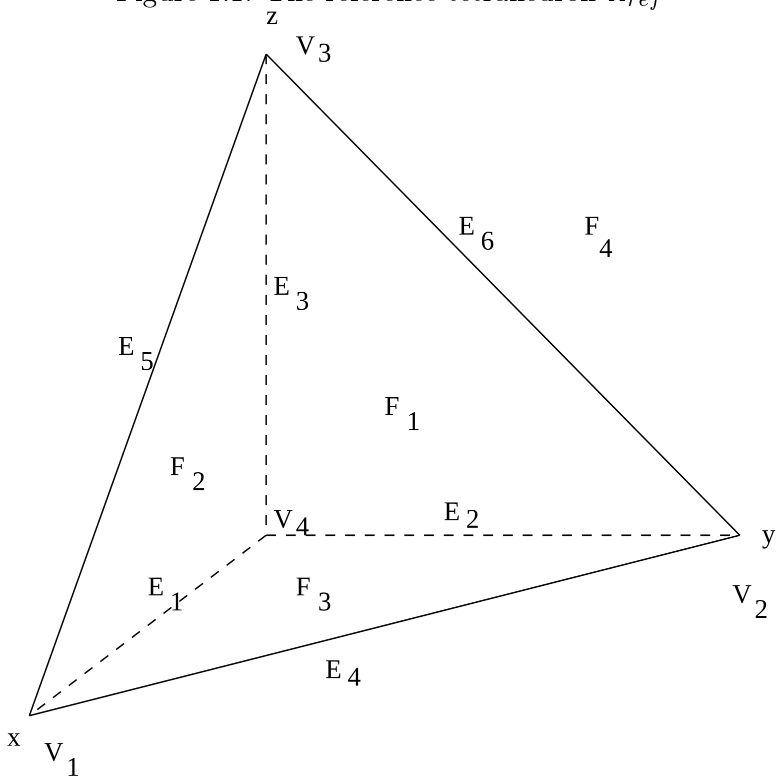
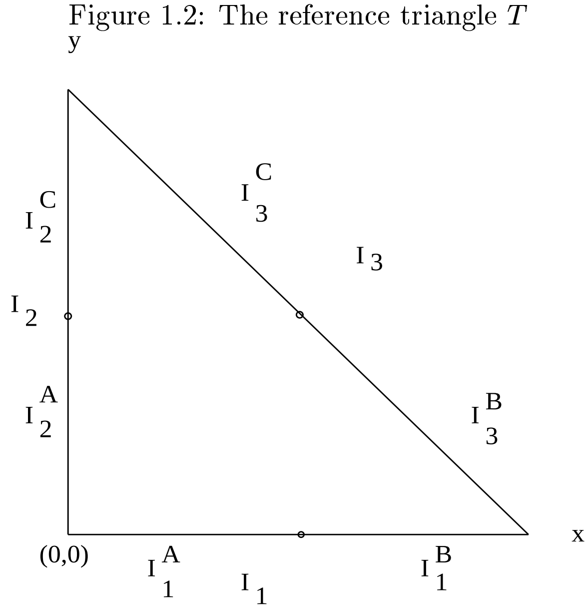


Figure 1.2: The reference triangle T



1.2.4 Explicit norms on a tetrahedral region

Notations. The main part of our analysis will be carried out on a tetrahedron. Let us consider the *reference tetrahedron*, shown in Fig. 1.1, defined by

$$\Omega_{ref} = \{(x, y, z); \quad x \geq 0,\ y \geq 0,\ z \geq 0,\ x + y + z \leq 1\}. \tag{1.5}$$

We use the following notations for its faces, edges, and vertices:

- F_i, $i = 1, \ldots, 4$, are the faces contained in the planes $\{x = 0\}$, $\{y = 0\}$, $\{z = 0\}$, and $\{x + y + z = 1\}$, respectively;

- E_i, $i = 1, \ldots, 6$, are the edges contained in the lines $\{y = z = 0\}$, $\{x = z = 0\}$, $\{x = y = 0\}$, $\{z = 0,\ x + y = 1\}$, $\{y = 0,\ x + z = 1\}$, and $\{x = 0,\ y + z = 1\}$, respectively;

- V_i, $i = 1, \ldots, 4$, are the vertices $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$, and $(0, 0, 0)$, respectively.

We define the barycentric functions on Ω_{ref} by $\mu_1 = x$, $\mu_2 = y$, $\mu_3 = z$, and $\mu_4 = 1 - x - y - z$. We denote by W the union of the closed edges of Ω_{ref} and call it the *wire basket* of the tetrahedron Ω_{ref}. We denote by $P^p(\Omega_{ref})$ the space of total degree p polynomials on Ω_{ref} and by $P_0^p(\Omega_{ref})$ the subspace of polynomials that vanish on the boundary $\partial\Omega_{ref}$ of the reference tetrahedron. If f is a function defined on $\partial\Omega_{ref}$, we denote its restriction to the face F_i by f_i. We define $P^p(\partial\Omega_{ref})$ as the space of continuous functions on $\partial\Omega_{ref}$, such that the f_i are polynomials of degree at most p, in the two variables corresponding to F_i.

Let T be the *reference triangle*, shown in Fig. 1.2, defined by

$$T = \{(x, y); \quad x \geq 0,\ y \geq 0,\ x + y \leq 1\}. \tag{1.6}$$

We use the following notations

- A, B, and C are the vertices $(0, 0)$, $(1, 0)$, and $(0, 1)$, respectively;

- I_i, $i = 1, \ldots, 3$, are the edges contained in the lines $\{y = 0\}$, $\{x = 0\}$, and $\{x + y = 1\}$;

- I_1^A, I_1^B, I_2^A, I_2^C, I_3^B, I_3^C, are half-edges, shown in the figure.

We define the barycentric functions on T by $\lambda_1 = x$, $\lambda_2 = y$, and $\lambda_3 = 1 - x - y$. We define $P^p(T)$, $P_0^p(T)$, and $P^p(\partial T)$ in the same way as $P^p(\Omega_{ref})$, $P_0^p(\Omega_{ref})$, and $P^p(\partial\Omega_{ref})$, respectively.

We use the notation Q^p for the spaces of polynomials of degree at most p in each variable.

Special norms. The principal goal of this subsection is to introduce explicit expressions for the $H^{1/2}$-norm (1.3) on the boundary $\partial\Omega_{ref}$ of the reference tetrahedron. We will use the following equivalent norm; see [30, 48]:

$$||u||^2_{H^{1/2}(\partial\Omega_{ref})} = \sum_{k=1}^{4} ||u_k||^2_{H^{1/2}(F_k)} + \sum_{i<j} C_{i,j}. \tag{1.7}$$

Here, the $C_{i,j}$ correspond to the cross terms $\int_{F_i}\int_{F_j}$ in the double integral in the definition of the $H^{1/2}$-norm (1.3). Thus,

$$C_{1,2} = C(u_1, u_2) = \int_0^1 \int_0^{1-z} \frac{|u_1(0,y,z) - u_2(y,0,z)|^2}{y} dydz. \tag{1.8}$$

The other terms are defined similarly.

We will also use the following subspaces of $H^{1/2}(T)$ related to one or two of its edges: $H_{00}^{1/2}(T, I_i)$ is the space of functions $u \in H^{1/2}(T)$, that vanish on the edge I_i and satisfy $\lambda_i^{-1/2} u \in L^2(T)$, with the norm

$$||u||^2_{H_{00}^{1/2}(T,\mathcal{I}_i)} = |u|^2_{H^{1/2}(T)} + ||\lambda_i^{-1/2} u||^2_{L^2(T)}, \tag{1.9}$$

and $H_{00}^{1/2}(T, I_i,\ I_j) = H_{00}^{1/2}(T, I_i) \cap H_{00}^{1/2}(T, I_j)$ for $i \neq j$, with the norm

$$||u||^2_{H_{00}^{1/2}(T,I_i,\ I_j)} = |u|^2_{H^{1/2}(T)} + ||\lambda_i^{-1/2} u||^2_{L^2(T)} + ||\lambda_j^{-1/2} u||^2_{L^2(T)}. \tag{1.10}$$

Similarly, $H_{00}^{1/2}(T, I_1,\ I_2,\ I_3) = H_{00}^{1/2}(T)$, and its norm is equivalent to the one given by (1.4).

We remark that, if $u \in H^{1/2}(\partial\Omega_{ref})$ is nonzero on the face F_k and equal to zero on the other three faces, then

$$||u||^2_{H^{1/2}(\partial\Omega_{ref})} = ||u||^2_{H_{00}^{1/2}(T)},$$

where we identify the face F_k of Ω_{ref} with the reference triangle T. If $u \in H^{1/2}(\partial\Omega_{ref})$ vanishes on F_4 and $u(x, y, 0) = u(x, 0, y) = u(0, x, y)$, then

$$||u||^2_{H^{1/2}(\partial\Omega_{ref})} = 3||u||^2_{H_{00}^{1/2}(T,I_3)}.$$

Finally, if $u \in H^{1/2}(\partial\Omega_{ref})$ vanishes on F_1 and F_4, and $u(x,y,0) = u(x,0,y)$, then

$$||u||^2_{H^{1/2}(\partial\Omega_{ref})} = 2||u||^2_{H^{1/2}_{00}(T,I_2,I_3)}.$$

1.3 Continuous and discrete problems

1.3.1 The model problem

We consider the following problem formulated variationally: Find $u \in V$ such that

$$a(u,v) = \int_{\Omega} \rho(x) \nabla u \nabla v \, dx = f(v) \quad \forall v \in V. \tag{1.11}$$

Here, V is a subspace of $H^1(\Omega)$, determined by boundary conditions, Ω is a polyhedral region triangulated with tetrahedra $\bar{\Omega}_i$, $\bar{\Omega} = \cup \bar{\Omega}_i$. We assume that the boundary conditions are of the same type within each face of any tetrahedra that is part of the boundary. The coefficient $\rho(x) > 0$ can be discontinuous across the interface between the subdomains, but varies only moderately within each Ω_i. Without further decreasing the generality, we assume $\rho(x) = \rho_i$ on Ω_i. We note that some of the bounds in Chapter 4 hold for arbitrary jumps in ρ_i.

1.3.2 The finite element method

We discretize the problem by applying the p-version finite element method: Find $u \in V^{FEM}$ such that

$$a(u,v) = f(v) \quad \forall v \in V^{FEM}, \tag{1.12}$$

where V^{FEM} is a finite dimensional subspace of V.

There are three main versions of the finite element method

- $h-$version. This is the standard version. The finite element space is the space of continuous, piecewise low order polynomials. The mesh is refined in order to increase the accuracy.

- $p-$version. The finite element space is the space of continuous piecewise higher order polynomials. In its pure form, the accuracy is increased by increasing the degree of the polynomials, while the mesh is fixed.

- $hp-$version. This is a combination of the two previous methods. Different degree polynomials are often used in different elements and the mesh is often far from uniform.

The $h-$version has been known and used extensively since the fifties. The development of the p and $hp-$ version has been stimulated by problems in structural mechanics and mechanics of solids. The first experimental and convergence results appeared in the late seventies; see [64], [63]. The systematic analysis of convergence has been worked out in the eighties; see [6], [4], [22, 23], [2], [5], [65], [49].

Error bounds for the finite element are based on approximation results and the structure and regularity of the solution of the elliptic problem (1.11). The following is Lemma 11 in Muñoz-Sola [48]. It is a $3D$ variant of the basic approximation results in two dimensions, proved in Babuška, Szabó, and Katz [6] and in Babuška and Suri [4].

Lemma 1.1 *Let $u \in H^k(\Omega_{ref})$. Then there exist a sequence $z_p \in P^p(\Omega_{ref})$, $p = 0, 1, 2, \ldots$, such that*

$$\begin{aligned}
||u - z_p||_{H^q(\Omega_{ref})} &\leq Cp^{-(k-q)}||u||_{H^k(\Omega_{ref})} \quad for \quad 0 \leq q \leq k; \\
||u - z_p||_{L^2(F_i)} &\leq Cp^{-(k-\frac{1}{2})}||u||_{H^k(\Omega_{ref})} \quad for \quad k > \frac{1}{2}, \quad i = 1, \ldots, 4; \\
||u - z_p||_{H^1(F_i)} &\leq Cp^{-(k-\frac{3}{2})}||u||_{H^k(\Omega_{ref})} \quad for \quad k > \frac{3}{2}, \quad i = 1, \ldots, 4; \\
||u - z_p||_{L^2(E_i)} &\leq Cp^{-(k-1)}||u||_{H^k(\Omega_{ref})} \quad for \quad k > 1, \quad i = 1, \ldots, 6; \\
||u - z_p||_{H^1(E_i)} &\leq Cp^{-(k-2)}||u||_{H^k(\Omega_{ref})} \quad for \quad k > 2, \quad i = 1, \ldots, 6; \\
||u - z_p||_{L^\infty(\Omega_{ref})} &\leq Cp^{-(k-\frac{3}{2})}||u||_{H^k(\Omega_{ref})} \quad for \quad k > \frac{3}{2}.
\end{aligned}$$

The next step is to obtain piecewise polynomial approximations which are continuous across the boundaries of the elements without degrading the order of approximation. The technical tools used for this are extension theorems for polynomials; see [4], [9] for $2D$ global approximation results and [48] for $3D$. We will use these theorems for two other purposes, to define equivalent norms on the finite element space, see Section 1.3.4, and to construct special basis functions for the finite element space, see Chapter 4.

A main result concerning the rate of convergence for the $hp-$version in $3D$ is given in Muñoz-Sola [48]. To simplify the statement of the theorem, we assume homogenous boundary conditions.

Theorem 1.7 *Let u be the solution to the model problem (1.11), and let u_{hp} be the solution of the problem (1.12), discretized by using the $hp-$version finite element method.*

We assume that $u \in H^k(\Omega),\ k > 2$. *Then,*

$$||u - u_{hp}||_{H^1(\Omega)} \leq C\ h^{\min(p+1,k)-1}\ p^{1-k}\ ||u||_{H^k(\Omega)}.$$

The author remarks that the above result is only a step towards the complete analysis of the $hp-$version in $3D$, since the regularity hypothesis $u \in H^k(\Omega),\ k > 2$ does not always hold if Ω is a polyhedron. Further analysis would involve the decomposition of the solution into singular and regular parts.

A complete $2D$ analysis has been done in Babuška and Suri in [4, 3]. Let Ω be a polygonal domain, with vertices $A_1, A_2, \ldots$, and set Dirichlet boundary conditions on a part of $\Gamma_1 \subset \partial\Omega$ that is union of some of the segments $A_i A_{i+1}$. The piecewise polynomial boundary values can change discontinuously at the vertices. The solution u of the model problem can then be decomposed as

$$u = u_1 + u_2 + \sum_i u_{3,i},$$

where

$$u_1 \in H^{k_1},\ k_1 > 1, \quad u_1 = 0 \text{ on } \Gamma_1;$$
$$u_2 \in H^{k_2},\ k_2 > \frac{3}{2};$$

and

$$u_{3,i} = r_i^{\alpha_i} |\log r_i|^{\gamma_i} \phi_i(\theta_i) \chi_i(r_i),$$

where $r_i,\ \alpha_i$ are polar coordinates with respect to an origin located at the vertex A_i, $\alpha_i > 0,\ \gamma_i \geq 0$ are integers, ϕ_i is an analytic function and χ_i is smooth and zero away from A_i. The following theorem, taken from [4], provides optimal bounds on the error in $2D$

Theorem 1.8 *Let* u *be the solution of (1.11), and let* u_{hp} *be the solution to the problem (1.12), discretized by using the* $hp-$*version finite element method. If* $u \in H^k(\Omega),\ k > \frac{3}{2}$, *then,*

$$||u - u_{hp}||_{H^1(\Omega)} \leq C\ h^{\min(p+1,k)-1}\ p^{1-k}\ ||u||_{H^k(\Omega)}.$$

If $u = u_1 + u_2,\ u_1 \in H^{k_1}\ k_1 > 1, \quad u_1 = 0$ *on* $\Gamma_1,\ k_1 \leq \frac{3}{2},\ u_2 \in H^{k_2}\ k_2 > \frac{3}{2}$. *Then,*

$$||u - u_{hp}||_{H^1(\Omega)} \leq C\ h^{\min(p+1,k)-1}\ p^{1-k}\ ||u||_{H^k(\Omega)},$$

where $k = \min(k_1, k_2)$*. Finally, if* $u = u_{3,i}$ *in the above decomposition, then,*

$$||u - u_{hp}||_{H^1(\Omega)} \leq C|\log p|^{\gamma} p^{-2\alpha}.$$

The best choice of the finite element version depends on the properties of the solution u of the model problem (1.11). According to Babuška and Szabó [65], the following are the best choices

1. If u is analytic in each finite element, including its boundary, the best method is the $p-$version since the rate of convergence is exponential.

2. If u is analytic in each finite element, including its boundary, with possible singularities at vertices, the best method is the $hp-$version.

3. If the mesh cannot be constructed such that the singularities are located at vertices, the best method is the $h-$version.

1.3.3 Basis functions for the $p-$version

Local basis functions. We will now define basis functions on the reference tetrahedron Ω_{ref}. There are many many ways to do so, and a proper choice is the key to obtaining an efficient iterative method. We present here only a general description. We distinguish between four types of basis functions, associated with the vertices, edges, faces, and interior.

1. A *vertex basis function* has value one at that vertex and vanishes on the face opposite to the vertex. There is only one vertex function per vertex.

2. An *edge basis function* vanishes on the two faces which do not share the edge. The traces of the edge functions associated with the same edge are chosen to be linearly independent on that edge. There are $p - 1$ such functions per edge.

3. A *face basis function* vanishes on the other three faces. The traces of the face functions associated with the same face, are chosen to be linearly independent on that face. There are $(p-1)(p-2)/2$ such functions per face.

4. An *interior basis function* vanishes on $\partial\Omega_{ref}$. They are linearly independent. There are $(p-1)(p-2)(p-3)/6$ interior functions.

The total number of vertex, edge, face, and interior functions is $(p+1)(p+2)(p+3)/6$. It is easy to see that they form a basis for $P^p(\Omega_{ref})$.

We now present the standard, hierarchic basis introduced in [65].

1. The vertex basis functions are the four barycentric functions

$$\Psi_{V_i} = \mu_i, \quad i = 1, \ldots 4. \tag{1.13}$$

2. The edge basis functions associated with edge $E_1, \ldots, E_6$ are

$$\Psi^i_{E_1} = \mu_1\mu_4\phi_i(\mu_1 - \mu_4), \;\; i = 1, \ldots p-1. \tag{1.14}$$
$$\Psi^i_{E_2} = \mu_2\mu_4\phi_i(\mu_2 - \mu_4), \;\; i = 1, \ldots p-1. \tag{1.15}$$
$$\Psi^i_{E_3} = \mu_3\mu_4\phi_i(\mu_3 - \mu_4), \;\; i = 1, \ldots p-1. \tag{1.16}$$
$$\Psi^i_{E_4} = \mu_1\mu_2\phi_i(\mu_2 - \mu_1), \;\; i = 1, \ldots p-1. \tag{1.17}$$
$$\Psi^i_{E_5} = \mu_1\mu_3\phi_i(\mu_1 - \mu_3), \;\; i = 1, \ldots p-1. \tag{1.18}$$
$$\Psi^i_{E_6} = \mu_2\mu_3\phi_i(\mu_2 - \mu_3), \;\; i = 1, \ldots p-1. \tag{1.19}$$

Here, $\phi_i(x) = L_i'(x)/x(1-x)$ and L_i is the i-th Legendre polynomial on $[0,1]$.

3. The face basis functions associated with the faces $F_1, \ldots, F_4$ are

$$\Psi^{i,j}_{F_1} = \mu_2\mu_3\mu_4 L_i(\mu_2 - \mu_3)L_j(2\mu_4 - 1), \;\; i+j = 0\ldots p-3, \tag{1.20}$$
$$\Psi^{i,j}_{F_2} = \mu_1\mu_3\mu_4 L_i(\mu_1 - \mu_3)L_j(2\mu_4 - 1), \;\; i+j = 0\ldots p-3, \tag{1.21}$$
$$\Psi^{i,j}_{F_3} = \mu_1\mu_2\mu_4 L_i(\mu_2 - \mu_1)L_j(2\mu_4 - 1), \;\; i+j = 0\ldots p-3, \tag{1.22}$$
$$\Psi^{i,j}_{F_4} = \mu_1\mu 2\mu_3 L_i(\mu_2 - \mu_1)L_j(2\mu_3 - 1), \;\; i+j = 0\ldots p-3. \tag{1.23}$$

4. The interior basis functions are

$$\Psi^{i,j,k}_I = \mu_1\mu_2\mu_3\mu_4 L_i(\mu_2 - \mu_1)L_j(2\mu_3 - 1)L_k(2\mu_4 - 1), \quad i+j+k = 0\ldots p-4. \tag{1.24}$$

It turns out that the standard vertex and edge functions (1.13 - 1.19) are not appropriate for our purpose because, as we will see in Chapter 5, they result in a slow rate of convergence of the iterative methods. In Chapter 4, we therefore construct *low energy* vertex and edge functions. At this time, we only remark that the low energy functions with highly oscillatory traces on the wire basket decay much more rapidly away from the edges than the standard ones that have the same trace. See Fig. 1.3 for a comparison of low energy and standard (high energy) basis function.

Global basis functions. Once we have defined the basis functions on a reference tetrahedron, we can define them on any other tetrahedron by using an affine transformation. The basis functions (1.14 - 1.23) are not symmetric, i.e. we obtain a different set of functions by a permutations of the variables. This will result in some problems in constructing the triangulation since the vertex, edge, and face functions must have consistent definitions on any pair of tetrahedra with a common boundary; this may cause conflicting definitions of global edge or face functions on a closed face shared by two tetrahedra. See Fig. 1.4 for a simple $2D$ example, where one cannot resolve this conflict. One way of avoiding this difficulty is to use a symmetric basis, i.e. a basis such that any permutation of the vertices maps a basis function into a basis function. An alternative would be to assign some orientation to each edge and face and recompute the stiffness matrix for each element. In the example in Fig. 1.4, we can choose the orientation of the edge DC given by Triangle 2. Now, Triangle 3 is a new element, which cannot be obtained from Triangle 1 or Triangle 2 by using an affine transformation.

We do not adopt the first approach because there is no hierarchic symmetric basis for $P^p(K)$; see Zumbusch [68]. The basis (1.13 - 1.24) is hierarchical, but not symmetric. We have found the second approach to be too expensive in our experimental work and have avoided the problem by considering particular triangulations. A sufficient condition to avoid conflicts is to make sure that each vertex has the same local number with respect to each tetrahedra that share it. See Fig. 1.5 for an example of compatible versus incompatible local numbering.

In our experiments, we will often consider regions Ω that are unions of cubes and divide each cube into 24 tetrahedra, see Fig. 1.6. We do have compatible local numbering of the 15 vertices in the cube, with respect to of each of the 24 tetrahedra. As a consequence, each edge or face have compatible local numberings. Furthermore, this is true for tetrahedra that belong to different cubes. We number the vertices of each of the 24 tetrahedra in the cube as in Table 1.1.

Figure 1.3: Basis functions on the face F_3 of the reference tetrahedron Ω_{ref}

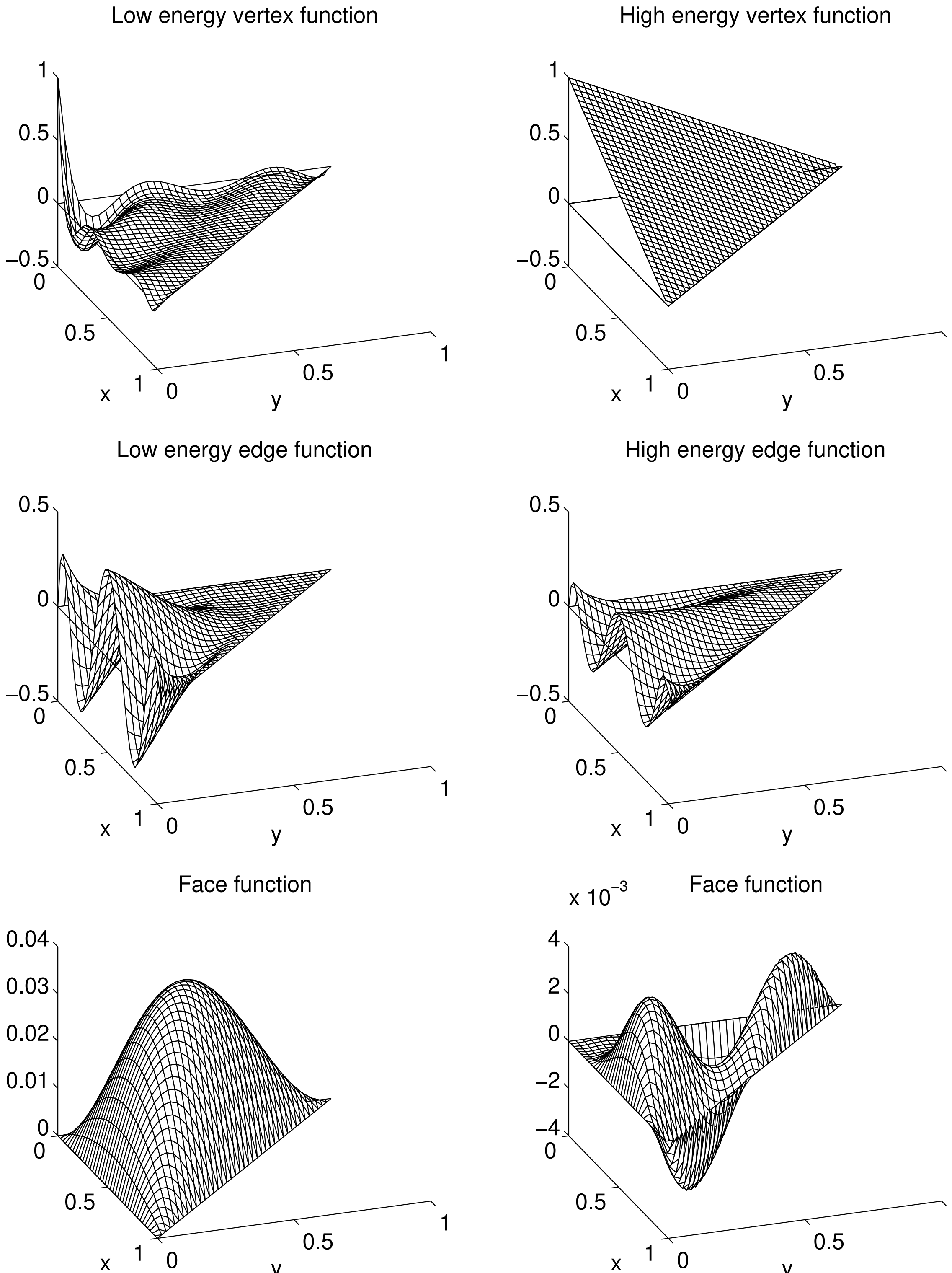


Figure 1.4: Conflicting orientations on the edge DC

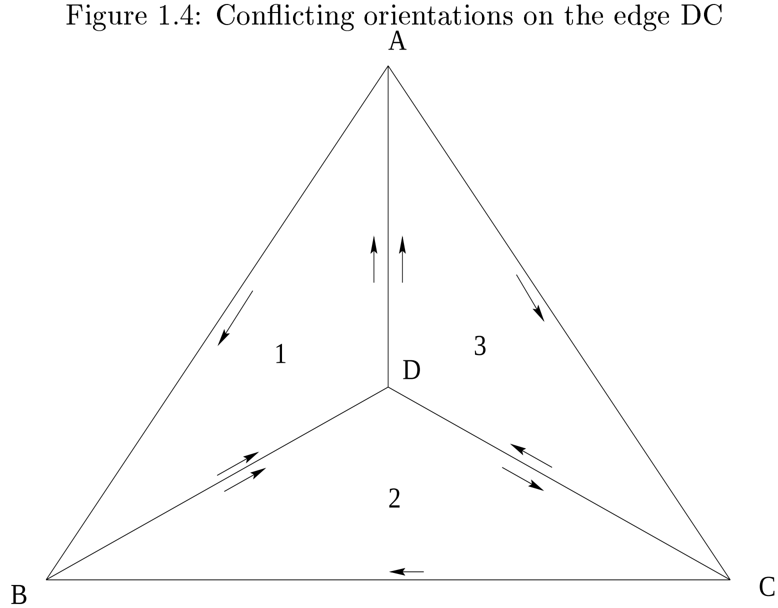


Figure 1.5: Compatible versus incompatible local numbering

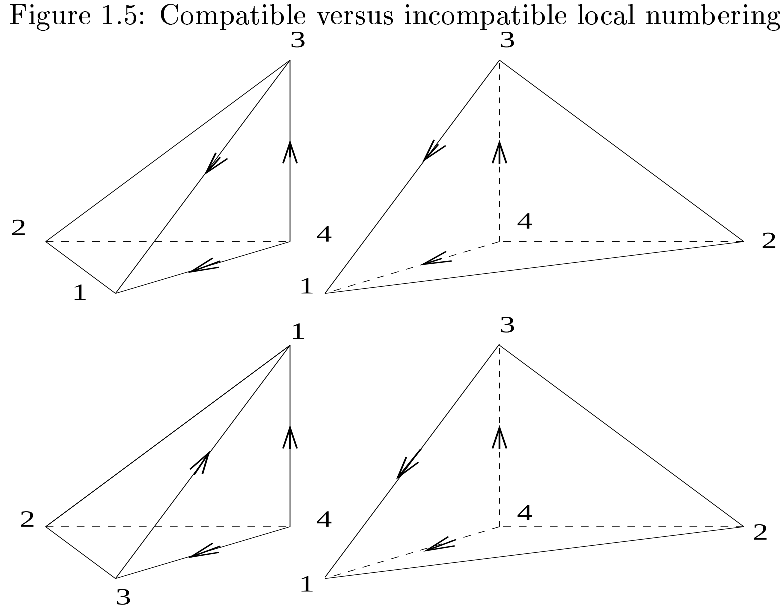
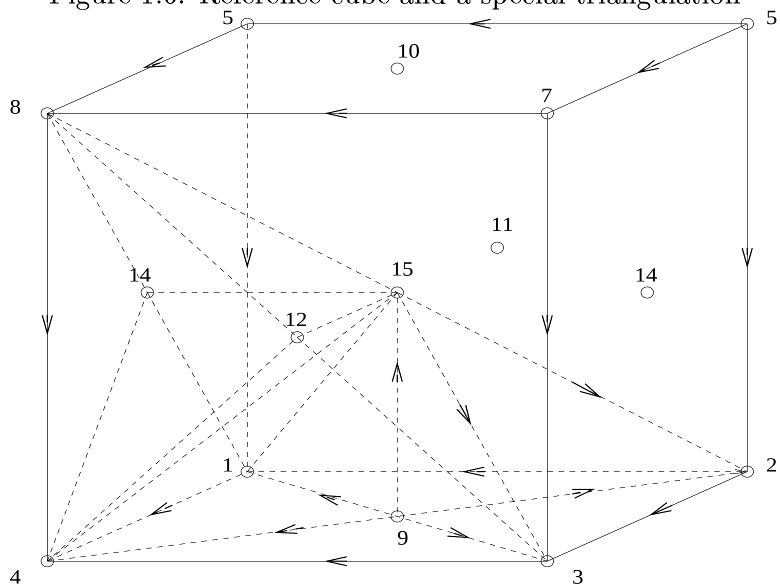


Table 1.1: Global numbering of the vertices of the cube with 24 subregions

tetrahedron	Global numbering of each vertex of the current tetrahedron			
	vertex 1	vertex 2	vertex 3	vertex 4
1	2	1	15	9
2	2	3	15	9
3	3	4	15	9
4	1	4	15	9
5	6	5	15	10
6	6	7	15	10
7	7	8	15	10
8	5	8	15	10
9	6	5	15	11
10	6	2	15	11
11	2	1	15	11
12	5	1	15	11
13	7	8	15	12
14	7	3	15	12
15	3	4	15	12
16	8	4	15	12
17	5	8	15	13
18	5	1	15	13
19	1	4	15	13
20	8	4	15	13
21	6	7	15	14
22	6	2	15	14
23	2	3	15	14
24	7	3	15	14

Figure 1.6: Reference cube and a special triangulation



1.3.4 Substructuring methods

Iterative substructuring methods are domain decomposition methods without overlap between subdomains. An overview of the direct and iterative substructuring methods for the $h-$version finite elements is given in Smith, Bjørstad, and Gropp [62]. A detailed analysis of many iterative substructuring methods for the $h-$version is given in Dryja, Smith, and Widlund [25]. For spectral elements, see Pavarino and Widlund [55, 56] and Mandel [42]. Algorithms for the $hp-$version in $2D$ and $3D$, hexahedral elements are analyzed in Guo and Cao [31, 33, 32].

The variational problem (1.12) can be rewritten as

$$K\underline{u} = \underline{f}.$$

Here, $K_{ij} = a(\phi_i, \phi_j)$, $f_i = (f, \phi_i)$, where ϕ_i are basis functions in V^p. The global system is built from local contributions, by subassembly,

$$K\underline{u} = \sum_i K^{(i)}\underline{u}^{(i)} = \sum_i \underline{f}^{(i)} = \underline{f}, \tag{1.25}$$

where the local stiffness matrices $K^{(i)}$ and the vectors $\underline{u}^{(i)}$ and $\underline{f}^{(i)}$ are expanded by zero vectors.

Table 1.2: Condition number of stiffness matrix and submatrices

p	W+F+I	W	F	I
4	7.8326e+04	783.3933	32.4031	1
5	1.0697e+06	803.2946	414.8193	35.3369
6	2.1524e+07	845.9709	3.4065e+03	1.0978e+03
7	5.3688e+08	867.4798	3.2074e+04	3.4047e+04
8	1.6179e+10	879.7329	3.5723e+05	1.2810e+06
9	5.2421e+11	894.8373	4.0552e+06	4.6596e+07
10	1.7524e+13	935.2027	4.5951e+07	1.6296e+09

p	W+F	W+I	F+I
4	1.6756e+04	2.4023e+04	405.5100
5	8.8998e+04	3.7748e+05	6.8090e+03
6	6.7713e+05	7.9859e+06	1.5728e+05
7	5.5259e+06	2.1086e+08	4.2174e+06
8	5.0418e+07	7.0602e+09	1.3336e+08
9	5.2306e+08	2.3492e+11	4.5341e+09
10	5.7463e+09	8.0466e+12	1.5231e+11

The global stiffness matrix K is very ill-conditioned. Following the ideas in [10], we present, in Table 1.2, the condition numbers of the local stiffness matrix as well as the submatrices associated to the wire basket (W), face (F), and the interior (I) degrees of freedom and of submatrices corresponding to several of these sets of degrees of freedom.

The stiffness matrix is computed in the basis (1.13 - 1.24), so the block that correspond to the vertex degrees of freedom, as well as any larger block containing it, is singular. In such cases, we define the condition number as $\kappa = \lambda_{\max}/\lambda_2$, where λ_2 is the smallest positive eigenvalue. It follows from the table that the main cause of the ill-conditioning is the coupling between the interior and boundary (wire basket and face) degrees of freedom.

This is the main motivation for eliminating the interior degrees of freedom by *static condensation*. We factor each local stiffness matrix $K^{(i)}$ as follows

$$K^{(i)} = \begin{pmatrix} I & 0 \\ {K_{IB}^{(i)}}^T {K_{II}^{(i)}}^{-1} & I \end{pmatrix} \begin{pmatrix} K_{II}^{(i)} & 0 \\ 0 & S^{(i)} \end{pmatrix} \begin{pmatrix} I & {K_{II}^{(i)}}^{-1} K_{IB}^{(i)} \\ 0 & I \end{pmatrix} \qquad (1.26)$$

where

$$S^{(i)} = K_{BB}^{(i)} - {K_{IB}^{(i)}}^T {K_{II}^{(i)}}^{-1} K_{IB}^{(i)},$$

is the *Schur complement* of $K^{(i)}$.

The system (1.25) reduces to

$$S\underline{u}_B = \sum_i S^{(i)} \underline{u}_B^{(i)} = \underline{f}_B - \sum_i K_{IB}^{(i)T} K_{II}^{(i)-1} \underline{f}_I^{(i)}, \tag{1.27}$$

where the matrices $S^{(i)}$ and vectors $\underline{u}_B^{(i)}$ and $\underline{f}_B^{(i)}$ are expanded by zeros.

Of particular importance to our work, and directly related to the Schur complement, is the space of *discrete harmonic functions* $\tilde{V}^p \subset V^p$, defined by

$$\tilde{V}^p = \{u \in V^p; \quad a(u,v) = 0 \quad \forall v \in V^p,\ v = 0 \text{ on } \Gamma\}.$$

Equivalently, $\tilde{V}^p$ can be defined as the functions with coefficients that satisfy

$$K_{II}\underline{u}_I + K_{IB}\underline{u}_B = 0.$$

The discrete harmonic functions are completely defined by their values on the interface Γ, and V^p can be decomposed into a direct sum of $a(\cdot,\cdot)-$orthogonal subspaces,

$$V^p = \tilde{V}^p + \sum_i P_0^p(\Omega_i).$$

It is easy to see that

$$\underline{u}_B^T S \underline{u}_B = a(u,u) = \min_{v=u \text{ on } \Gamma} a(v,v) = \min_{\underline{v}_B = \underline{u}_B} \underline{v}^T K \underline{v},$$

where $\underline{u}$ and $\underline{v}$ contain the coefficients of u and v, respectively.

The middle factor in (1.26) is the expression of the stiffness matrix in a new basis, with discrete harmonic vertex, edge, and face functions, defined by

$$\begin{aligned} \tilde{\Psi}^i &= \Psi_B^i + \sum_j \alpha_{ij} \Psi_I^j, \\ \tilde{\Psi}_I^j &= \Psi_I^j. \end{aligned}$$

Here, the Ψ_B^i are the original vertex, edge, and face basis functions, and the Ψ_I^j are the original interior functions. The α_{ij} are the entries in $-K_{II}^{-1}K_{IB}$. We remark that since the $\tilde{\Psi}_B^i$ are obtained from the Ψ_B^i by adding interior functions, their type (vertex, edge, or face) has not changed after the orthogonalization.

In Chapter 4, we will use the equivalent norm $\sum_i ||u||^2_{H^{1/2}(\Omega_i)}$ on the space $\tilde{V}^p$. In general, the proof of this equivalence is based on trace and extension theorems. As mentioned in Section 1.3.2, this is still another application of extension theorems for polynomials; cf. Section 1.3.2. We also use them in Chapter 4 in the study of our iterative methods to solve problem (1.12).

The goal of this work is to design iterative methods for the system (1.27). Such methods are known as iterative substructuring methods in the literature. The system is still very ill-conditioned, see the column W+F in Table 1.2, so an iterative method without preconditioning is potentially very expensive and is likely to be completely useless. We will use the preconditioned conjugate gradient method, presented in Section 1.4.

1.4 Iterative methods

The preconditioned conjugate gradient algorithm (PCG) is the standard choice for our problem.

Let A be a $N \times N$ symmetric, positive definite matrix, and consider the system

$$Ax = b.$$

Given a tolerance $\epsilon > 0$, the conjugate gradient algorithm (CG) is given by

Set $k = 0$; $x_0 = 0$; $r_0 = b$;
While $||r_k||_2 \geq \epsilon ||r_0||_2$
  $k = k + 1$
  if $k = 1$
    $p_1 = r_0$
  else
    $\beta_k = (r_{k-1}, r_{k-1})/(r_{k-2}, r_{k-2})$
    $p_k = r_{k-1} + \beta_k p_{k-1}$
  end
  $\alpha_k = (r_{k-1}, r_{k-1})/(P_k, Ap_k)$
  $x_k = x_{k-1} + \alpha_k p_k$
  $r_k = r_{k-1} - \alpha_k Ap_k$
end.

Here, $(x, y) = x^T Ay$. The conjugate gradient algorithm gives the exact solution after at most N steps. The reason is that the search diretions p_k are $S-$conjugate:

$$p_i^T Sp_j = 0, \;\; i \neq j.$$

It is well known that

$$||x - x_k||_S = 2\left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1}\right)^k ||x - x_0||_S,$$

where $\kappa(A)$ is the condition number of A, given by

$$\kappa(A) = \frac{\lambda_{max}(A)}{\lambda_{min}(A)}.$$

If $\kappa(A)$ is very large, we need to introduce a preconditioner B and solve the system

$$BAx = Bb.$$

The preconditioned conjugate gradient algorithm is obtained by applying CG to the transformed system

$$B^{1/2}AB^{1/2}\tilde{x} = B^{1/2}b,$$

where $\tilde{x} = B^{-1/2}x$. The explicit references to the matrix $B^{1/2}AB^{1/2}$ can be eliminated. The PCG algorithm becomes

Set $k = 0$; $x_0 = 0$; $r_0 = b$;
While $||r_k||_2 \geq \epsilon ||r_0||_2$
  Solve $z_k = Br_k$
  $k = k + 1$
  if $k = 1$
    $p_1 = z_0$
  else
    $\beta_k = (r_{k-1}, z_{k-1})/(r_{k-2}, z_{k-2})$
    $p_k = z_{k-1} + \beta_k p_{k-1}$
  end
  $\alpha_k = (r_{k-1}, r_{k-1})/(P_k, Ap_k)$
  $x_k = x_{k-1} + \alpha_k p_k$

$r_k = r_{k-1} - \alpha_k A p_k$

end.

The preconditioner B should be chosen such that

- $\kappa(B^{1/2}AB^{1/2})$ is much smaller than $\kappa(A)$,
- the solution of the system system $Bz_k = r_k$ is easy to compute.

Standard preconditioners are diagonal scaling and incomplete factorization. For domain decomposition methods, a standard choice is preconditioners based on the Schwarz framework.

Approximate values of the condition number of A can be obtained during the CG iteration by using a variant of an eigenvalue algorithm due to Lanczos [37]. Let R_k be the matrix with the normalized residual vectors $r_0/||r_0||_2, \ldots, r_k/||r_k||_2$ as columns. It is possible to prove that

$$R_k^T A R_k = \begin{pmatrix} 1/\alpha_1 & -\sqrt{\beta_2}/\alpha_1 & & \\ -\sqrt{\beta_2}/\alpha_1 & \beta_2/\alpha_1 + 1/\alpha_2 & -\sqrt{\beta_3}/\alpha_2 & \\ & -\sqrt{\beta_3}/\alpha_2 & \ddots & \ddots \\ & & \ddots & \ddots \end{pmatrix}.$$

We have obtained our experimental values of the condition number of our different iterative methods by computing the extreme eigenvalues of these tridiagonal matrices.

Chapter 2

Abstract Schwarz theory

In Section 1.3, we have derived the linear system that arises from the discretization of the continuous problem (1.11) and indicated that we would build preconditioners for this system, based on the decomposition of the finite element space into a sum of subspaces.

In order to analyze the convergence of the algorithms, we introduce a variational framework, known as the abstract Schwarz theory. This approach has been adopted by many authors; cf. [25, 55, 56]. For a comprehensive treatment, see the monograph by Smith, Bjørstad, and Gropp [62].

2.1 The classical Schwarz alternating method

This method was introduced by Schwarz in 1870 [58] and is the oldest domain decomposition method.

Let $\Omega = \Omega_1 \cup \Omega_2$ be a plane domain. The goal is to find the solution u to the problem

$$\begin{aligned} -\Delta u &= f \quad \text{in} \quad \Omega \\ u &= 0 \quad \text{on} \quad \partial\Omega \end{aligned}$$

To this end, we compute a sequence u_n that converges to u. We choose an initial guess u_0. Then, for $n = 1, 2, \ldots$, we solve the Dirichlet problems

$$\begin{aligned} -\Delta u_{n+1/2} &= f \quad \text{in} \quad \Omega_1 \\ u_{n+1/2} &= u_n \quad \text{on} \quad \partial\Omega_1; \end{aligned}$$

and

$$\begin{aligned} -\Delta u_{n+1} &= f \quad \text{in} \quad \Omega_2 \\ u_{n+1} &= u_{n+1/2} \quad \text{on} \quad \partial\Omega_2. \end{aligned}$$

The sequence

$$u_n = \begin{cases} u_{n+1/2} & \text{in} \quad \Omega_1 \setminus \Omega_2 \\ u_{n+1} & \text{in} \quad \Omega_2 \end{cases}$$

converges to the solution u of the original Dirichlet problem,

This algorithm can be rewritten in variational form; see [40]. Find $u \in H_0^1(\Omega)$ such that

$$\int_\Omega \nabla u \cdot \nabla v dx = \int_\Omega f v dx \quad \forall v \in H_0^1(\Omega).$$

We use the notations

$$a(u,v) = \int_\Omega \nabla u \cdot \nabla v dx; \quad f(v) = \int_\Omega f v dx.$$

The two fractional steps of the method are formulated variationally:
Find $u_{n+1/2} - u_n \in H_0^1(\Omega_1)$ such that

$$a(u_{n+1/2} - u_n, v) = f(v) - a(u_n, v), \quad \forall v \in H_0^1(\Omega_1);$$

and: Find $u_{n+1} - u_{n+1/2} \in H_0^1(\Omega_2)$ such that

$$a(u_{n+1} - u_{n+1/2}, v) = f(v) - a(u_{n+1/2}, v), \quad \forall v \in H_0^1(\Omega_2).$$

We define the orthogonal projections $P_i : H_0^1(\Omega) \to H_0^1(\Omega_1)$ by

$$a(P_i v, \phi) = a(v, \phi), \quad \forall \phi m \in V_i,$$

and we can show that

$$\begin{aligned} u - u_{n+1/2} &= (I - P_1)(u - u_n) \\ u - u_{n+1} &= (I - P_2)(u - u_{n+1/2}). \end{aligned}$$

It follows that

$$u - u_{n+1} = (I - P_2)(I - P_1)(u - u_n).$$

Thus the error propagation operator E for a complete step of the classical alternating Schwarz algorithm is

$$E = (I - P_2)(I - P_1).$$

2.2 Abstract Schwarz theory

Schwarz methods are generalizations, in an abstract setting, of the alternating Schwarz method.

Our abstract problem is: Find $u \in V$ such that

$$a(u, v) = f(v) \quad \forall v \in V. \tag{2.1}$$

Here, V is a finite-dimensional vector space, $a(\cdot, \cdot)$ a symmetric, positive definite bilinear form on V, and f a linear functional on V.

A Schwarz method is closely related to a decomposition of V into subspaces

$$V = V_0 + V_1 + \cdots + V_N.$$

In the case of our finite element problems, the subspaces correspond to subsets of degrees of freedom. It is often the case that these subsets consist of degrees of freedom associated with subdomains Ω_i of Ω. We also need a subspace that corresponds to a coarse mesh, to be responsible for the global transfer of information in each step of the iterative methods we are going to define.

To each subspace, we associate an operator T_i defined by

$$\hat{a}_i(T_i u, v) = a(u, v) \quad \forall v \in V.$$

Here, $\hat{a}_i(\cdot, \cdot)$ is a positive definite, symmetric bilinear form on the subspace V_i.

Each bilinear form $\hat{a}_i(\cdot, \cdot)$ uniquely defines the operator T_i and vice-versa. If $\hat{a}_i(\cdot, \cdot) = a(\cdot, \cdot)$ then $T_i = P_i$, the $a(\cdot, \cdot)$-orthogonal projection on V_i. We say that T_i is an approximate projection. For specific problems, we often choose T_i to be almost spectrally equivalent to P_i, but cheaper to compute.

We construct the relevant operator T for our iterative method from the operators T_i. **The classical multiplicative Schwarz algorithm** is a straightforward generalization of the two-domain Schwarz method:

1. Compute $g_i = T_i u$, $i = 1, \ldots, N$.
2. Given u_n, compute u_{n+1} in N steps:

$$u_{n+\frac{i+1}{N+1}} = u_{n+\frac{i}{N+1}} + (g_i - T_i u_{n+\frac{i}{N+1}}).$$

It is often a good idea to minimize the number of intermediate steps by *coloring* the subdomains. For each subdomain, we associate a color so that domains that share common points have different colors. We attempt to keep the number of colors small, or even to minimize it. We then merge the subspaces that correspond to domains having the same color into a single subspace and apply the multiplicative method to the new decomposition into subspaces.

This algorithm can be viewed as an iterative method for solving the equation

$$T_m u = g_m,$$

where $T_m = I-(I-T_N)\dots(I-T_1)$ and $g_m = T_m u$. The right hand side can be computed without knowledge of the solution u since T_m is a polynomial without a constant term.

The accelerated multiplicative Schwarz method: The nonsymmetric equation $T_m u = g_m$ can be solved by the GMRES method, which is a preconditioned GMRES method for the initial system; the preconditioning step is the iteration above; see [62].

The symmetrized multiplicative Schwarz method. In this case, the system we are solving is

$$T_{sm} = g_{sm},$$

where

$$\begin{aligned} T_{sm} &= I-(I-T_m)^T(I-T_m) = T_m + T_m^T - T_m^T T_m \qquad (2.2) \\ &= I-(I-T_1)\dots(I-T_{N-1})(I-T_N)(I-T_N)(I-T_{N-1})\dots(I-T_1) \end{aligned}$$

and

$$g_{sm} = T_{sm} u.$$

Since the operator T_{sm} is symmetric and positive definite, we can use the conjugate gradient method to solve the system. We can simplify the algorithm by removing one of the factors $(I-T_N)$.

The additive Schwarz method: The relevant operator is

$$T_a = T_0 + \dots T_N. \qquad (2.3)$$

Other Schwarz methods: Cai [17] uses the operator

$$\gamma T_0 + I - (I-T_N)\dots(I-T_1),$$

where $\gamma > 0$ is chosen appropriately.

We give lower and upper bounds on the spectrum of the operator (2.2) and (2.3), in this abstract context. They depend on a few parameters which we will now define.

• Let C_0^2 be the minimum constant such that for all $u \in V$ there exists a representation $u = \sum u_i$, $u_i \in V_i$ with

$$\sum_i \hat{a}_i(u_i, u_i) \le C_0^2 a(u, u).$$

• Let ϵ_{ij} be the smallest positive coefficients such that

$$|a(v_i, v_j)| \le \epsilon_{ij} a(v_i, v_i)^{1/2} a(v_j, v_j)^{1/2} \quad \forall v_i \in V_i, \ \ \forall v_j \in V_j, \ \ i, j = 1, \ldots, N.$$

Let $\mathcal{E}$ be the matrix of the coefficients ϵ_{ij} and let $\rho(\mathcal{E})$ be its spectral radius.

• Let ω be the minimum constant such that

$$a(u, u) \le \omega \hat{a}_i(u, u), \quad \forall u \in V_i, \quad i = 0, \ldots, N.$$

We can choose $\omega = \max ||T_i||_a$. We can always scale $\hat{a}_i$ such that $1 \le \omega < 2$. However, we note that this scaling will affect the value of C_0^2.

The following is a result of Bramble, Pasciak, Wang, and Xu [15], and Xu [66].

Theorem 2.1 *An upper bound on the condition number of the operator T_m is*

$$\kappa(T_{sms}) \le \frac{(1 + 2\hat{\omega}^2 \rho(\mathcal{E}^2) C_0^2}{2 - \hat{\omega}},$$

where $\hat{\omega} = \max(1, \omega)$.

A basic result for the additive case has been proven by Dryja and Widlund [26, 27] and Nepomnyaschikh [46].

Theorem 2.2 *A lower bound on the spectrum of the operator T_a is*

$$\lambda_{\min}(T_a) \ge C_0^{-2};$$

and an upper bound is

$$\lambda_{\max}(T_a) \le \omega(\rho(\mathcal{E}) + 1).$$

We note that richer choices in decomposition of u potentially decrease C_0^2.

The upper bound can be derived more directly, outside of the abstract setting; see Dryja and Widlund [28]. Suppose that we use exact solvers on subspaces. By construction, there is always an upper bound N_c on the number of functions that belong to different subspaces and are simultaneously nonzero. Then

$$\sum_{i=1}^{N} a_i(u,u) \leq N_c a(u,u),$$

where $a_i(\cdot,\cdot)$ is the restriction of $a(\cdot,\cdot)$ to the subdomain where the functions in V_i are supported. Then

$$\lambda_{\max}(P) \leq N_c. \tag{2.4}$$

Because of this, we skip the proof of the upper bound of Theorem 2.2.

We now present a proof of the lower bound in Theorem 2.2; see Dryja and Widlund [26, 27], and Zhang [67]. We use the following lemma

Lemma 2.1 *The operator $T_a = \sum_i T_i$ is invertible and*

$$a(T^{-1}v, v) = \min \sum_i \hat{a}_i(v_i, v_i),$$

where the minimum is being taken among all possible decompositions $v = \sum_i v_i$, $v_i \in V_i$. The minimum is attained at $v_i = T_i T^{-1} v$.

Proof.

$$\begin{aligned} a(T^{-1}v, v) &= \sum_i a(T^{-1}v, v_i) = \sum_i \hat{a}_i(T_i T^{-1} v, v_i) \leq \\ &\leq \Big(\sum_i \hat{a}_i(T_i T^{-1}v, T_i T^{-1}v)\Big)^{1/2} \Big(\sum_i \hat{a}_i(v_i, v_i)\Big)^{1/2} = \\ &= \Big(\sum_i a(T^{-1}v, T_i T^{-1}v)\Big)^{1/2} \Big(\sum_i \hat{a}_i(v_i, v_i)\Big)^{1/2} = \\ &= \big(a(T^{-1}v, v)\big)^{1/2} \Big(\sum_i \hat{a}_i(v_i, v_i)\Big)^{1/2}. \end{aligned}$$

It follows that

$$a(T^{-1}v, v) \leq \sum_i \hat{a}_i(v_i, v_i).$$

We remark that, for $v_i = T_i T^{-1} v$, we obtain equality. $\square$

The proof of the lower bound in Theorem 2.2 is a direct consequence of this lemma. Another direct consequence is

$$\lambda_{\min}^{-1}(T) = \lambda_{\max}(T^{-1}) = \max_v \frac{a(T^{-1}v, v)}{a(v,v)} = \max_v \min_{v=\sum v_i} \frac{\sum_i \hat{a}_i(v_i, v_i)}{a(v,v)}.$$

2.3 Matrix form of the preconditioners

Let us fix bases in each V_i and V. Let $\hat{A}_i$, A be the matrix expressions for $\hat{a}_i(\cdot,\cdot)$ and $a(\cdot,\cdot)$, respectively. Let Π_i be the matrix that contains the coefficients of the basis in V_i in terms of the basis in V. We rewrite (2.1) as

$$Ax = b.$$

Then the matrix expression of the operator T_i, in the basis of V, is

$$\Pi_i^T \hat{A}_i^{-1} \Pi_i A.$$

In the simplest case, the $\hat{a}_i(\cdot,\cdot)$ is the restriction of $a(\cdot,\cdot)$ to the subspace V_i, and V_i is spanned by a subbasis of V. The above expression then becomes, after a reordering of the basis functions,

$$\begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{pmatrix}.$$

To fix the ideas, consider the additive Schwarz method. We apply the conjugate gradient method to the system

$$(\sum_i \Pi_i^T \hat{A}_i^{-1} \Pi_i) A x = (\sum_i \Pi_i^T \hat{A}_i^{-1} \Pi_i) b,$$

i.e. we apply the preconditioned conjugate gradient method to the system $Ax = b$, with the preconditioner A_{prec}, defined by

$$A_{prec}^{-1} = \sum_i \Pi_i^T \hat{A}_i^{-1} \Pi_i. \tag{2.5}$$

Chapter 3

Algorithms

Our algorithms are preconditioned conjugate gradient methods based on the decomposition of the finite element space $\tilde{V}^p$, of discrete harmonic functions, into a sum of subspaces. There are many local spaces and one global, coarse space. The coarse space must contain the constants, otherwise the condition number will deteriorate with a growing number of subregions; cf. Dryja, Smith, and Widlund [25].

3.1 A wire basket algorithm

3.1.1 Local solvers

This algorithm is analogous to the wire basket algorithms defined by Dryja, Smith, and Widlund [25, Section 6.2], and Pavarino and Widlund [55, 56, Section 6].

An interesting theoretical feature of this algorithm is that the bound on the condition number of the global preconditioned system is the same as the local one.

We will define the coarse space $\tilde{V}_W$ as the space spanned by a special set of discrete harmonic vertex and edge functions such that it contains the constants. In Chapter 4, we will construct special, low energy vertex and edge functions for which we can prove logarithmic bounds on the condition number. In Chapter 5, we will compare the condition number and rate of convergence for algorithms based on preconditioners constructed by using low energy and standard (high energy) vertex and edge functions.

All the local spaces are associated with individual faces of the tetrahedra. For each face F_k, we define this face space $\tilde{V}_{F_k}$ as the space of discrete harmonic functions in $\tilde{V}^p$

that vanish on all the faces of the interface Γ except F_k. We obviously have

$$\tilde{V}^p = \tilde{V}_W + \sum_k \tilde{V}_{F_k}.$$

The decomposition of any $u \in \tilde{V}^p$ as a sum of functions in these subspaces is unique.

We will now explain how we can construct vertex and edge functions such that $\tilde{V}_W$ contains the constants, starting from vertex and edge functions for which $\tilde{V}_W$ does not satisfy this condition. This is the general situation, see Fig. 4.2. Let

$$\tilde{v}^{(i)}, \quad \tilde{e}_l^{(j)}, \quad \tilde{f}_m^{(k)} \tag{3.1}$$

be a given preliminary basis for $\tilde{V}^p$, where the upper index refers to the vertex, edge, or face, and the lower one to the local numbering within that edge or face.

We now present a construction given in Mandel [42], which provides an alternative approach to obtaining a low condition number. It is not essential for our work, and we could proceed without it, but is nevertheless of interest. In Chapter 5, we will see how much this approach can reduce the condition numbers.

We orthogonalize each vertex function with respect to the face spaces associated with the three faces that share the vertex and each edge function with respect to the two face spaces associated with the faces that share the edge. Returning to the reference tetrahedron, Fig. 1.1, and the local numbering of the basis functions, the new vertex function associated with the vertex V_4 is

$$\hat{v}^{(1)} = \tilde{v}^{(1)} + \sum_{m=1}^{p(p-1)/2} \alpha_{1,m}^{(2)} \tilde{f}_m^{(2)} + \sum_{m=1}^{p(p-1)/2} \alpha_{1,m}^{(3)} \tilde{f}_m^{(3)} + \sum_{m=1}^{p(p-1)/2} \alpha_{1,m}^{(4)} \tilde{f}_m^{(4)}, \tag{3.2}$$

where the $\alpha_{1,m}^{(2)}$, $\alpha_{1,m}^{(3)}$, and $\alpha_{1,m}^{(4)}$ are uniquely determined such that $\hat{v}^{(1)}$ is $a(\cdot,\cdot)$-orthogonal to the subspaces $\tilde{V}_{F_i}$, $i = 1 \ldots 3$. If we set

$$(X_{V_1F_2}\ X_{V_1F_3}\ X_{V_1F_4}) = (\alpha_{1,1}^{(2)} \ldots \alpha_{1,p(p-1)/2}^{(2)}\ \alpha_{1,1}^{(3)} \ldots \alpha_{1,p(p-1)/2}^{(3)}\ \alpha_{1,1}^{(4)} \ldots \alpha_{1,p(p-1)/2}^{(4)}),$$

then,

$$\begin{pmatrix} X_{V_1F_2}^T \\ X_{V_1F_3}^T \\ X_{V_1F_4}^T \end{pmatrix} = - \begin{pmatrix} S_{F_2F_2} & S_{F_2F_3} & S_{F_2F_4} \\ S_{F_2F_3}^T & S_{F_3F_3} & S_{F_3F_4} \\ S_{F_2F_4}^T & S_{F_3F_4}^T & S_{F_4F_4} \end{pmatrix}^{-T} \begin{pmatrix} S_{V_1F_2}^T \\ S_{V_1F_3}^T \\ S_{V_1F_4}^T \end{pmatrix}.$$

The new edge functions associated with the edge E_1 are

$$\hat{e}_l^{(1)} = \tilde{e}_l^{(1)} + \sum_{m=1}^{p(p-1)/2} \beta_{1,l,m}^{(2)} \tilde{f}_m^{(2)} + \sum_{m=1}^{p(p-1)/2} \beta_{1,l,m}^{(3)} \tilde{f}_m^{(3)}, \quad l = 1, \ldots, p-1, \tag{3.3}$$

where the $\beta_{1,l,m}^{(2)}$, $\beta_{1,l,m}^{(3)}$ are uniquely determined such that $\hat{e}_l^{(1)}$ is $a(\cdot,\cdot)$-orthogonal to the subspaces $\tilde{V}_{F_i}$, $i = 1 \ldots 3$. If we set

$$(X_{E_1F_2}\ X_{E_1F_3}) = \begin{pmatrix} \beta_{1,1,1}^{(2)} & \cdots & \beta_{1,1,p(p-1)/2}^{(2)} & \beta_{1,1,1}^{(3)} & \cdots & \beta_{1,1,p(p-1)/2}^{(3)} \\ \beta_{1,p-1,1}^{(2)} & \cdots & \beta_{1,p-1,p(p-1)/2}^{(2)} & \beta_{1,p-1,1}^{(3)} & \cdots & \beta_{1,p-1,p(p-1)/2}^{(3)} \end{pmatrix},$$

then,

$$\begin{pmatrix} X_{E_1F_2}^T \\ X_{E_1F_3}^T \end{pmatrix} = -\begin{pmatrix} S_{F_2F_2} & S_{F_2F_3} \\ S_{F_2F_3}^T & S_{F_3F_3} \end{pmatrix}^{-T} \begin{pmatrix} S_{E_1F_2}^T \\ S_{E_1F_3}^T \end{pmatrix}.$$

The other new vertex and edge functions are defined analogously. Thus, we obtain a new basis

$$\hat{v}^{(i)}, \quad \hat{e}_l^{(j)}, \quad \hat{f}_m^{(k)}, \tag{3.4}$$

where $\hat{f}_m^{(k)} = \tilde{f}_m^{(k)}$, $k = 1, \ldots, 4$, $m = 1, \ldots, p(p-1)/2$.

It is easy to see that, for symmetry reasons, the definition of the basis functions (3.4) on faces shared by two tetrahedra is consistent. It is not necessary to compute these new vertex and edge functions explicitly. If S is the local Schur complement in the preliminary basis (3.1) then the Schur complement in the new basis is

$$USU^T,$$

where

$$U = \left(\begin{array}{cccc|cccccc|cccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X_{V_1F_2} & X_{V_1F_3} & X_{V_1F_4} \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X_{V_2F_1} & 0 & X_{V_2F_3} & X_{V_2F_4} \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X_{V_3F_1} & X_{V_3F_2} & 0 & X_{V_3F_4} \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & X_{V_4F_1} & X_{V_4F_2} & X_{V_4F_3} & 0 \\
0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & X_{E_1F_2} & X_{E_1F_3} & 0 \\
0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & X_{E_2F_1} & 0 & X_{E_2F_3} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & X_{E_3F_1} & X_{E_3F_2} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & X_{E_4F_3} & X_{E_4F_4} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & X_{E_5F_2} & 0 & X_{E_5F_4} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & X_{E_6F_1} & 0 & 0 & X_{E_6F_4} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I
\end{array} \right).$$

The inverse U^{-1} is obtained by changing all the signs of the off-diagonal blocks of U. The Schur complement in the new basis has a number of null blocks below and above the diagonal. Those above the diagonal are located exactly in the same places as the nonzero blocks of U.

We now construct vertex and edge functions such that $\tilde{V}_W$ contains the constants. We proceed as in [25, 55, 56]. We consider the expansion of the constant function 1 in the basis (3.4)

$$1 = \sum_{i=1}^{4} a^{(i)} \hat{v}^{(i)} + \sum_{j=1}^{6} \sum_{l=1}^{p-1} b_l^{(j)} \hat{e}_l^{(j)} + \sum_{k=1}^{4} \sum_{m=1}^{p(p-1)/2} c_m^{(k)} \hat{f}_m^{(k)}.$$

We define the special face functions by

$$\mathcal{F}_k = \sum_{m=1}^{p(p-1)/2} c_m^{(k)} \hat{f}_m^{(k)}, \tag{3.5}$$

and the new vertex, edge, and face basis functions by

$$v^{(i)} = \hat{v}^{(i)} + \sum_{k=1}^{4} \mathcal{F}_k \frac{\int_{\partial F_k} v^{(i)}}{\int_{\partial F_k} 1}, \quad k = 1, \ldots, 4,\ i = 1, \ldots, 4, \tag{3.6}$$

$$e_l^{(j)} = \hat{e}_l^{(j)} + \sum_{k=1}^{4} \mathcal{F}_k \frac{\int_{\partial F_k} e_l^{(j)}}{\int_{\partial F_k} 1}, \quad k = 1, \ldots, 4,\ j = 1, \ldots, 6,\ l = 1, \ldots, 6, \tag{3.7}$$

$$f_m^{(k)} = \hat{f}_m^{(k)}, \quad m = 1, \ldots, p(p-1)/2,\ k = 1, \ldots, 4. \tag{3.8}$$

It is easy to see that the wire basket space spanned by these vertex and edge functions contains the constants.

Once again, for symmetry reasons, the definitions of the basis functions (3.6 - 3.8) on faces shared by two tetrahedra is consistent. If S is the local Schur complement in the preliminary basis (3.4) then the Schur complement in the new basis is

$$USU^T,$$

where

$$U = \begin{pmatrix} I & U_{WF} \\ 0 & I \end{pmatrix}$$

with

$$U_{WF} = \sum_{k=1}^{4} \frac{z_{F_k} z_W^T M_{F_k}}{z_W^T M_{F_k} z_W}.$$

Here M_{F_k} is the mass matrix of the boundary of face F_k, and z_{F_k} and z_W are the vectors containing the coefficients of the constant function 1 that correspond to the face F_k and the wire basket W, respectively.

On the coarse space, we can use the exact solver $a(\cdot,\cdot)$, or, more economically, an inexact solver based on the bilinear form

$$\hat{a}_W(u,u) = (1+\log p)\rho_i \sum_i \inf_{c_i} ||u - c_i||^2_{L^2(W_i)}.$$

The preconditioner for the additive Schwarz method, using the exact solver on $\tilde{V}_W$, has the following matrix form:

$$S_{prec} = \begin{pmatrix} S_{WW} & 0 & 0 & 0 \\ 0 & S_{F_1F_1} & 0 & 0 \\ 0 & 0 & S_{F_2F_2} & 0 \\ 0 & 0 & 0 & \ddots \end{pmatrix}. \tag{3.9}$$

If we use the inexact solver $\hat{a}_W(\cdot,\cdot)$, the block S_{WW} is replaced by

$$\hat{S}_{WW} = (1+\log p)\left(M - \sum_i \frac{(M^{(i)}z^{(i)}) \cdot (M^{(i)}z^{(i)})^T}{z^{(i)T} M^{(i)} z^{(i)}} \right),$$

where $M^{(i)}$ is the mass matrix of the wire basket W_i, and $z^{(i)}$ is the vector containing the coefficients of the constant function 1. The edge blocks of the mass matrix M are

Table 3.1: The number of vertex and wire basket degrees of freedom

p	Number of vertices			Number of wire basket d.o.f.		
	24 tetrahedra	192 tetrahedra	684 tetrahedra	24 tetrahedra	192 tetrahedra	684 tetrahedra
4	15	71	199	165	1001	3061
5	15	71	199	215	1311	4015
6	15	71	199	265	1621	4969
7	15	71	199	315	1931	5923
8	15	71	199	365	2241	6877
9	15	71	199	415	2551	7831
10	15	71	199	465	2861	8785

tridiagonal because the restrictions of the edge functions to the edges are integrated Legendre polynomials, for which a recurrence relation that involves only three terms holds.

In Chapter 4, we will construct vertex functions with traces on the wire basket which are L^2-orthogonal to the traces of the edge functions and with mutual L^2-scalar products that are negligible (4.20). We can then simplify the preconditioner further by replacing each $M^{(i)}$ and M by a block diagonal matrix $D^{(i)}$ and D, obtained from $M^{(i)}$ and M, respectively, by dropping the couplings between pairs of vertices. Since these couplings are negligible, there are c and C such that

$$cx^{(i)}D^{(i)}x^{(i)} \leq x^{(i)}M^{(i)}x^{(i)} \leq Cx^{(i)}D^{(i)}x^{(i)}.$$

3.1.2 The coarse problem

The coarse problem, i.e. solving $\hat{S}_{WW}x_W = r_W$ or $S_{WW}x_W = r_W$ can be quite expensive if we use a direct method. See Table 3.1 for a comparison of the sizes of the coarse problem for a method with a vertex-based coarse space (see Section 3.2) and a wire basket-based space, respectively.

The following method, introduced in [13, 14], allow us to solve an auxiliary problem which first provides the local averages of x_W. See also Mandel [42] and Dryja, Smith, and Widlund [25] for further details. It is quite cheap to solve this problem, since it has only one degree of freedom per element.

Solving the system $\hat{S}_{WW}x_W = r_W$ is equivalent to solving the minimization problem

$$\min_x \sum_i \min_{c_i} \frac{1}{2}(x^{(i)} - c_i z^{(i)})^T D_i (x^{(i)} - c_i z^{(i)}) - x^T r.$$

Taking derivatives with respect to the $x^{(i)}$ and c_i, we obtain the system:

$$(z^{(i)^T} D^{(i)} z^{(i)})c_i - c_i^T D^{(i)} M^{-1} \sum_j D^{(j)} z^{(j)} c_j = z^{(i)^T} D^{(i)} D^{-1} r.$$

$$Dx - \sum_i D^{(i)} z^{(i)} c_i = r. \tag{3.10}$$

Eliminating x, we obtain a system for the c_i:

$$(z^{(i)^T} D^{(i)} z^{(i)})c_i - z^{(i)^T} D^{(i)} D^{-1} \sum_j z^{(i)} c_j = z^{(i)^T} D^{(i)} D^{-1} r. \tag{3.11}$$

Finally, we compute x by solving the system (3.10).

We remark that it is quite cheap to solve the systems (3.11) and (3.10) since the mass matrix D is tridiagonal, unlike the wire basket block $\hat{S}_{WW}$.

3.2 A vertex-based algorithm

In this section, we consider the method defined by the decomposition into subspaces

$$\tilde{V}^p = \tilde{V}_0 + \sum_{i=1}^{n_V} \tilde{V}_{V_i} + \sum_{j=1}^{n_E} \tilde{V}_{E_j} + \sum_{k=1}^{n_F} \tilde{V}_{F_k},$$

where $\tilde{V}_0$ is the space of piecewise linear functions on the whole region Ω. This a block-Jacobi method using the subspaces associated with the vertices, edges, faces, augmented by a coarse solver.

The piecewise linear functions are degrees of freedom associated with the vertices, but it turns out, see Theorem 4.8, that it is necessary to introduce one extra degree of freedom for each vertex, in order to make the condition number independent of the number of substructures.

Finally, in this case, the space spanned by the vertex and edge functions need not contain the constant functions, since the coarse space of piecewise linear functions already contains them.

The preconditioner for the additive method is defined by

$$S_{prec}^{-1} = \Pi S_{coarse}^{-1} \Pi + S_J^{-1},$$

where

$$S_J = \begin{pmatrix} S_{V_1V_1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 \\ 0 & 0 & S_{E_1E_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & S_{F_1F_1} & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots \end{pmatrix} \tag{3.12}$$

and Π is the interpolation matrix from $\tilde{V}^p$ to $\tilde{V}_0$.

3.3 Algorithms using overlap

It easy to see, from the abstract theory that a richer choice in a decomposition of any $u \in \tilde{V}^p$ as a sum of functions in the subspaces potentially improves the lower bound on the eigenvalues of the relevant operator T for our method; see Theorem 2.2.

For the previous two algorithms, we can enlarge the face spaces by adding some of the edge degrees of freedom that are coupled to the face in question in the original finite element model. We can, e.g., add all the first and the second edge basis functions, or we can do this only for the edges that are coupled to the face in question, but do not belong to the face, if we use the basis (3.4). To fix the ideas, given the face F_3, of the reference tetrahedron, we can add to its face space some of the degrees of freedom associated with the edges E_3, E_5, and E_6. In Chapter 5, we will investigate which edge degrees of freedom are the main cause of the ill conditioning and add them to the face spaces to which they are coupled.

There are disadvantages with such a strategy. The face spaces no longer consist solely of functions that are nonzero on that face only, but also contain functions that are nonzero on more that one face. We now use the notations in Fig. 1.6. Denote a face by the triplet of its vertices and an edge by the ordered pair of its endpoints. Consider the edge $(4, 15)$. The faces that are coupled but adjacent to this edge, are: $(1, 4, 9)$, $(1, 9, 15)$, $(1, 4, 14)$, $(1, 14, 15)$, $(4, 8, 14)$, $(8, 14, 15)$, $(3, 4, 9)$, and $(3, 9, 15)$. This means that, for any point on the edge $(4, 15)$, there are be nine functions that are nonzero, and belong to different spaces. This can result in an upper bound that is nine times as large as before the overlap was introduced.

Moving now to the global numbering of the faces, the new decomposition into subspaces is

$$\tilde{V}^p = \tilde{V}_W + \sum_k \tilde{V}'_{F_k},$$

where the $\tilde{V}'_{F_k}$ are the subspaces spanned by the $\tilde{V}_{F_k}$, and the edge functions as just indicated. The decomposition into subspaces $\tilde{V}^p$ is no longer a direct decomposition.

In conclusion, we expect to obtain

- a larger upper bound, which is bad for our purpose.

- a larger lower bound, which is good.

We hope that the increase of lower bound might compensate for the increase of the upper bound.

The preconditioner is no longer block-diagonal. Instead, we have

$$S_{prec}^{-1} = \tilde{S}_{WW}^{-1} + \sum_{k=1}^{n_F} \Pi_{F_k}^T \; S'^{-1}_{F_k F_k} \; \Pi_{F_k},$$

where Π_{F_k} is the matrix form of the identity operators $\tilde{V}_{F_k} \rightarrow \tilde{V}^p$.

3.4 Neumann-Neumann algorithms

In the previous algorithms, we eliminate the couplings between all pairs of faces and the wire basket or between all pairs of faces, edges, and vertices. In this section, we adopt a different approach. We keep all the couplings between the vertices, edges, and faces of the individual tetrahedra, but decouple all the tetrahedra. For each tetrahedron Ω_i, the space $\tilde{V}_{\Omega_i}$ is spanned by the vertex, edge, and face basis functions associated with its vertices, edges, and faces. The preconditioner is based on the decomposition into subspaces given by

$$\tilde{V}^p = \sum_i \tilde{V}_{\Omega_i},$$

and is built from the local Schur complements S_i.

The auxiliary bilinear forms are given by

$$b_i(u,v) = a_i(\mathcal{H}_i(\nu_i(u)), \mathcal{H}_i(\nu_i(v))).$$

Here $\mathcal{H}_i(w)$ is the discrete harmonic extension of the finite element interpolant of a continuous function w, defined on $\partial\Omega_i$. For the $h-$version, the counting functions ν_i are defined on the set of degrees of freedom by:

$$\begin{aligned} \nu_i(x) &= \text{number of } \partial\Omega_j \text{ to which } x \in \partial\Omega_i \text{ belongs}, \\ \nu_i(x) &= 0 \;\; x \in (\Gamma \cup \partial\Omega) \setminus \partial\Omega_i. \end{aligned}$$

The definition of the counting functions for the $p-$version is similar, except that the degrees of freedom are no longer associated with nodes. Let ϕ be any basis function associated with a vertex, edge, or face of the tetrahedron Ω_i. Then $\nu_i(\phi)$ is the number of substructures Ω_j to which that vertex, edge, or face belongs. If ϕ is a face function associated to a face that is not part of the boundary, then $\nu_i(\phi) = 2$. If the face is part of the boundary, then $\nu_i(\phi) = 1$. If ϕ is a vertex or edge function associated to a vertex or edge that are not part of the boundary, then $\nu_i(\phi) > 2$.

If $u = \sum_i \alpha_i \phi_i$, we define $\nu_i(u)$ by

$$\nu_i(u) = \sum_i \alpha_i \nu_i(\phi_i).$$

We define the pseudoinverses $\nu_i^\dagger$ by

$$\begin{aligned} \nu_i^\dagger(\phi) &= \nu_i^{-1}(\phi) \text{ if } \nu_i(\phi) \neq 0, \\ \nu_i^\dagger(\phi) &= 0 \text{ otherwise }. \end{aligned}$$

The preconditioner is given by

$$S_{prec}^{-1} = \sum_i N_i {S^{(i)}}^\dagger N_i,$$

where the N_i are diagonal matrices and the diagonal entry corresponding to the basis function ϕ is $\nu_i^\dagger(\phi)$.

For previous work, see Dryja and Widlund [29], Bourgat, Glowinski, Le Tallec, and Vidrascu [12] Le Tallec, De Roeck, and Vidrascu [38], Le Tallec and De Roeck [21], Cowsar, Mandel, and Wheeler [20], Kuznetsov, Manninen, and Vassilevski [36], Mandel [43], and Mandel and Brezina [45, 44].

A major technical difficulty stems from the fact that the local components S_i of the preconditioner, corresponding to the tetrahedra that do not touch the Dirichlet part of the boundary are singular. The components corresponding to tetrahedra that touch the Dirichlet part of the boundary, are not singular. There are several ways to deal with this.

1. We can work with pseudoinverses ${S^{(i)}}^{\dagger}$ of the local Schur complements.

2. We can follow the approach taken in Dryja and Widlund [29], and solve local problems using a different elliptic operator, for the substructures that do not touch the boundary:

$$\hat{a}_i(u,u) = \int_{\Omega_i} (\nabla u)^2 dx + \frac{1}{H_i^2} \int_{\Omega_i} u^2 dx.$$

Dryja and Widlund [29] contains a detailed discussion concerning those substructures that touch the boundary. If a given substructure touches the Dirichlet part of the boundary along an edge or face, then the bilinear form $\hat{a}_i(\cdot,\cdot)$ is used on that substructure. If it touches the boundary at a vertex only, then the bilinear for $a(\cdot,\cdot)$ is used.

3. We can impose zero Dirichlet boundary conditions at some points of the wire basket of the local substructures, e.g., one vertex, all the vertices, or the whole wire basket of each substructure. In such a case, we must add a component to the preconditioner that corresponds to the degrees of freedom set to zero by these local Dirichlet boundary conditions.

Chapter 4

Auxiliary results and main theorems

In this chapter, we construct the low energy vertex and edge functions announced in Section 1.3 and prove upper bounds on the H^1-norm of the vertex, edge, and face components of a discrete harmonic function. We then use these technical tools to prove bounds on the condition numbers of the algorithms defined in Chapter 3.

4.1 Technical tools

We begin by collecting several general results concerning Sobolev norms and polynomials.

The following construction and lemma are given in [9, Lemma 7.1] and [41, formula (4)], [11, Theorem 2.2]. For $f \in P^p(I_1)$, we define

$$F_1^{[f]}(x,y) = \frac{1}{y}\int_x^{x+y} f(t)dt, \tag{4.1}$$

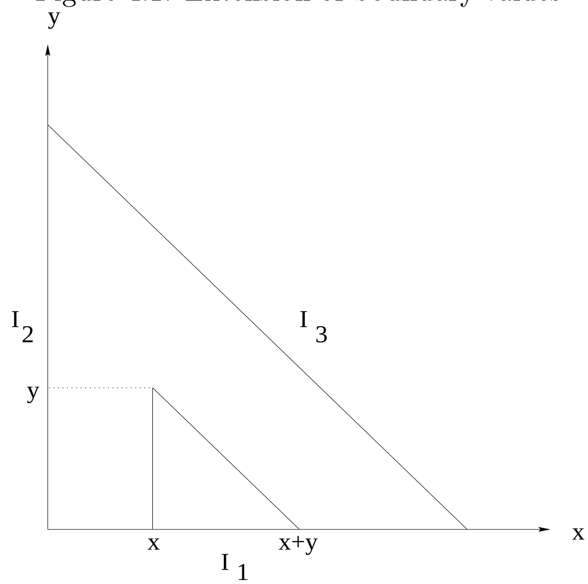
i.e. the value of $F_1^{[f]}$ at the point (x, y) is the average of the values of f on the segment $[(x, 0), (x + y, 0)]$; see Fig. 4.1. We denote the similar extensions of $f \in I_2$ and $f \in I_3$ by $F_2^{[f]}$ and $F_3^{[f]}$, respectively. It is easy to see that $F_1^{[f]} \in P^p(T)$ and that $F_1^{[f]}(x, 0) = f(x)$.

Lemma 4.1 *Let $f \in P^p(I_1)$ and let $F_1^{[f]}$ be defined by (4.1). Then,*

$$||F_1^{[f]}||_{H^{s+1/2}(T)} \;\; \leq \;\; C||f||_{H^s(I_1)}. \tag{4.2}$$

The following is the classical Markov inequality; see, e.g., Rivlin [57].

Figure 4.1: Extension of boundary values



Lemma 4.2 *Let $f \in Q^p([0,1])$. Then,*

$$\max_{[0,1]} \left| \frac{df}{dx}(x) \right| \leq 2p^2 \max_{[0,1]} |f(x)|.$$

The following result is Theorem 6.2 in Babuška, Craig, Mandel, and Pitkäranta [9].

Lemma 4.3 *Let $u \in Q^p([0,1] \times [0,1])$. Then,*

$$||u||^2_{L^\infty([0,1]\times[0,1])} \leq C(1 + \log p)||u||^2_{H^1([0,1]\times[0,1])}.$$

If $x_0 \in [0,1] \times [0,1]$, then

$$||u - u(x_0)||^2_{L^\infty([0,1]\times[0,1])} \leq C(1 + \log p)|u|^2_{H^1([0,1]\times[0,1])}.$$

The following is Lemma 5.3. in Pavarino and Widlund [55].

Lemma 4.4 *Let I be any line segment in the closed reference tetrahedron $\bar{\Omega}_{ref}$ and let $u \in P^p(\Omega_{ref})$. Then,*

$$||u||^2_{L^2(I)} \leq C(1 + \log p)||u||^2_{H^1(\Omega_{ref})}.$$

If $\bar{u}_W$ is the average of u over the wire basket W, then,

$$||u - \bar{u}_W||^2_{L^2(I)} \leq C(1 + \log p)|u|^2_{H^1(\Omega_{ref})}.$$

We remark that the original Lemma 5.3 in [55] requires I to be parallel to a coordinates axis, and that $u \in Q^p([-1,1]^3)$. Since we only have $u \in P^p(\Omega_{ref})$, we can let I be any line segment, without violating the hypotheses of that lemma.

The following is analogue to Lemma 1 in Dryja [24].

Lemma 4.5 *Let $u \in P^p([0,1])$. Then,*

$$||u||^2_{L^\infty([0,1])} \leq C(1+\log p)||u||^2_{H^{1/2}([0,1])}.$$

Proof. Let U be the extension (4.1) of u to the reference triangle T. From Lemma 4.1, we have

$$||U||_{H^1(T)} \leq C||u||_{H^{1/2}(I_1)}.$$

We then combine this inequality and the one provided by Lemma 4.3. □

The following is the classical Hardy's inequality [34].

Lemma 4.6 *Let $f \geq 0$. Then,*

$$\int_0^{1-x} \frac{1}{y^2}\left(\int_x^{x+y} f(t)dt\right)^2 dy \leq 4\int_x^1 f(t)^2 dt.$$

The following is a weaker variant of Lemma 5.9. in Pavarino and Widlund [55]. It is important because it gives a bound of the $H^{1/2}_{00}-$norm in terms of the $H^{1/2}-$norm.

Lemma 4.7 *Let $u \in P^p(T)$ and $u = 0$ on I_1. Then,*

$$\int_T \frac{u(x,y)^2}{y} dydx \leq C(1+\log p)^2 ||u||^2_{H^{1/2}(T)}.$$

Proof. We divide the double integral into an integral over a thin slice along the x-axis and over the rest of the triangle.

$$\int_0^1 \frac{1}{y}\int_0^{1-y} u^2 dxdy = \int_0^{1/p^2} \frac{1}{y}\int_0^{1-y} u^2 dxdy + \int_{1/p^2}^1 \frac{1}{y}\int_0^{1-y} u^2 dxdy. \tag{4.3}$$

The second integral is bounded by

$$\int_{1/p^2}^1 \frac{1}{y}\int_0^{1-y} u^2 dxdy \leq \max_y\left(\int_0^{1-y} u^2 dx\right)\int_{1/p^2}^1 \frac{1}{y}dy \leq C\ \log p\ \max_y\left(\int_0^{1-y} u^2 dx\right).$$

From Lemma 4.4, we obtain

$$\max_y\left(\int_0^{1-y} u^2 dx\right) \leq C(1+\log p)\int_0^1 ||u(\cdot,y)||^2_{H^{1/2}([0,1-y])}dy. \tag{4.4}$$

By using an equivalent formula for the $H^{1/2}-$norm; see [47], [24], we obtain

$$\max_y \left(\int_0^{1-y} u^2 dx \right) \leq C(1 + \log p) ||u||^2_{H^{1/2}(T)}.$$

We now bound the first integral in (4.3). We observe that $\int_0^{1-y} u(x,y)^2 dx$ is a polynomial of degree $2p$ in y and use the mean value theorem and Lemma 4.2 to obtain

$$\begin{aligned}
& \int_0^{1/p^2} \frac{1}{y} \int_0^{1-y} u(x,y)^2 dx dy \\
= & \int_0^{1/p^2} \frac{1}{y} \left(\int_0^{1-y} u(x,y)^2 dx - \int_0^1 u(x,0)^2 dx \right) dy \\
= & \int_0^{1/p^2} y \frac{d}{dy} \int_0^{1-\tilde{y}} u(x,\tilde{y})^2 dx dy \\
\leq & \; Cp^2 \max_y \left(\int_0^{1-y} u(x,y)^2 dx \right) \int_0^{1/p^2} dy \\
\leq & \; C \max_y \left(\int_0^{1-y} u(x,y)^2 dx \right).
\end{aligned}$$

The last member of the sequence of inequalities can be bounded as in (4.4). □

The following is Lemma 4.2 in Bramble and Xu [16].

Lemma 4.8 *Let Q be the L^2-projection operator onto the coarse space V_0. Then*

$$||u - Qu||^2_{L^2(\Omega)} \leq CH^2 |u|^2_{H^1(\Omega)}.$$

and

$$|Qu|^2_{H^1(\Omega)} \leq C|u|^2_{H^1(\Omega)}$$

4.2 Extension theorems and vertex and edge functions

In this section, we construct the low energy vertex and edge functions on the reference tetrahedron Ω_{ref} (1.5). Our work involves the use and further development of extension theorems for polynomials given in Muñoz-Sola [48], Babuška, Craig, Mandel, and Pitkäranta [9], Babuška and Suri [3], and Maday [41].

The corresponding results are well known for Sobolev spaces, but their extension to finite element spaces is quite intricate. Similar constructions for the h-version and for spectral elements have been given in Dryja, Smith, and Widlund [25] and Pavarino and Widlund in [56, 55], respectively. The difference between the case of spectral elements and our case, of tetrahedral substructures, is that our wire basket functions are polynomials in P^p, whereas in the other case they are polynomials in Q^p. The use of separation of variables makes the Q^p case much easier. Different ways of constructing these functions as well as different stability results are needed.

4.2.1 Extension theorems (I)

The constructions in this section are basically those of Babuška, Craig, Mandel, and Pitkäranta [9] and Maday [41]. Let T be the reference triangle (1.6), shown in Fig. 1.2. The following is Theorem 7.4 in [9]:

Theorem 4.1 *Let f be a continuous function on the reference triangle T such that $f_i \in P^p(I_i)$, $i = 1, 2, 3$. Then there exists a $U \in P^p(T)$ such that $U = f$ on ∂T and*

$$||U||_{H^1(T)} \leq C||f||_{H^{1/2}(\partial T)}.$$

Maday [41, Theorem 4] proves a more general result:

Theorem 4.2 *Let f be a continuous function on the reference triangle T such that $f_i \in P^p(I_i)$, $i = 1, 2, 3$. Then there exists a $U \in P^p(T)$ such that $U = f$ on ∂T and*

$$||U||_{H^{s+1/2}(T)} \leq C||f||_{H^s(\partial T)} \;\; \forall s \geq 0.$$

The proof of the main theorems of Section 4.4 require similar bounds on some $H_{00}^{1/2}$ -norms.

Theorem 4.3 *Let f be a continuous function on the reference triangle T such that $f_i \in P^p(I_i)$, $i = 1, 2, 3$. Then there exists a $U \in P^p(T)$ such that $U = f$ on ∂T and*

$$||U||^2_{H^{1/2}(T)} \leq C||f||^2_{L^2(\partial T)}. \tag{4.5}$$

If $f_3 = 0$, then

$$||U||^2_{H_{00}^{1/2}(T,I_3)} \leq C(1 + \log p)^2||f||^2_{L^2(\partial T)}. \tag{4.6}$$

If $f_2 = f_3 = 0$, then

$$||U||^2_{H_{00}^{1/2}(T,I_2,I_3)} \leq C(1 + \log p)^2||f||^2_{L^2(\partial T)}. \tag{4.7}$$

Proof. Inequality (4.5) is proven in Maday [41, Theorem 4]. Inequalities (4.6-4.7) follow immediately from (4.5) by using Lemma 4.7. □

Our conjecture is that the logarithms in (4.6-4.7) can be dropped. In Chapter 5, we present numerical experiments that support this conjecture.

The proof of the theorems cited above are based on the extension (4.1) and Lemma 4.1, case $s = 0$.

4.2.2 Extension theorems (II)

The construction in this section is basically given in Muñoz-Sola [48]. The following is Theorem 1 in [48]:

Theorem 4.4 *Let f be a continuous function on the reference tetrahedron Ω_{ref} such that $f_i \in P^p(F_k)$, $i = 1, 2, 3, 4$. Then there exist $U \in P^p(\Omega_{ref})$ such that $U = f$ on $\partial\Omega_{ref}$ and*

$$||U||_{H^1(\Omega_{ref})} \leq C||f||_{H^{1/2}(\partial\Omega_{ref})}.$$

This result is important for several different reasons. First, it is used to obtain approximation results for the $p-$version finite element, as well as to deal with nonhomogeneous boundary conditions [48]. It also provides a necessary inequality in the equivalence between the $a(\cdot,\cdot)-$ and $H^{1/2}-$norms in the space of discrete harmonic functions. We have briefly reviewed these matters in Section 1.3. Finally, we will also use a $2D$ analogue, Theorem 4.6 given below, to provide an alternative recipe for the construction of our vertex and edge functions.

We also consider the simpler, $2D$ version of this theorem. Here, T is the reference triangle (1.6).

Theorem 4.5 *Let f be a continuous function on the reference triangle T such that $f_i \in P^p(I_i)$, $i = 1, 2, 3$. Then there exist $U \in P^p(T)$ such that $U = f$ on ∂T and*

$$||U||_{H^1(T)} \leq C||f||_{H^{1/2}(\partial T)}.$$

We are not going to give a proof, since this is a much simpler variant of Theorem 4.4. The result we prove depends on a factor $N(p)$ to be defined later in the section.

Theorem 4.6 *Let f be a continuous function on the reference triangle T such that $f_i \in P^p(I_i)$, $i = 1, 2, 3$. Then there exist $U \in P^p(T)$ such that $U = f$ on ∂T and*

$$||U||_{H^{1/2}(T)} \leq C\ N(p)||f||_{L^2(\partial T)}. \tag{4.8}$$

If $f_3 = 0$, then

$$||U||_{H_{00}^{1/2}(T, I_3)} \leq C\ N(p)||f||_{L^2(\partial T)}. \tag{4.9}$$

If $f_2 = f_3 = 0$, then

$$||U||_{H_{00}^{1/2}(T, I_2, I_3)} \leq C\ N(p)||f||_{L^2(\partial T)}. \tag{4.10}$$

A proof of this theorem is given later in this section. We believe that $N(p)$ can be eliminated. In Chapter 5, we present numerical experiments that support this conjecture.

Lemma 4.9 *Let $f \in P^p(I_1)$ and $F^{[f]}$ be defined by (4.1). Then,*

$$\begin{aligned} ||F_1^{[f]}||_{H^{1/2}(T)} &\leq C||f||_{L^2(I_1)}. & (4.11)\\ ||F_1^{[f]}||_{L^2(I_2)} &\leq C||x^{1/2}f||_{L^2(I_1)}. & (4.12)\\ ||F_1^{[f]}||_{L^2(I_3)} &\leq C||(1-x)^{1/2}f||_{L^2(I_1)}. & (4.13) \end{aligned}$$

Proof. For (4.11), see Lemma 4.1.

We now prove (4.12):

$$||F_1^{[f]}||^2_{L^2(T)} = \int_0^1 \int_0^{1-x} \frac{1}{y^2} \left(\int_x^{x+y} f(t)dt \right)^2 dydx$$

From Lemma 4.6, we deduce

$$\int_0^{1-x} \frac{1}{y^2} \left(\int_x^{x+y} f(t)dt \right)^2 dz \leq 4 \int_x^1 f(z)^2 dz.$$

It follows that

$$\begin{aligned} ||F_1^{[f]}||^2_{L^2(T)} &\leq 4 \int_0^1 \int_x^1 f(z)^2 dzdx = 4 \int_0^1 f(z)^2 \int_0^z dxdz = \\ &= 4 \int_0^1 f(x)^2 x dx = 4||x^{1/2}f||^2_{L^2([0,1])}. \end{aligned}$$

We prove (4.13) similarly. $\square$

The following constructions and lemmas provide the main result of this subsection. For $f \in P^p(I_1)$, $f(0) = 0$, let

$$E_1^{[f]}(x,y) = \frac{x}{y}\int_x^{x+y} \frac{f(t)}{t}dt. \tag{4.14}$$

For $f \in P^p(I_1)$, with $f(1) = 0$, let

$$E_2^{[f]}(x,y) = \frac{1-x-y}{y}\int_x^{x+y} \frac{f(t)}{1-t}dt. \tag{4.15}$$

Finally, if $f \in P^p(I_1)$, with $f(0) = f(1) = 0$, let

$$E^{[f]}(x,y) = \frac{x(1-x-y)}{y}\int_x^{x+y} \frac{f(t)}{t(1-t)}dt. \tag{4.16}$$

Lemma 4.10 *If $f \in P^p(I_1)$, $f(0) = 0$, then $E_1^{[f]} = 0$ on I_2, $E_1^{[f]} = f$ on I_1, and*

$$||E_1^{[f]}||_{H_{00}^{1/2}(T,I_2)} \leq CN(p)||f||_{L^2(I_1)}. \tag{4.17}$$

If $f \in P^p(I_1)$, $f(1) = 0$, then $E_2^{[f]} = 0$ on I_3, $E_2^{[f]} = f$ on I_1, and

$$||E_2^{[f]}||_{H_{00}^{1/2}(T,I_3)} \leq CN(p)||f||_{L^2(I_1)}. \tag{4.18}$$

If $f \in P^p(I_1)$, $f(0) = f(1) = 0$, then $E^{[f]} = 0$ on I_2 and I_3, $E^{[f]} = f$ on I_1, and

$$||E^{[f]}||_{H_{00}^{1/2}(T,I_2,I_3)} \leq CN(p)||f||_{L^2(I_1)}. \tag{4.19}$$

Here, $N(p)$ is the smallest constant such that

$$||E_1^{[f]}||_{L^2(T)} \leq N(p)||F_1^{[f]}||_{L^2(T)}.$$

Proof. First, we prove (4.17).

$$||E_1^{[f]}||_{H^{1/2}(T)} \leq C\ N(p)||f||_{L^2(\partial T)}.$$

By Theorem 4.5 and a trace theorem, we have

$$||E_1^{[f]}||_{H^1(T)} \leq C||f||_{H^{1/2}(\partial T)} \leq C||F_1^{[f]}||_{H^1(T)}.$$

By an interpolation argument and Lemma 4.9

$$||E_1^{[f]}||_{H^{1/2}(T)} \leq CN(p)||f||_{L^2(\partial T)}.$$

In order to complete the proof of (4.17), we must bound the weighted L^2-norms, in the expression of $||E_1^f||_{H_{00}^{1/2}(T,I_2)}$, from above.

$$||x^{-1/2}E_1^{[f]}||_{L^2(T)}^2 \quad = \quad \int_0^1\int_0^{1-x}\frac{x}{y^2}\left(\int_x^{x+y}\frac{f(t)}{t}dt\right)^2 dydx.$$

From Lemma 4.6, we deduce

$$\int_0^{1-x}\frac{1}{y^2}\left(\int_x^{x+y}\frac{f(t)}{t}dt\right)^2 dz \le 4\int_x^1\frac{f(z)^2}{z^2}dz.$$

It follows

$$\begin{aligned} ||x^{-1/2}E_1^{[f]}||_{L^2(T)}^2 &\le 4\int_0^1 x\int_x^1\frac{f(z)^2}{z^2}dzdx = 4\int_0^1\frac{f(z)^2}{z^2}\int_0^z xdxdz = \\ &= 4\int_0^1\frac{f(z)^2}{z^2}\frac{z^2}{2}dz = 2||f||_{L^2([0,1])}^2. \end{aligned}$$

The proof of (4.18) is analogous to the proof of (4.17), given above.

We now prove (4.19). Because $E^{[f]}(x,y) = (1-x-y)E_1^{[f]}(x,y) + xE_2^{[f]}(x,y)$, we need to look at the first term only.

$$\begin{aligned} &|(1-x-y)E_1^{[f]}(x,y) - (1-x'-y')E_1^{[f]}(x',y')|^2 \le \\ &2(1-x-y)^2|E_1^{[f]}(x,y) - E_1^{[f]}(x',y')|^2 + 2(x'-x+y'-y)^2|E_1^{[f]}(x',y')|^2. \end{aligned}$$

It follows

$$\begin{aligned} &|(1-x-y)E_1^{[f]}|_{H^{1/2}}^2 \\ \le\ & C\Big(|E_1^{[f]}|_{H^{1/2}}^2 + \int_0^1\int_0^{1-y}\int_0^1\int_0^{1-y'}\frac{(x'-x+y'-y)^2E_1^{[f]}(x,y)^2}{((x'-x)^2+(y'-y)^2)^{3/2}}dx'dy'dxdy\Big) \\ \le\ & C(|E_1^{[f]}|_{H^{1/2}}^2 + ||E_1^{[f]}||_{L^2}^2). \end{aligned}$$

Also,

$$\begin{aligned} ||x^{-1/2}(1-x-y)E_1^{[f]}||_{L^2}^2 &\le \int_0^1\int_0^{1-y}\frac{x(1-x-y)^2}{y^2}\left(\int_x^{x+y}\frac{f(t)}{t(1-t)}\right)^2 dxdy \\ &\le \int_0^1\int_0^{1-y}\frac{x}{y^2}\left(\int_x^{x+y}\frac{f(t)}{t}\right)^2 dxdy \le C||f||_{L^2}. \end{aligned}$$

□

We next prove Theorem 4.6.

Proof. The inequality (4.10) is exactly (4.19). We now prove (4.9). Let $U_1 = E_2^{[f_1]}$ and let g_2 be its trace on I_2. Consider now $U_2 \in P^p(T)$ such that $U_2 = g_2 - f_2$ on I_2, $U_2 = 0$ on I_1 and I_3, and

$$||U_2||_{H_{00}^{1/2}(T,I_1,I_3)} \leq C\ N(p)||g_2 - f_2||_{L^2(I_2)}.$$

Finally, let $U = U_1 + U_2$. We only have to prove that

$$||g_2||_{L^2(I_2)} \leq C||f||_{L^2(I_1)}.$$

This follows from Lemma 4.6 in the same way as in the previous lemmas.

Finally, we prove (4.8). Let $U_1 = F_1^{[f_1]}$ and let g_2 be its trace on I_2. Consider now $U_2 \in P^p(T)$ such that $U_2 = f_2 - g_2$ on I_2, $U_2 = 0$ on I_1 and

$$||U_2||_{H_{00}^{1/2}(T,I_1)} \leq C\ N(p)||f_2 - g_2||_{L^2(I_2)}.$$

Let g_3 be the trace of $U_1 + U_2$ on I_3. Finally, consider $U_3 \in P^p(T)$ such that $U_3 = f_3 - g_3$ on I_2, $U_2 = 0$ on I_1 and I_3. We obtain

$$||U_3||_{H_{00}^{1/2}(T,I_1,I_2)} \leq C\ N(p)||f_3 - g_3||_{L^2(I_3)}.$$

We finally set $U = U_1 + U_2 + U_3$. □

4.2.3 Construction of the vertex and edge functions

We now turn to the construction of the vertex functions. We need a preliminary result concerning a special polynomial in one variable; cf. Pavarino and Widlund [55]. We denote by $\phi_0(x)$ the degree p polynomial of minimal L^2 norm satisfying $\phi_0(0) = 1,\ \phi_0(1) = 0$. It follows immediately from the definition that ϕ_0 is L^2-orthogonal to P_0^p. The Legendre expansion of ϕ_0 can be calculated explicitly, as in Lemma 4.1 of [55]. Based on this expansion, it is shown that $||\phi_0||^2_{L^2([0,1])} = 1/(p^2 + p)$. The degree p polynomial of minimal L^2 norm satisfying $\phi_0(0) = 0$ and $\phi_0(1) = 1$ is $\phi_0^-(x) = \phi_0(1 - x)$ and we have the following equality, given as Lemma 4.2 in [55]:

$$(\phi_0, \phi_0^-)_{L^2([0,1])} = \frac{(-1)^{p+1}}{2(p+1)}||\phi_0||^2_{L^2([0,1])}. \tag{4.20}$$

We now construct the vertex function associated with the vertex V_4. We start by defining it on the face F_3. To this end, we identify the face F_3 of Ω_{ref} with T. Let f be a continuous function on ∂T such that $f = \phi_0$ on I_1 and I_2, and $f = 0$ on I_3. We define $\Phi_{V^4}(x, y, 0) = U(x, y)$, where U is the extension of f given in Theorem 4.3 or 4.6.

We similarly define the vertex function on the faces F_2 and F_1 and set it to zero on the face F_4. Finally, we consider its discrete harmonic extension to the interior of Ω_{ref}.

Lemma 4.11 *If the function Φ_{V^4} is constructed by using the extension of Theorem 4.3, then*

$$||\Phi_{V^4}||_{H^1(\Omega_{ref})} \leq C(1 + \log p)||\phi_0||_{L^2(I_1)}. \tag{4.21}$$

If it is constructed by using the extension of Theorem 4.5, then

$$||\Phi_{V^4}||_{H^1(\Omega_{ref})} \leq CN(p)||\phi_0||_{L^2(I_1)}. \tag{4.22}$$

Proof. To fix the ideas let Φ_{V^4} be constructed by using the extension of Theorem 4.3. By a trace theorem, we have

$$||\Phi_{V^4}||_{H^1(\Omega_{ref})} \leq C||\Phi_{V^4}||_{H^{1/2}(\partial\Omega_{ref})}.$$

The $H^{1/2}$-norm can be broken into a sum of $H^{1/2}$-norms on faces and weighted L^2-norms that represent the $H^{1/2}$ -couplings between faces. The $H^{1/2}$-couplings between F_1, F_2, and F_3 are zero, for symmetry reasons. The $H^{1/2}$-coupling between F_3 and F_4 is the weighted L^2-norm in the expression of $||U||_{H^{1/2}_{00}(T,I_3)}$. The $H^{1/2}$-couplings between F_1 and F_4, and between F_2 and F_4, can be characterized similarly. The $H^{1/2}$-norms can be bounded from above as in Theorem 4.3. In order to prove the theorem, we need to bound the $H^{1/2}_{00}(T, I_3)$-norm by the $H^{1/2}$-norm. From Lemma 4.7, we have

$$||U||_{H^{1/2}_{00}(F_3,I_3)} \leq C(1 + \log p)||U||_{H^{1/2}(F_3)}.$$

and, consequently,

$$||\Phi_{V^4}||_{H^{1/2}(\partial\Omega_{ref})} \leq C(1 + \log p)||\phi_0||_{L^2(E_1)}.$$

□

We now construct the edge functions associated with the edge E_1. As before, we start by defining it on the face F_3, which we identify with T. Let f be any continuous function on ∂T such that $f = 0$ on I_2 and I_3. We define $\Phi^{[f]}_{E_1}(x, y, 0) = U(x, y)$, where U is the extension of f given in Theorem 4.3 or 4.6. We define it similarly on the face F_2 and set it to zero on the faces F_1 and F_4. Finally, we consider the discrete harmonic extension to the interior of Ω_{ref}. We remark that the edge functions are uniquely defined by their traces on the edge E_1. It follows that the dimension of the space of edge functions associated with one edge is $p-1$.

Lemma 4.12 *If the $\Phi_{E_1}^{[f]}$ are constructed by using the extension of Theorem 4.3, then*

$$||\Phi_{E_1}^{[f]}||_{H^1(\Omega_{ref})} \leq C(1+\log p)||f||_{L^2(W)}. \tag{4.23}$$

If they are constructed by using the extension of Theorem 4.6, then

$$||\Phi_{E_1}^{[f]}||_{H^1(\Omega_{ref})} \leq CN(p)||f||_{L^2(W)}. \tag{4.24}$$

Proof. Again, let $\Phi_{E_1}^{[f]}$ be constructed by using the extension of Theorem 4.3. By a trace theorem, we have

$$||\Phi_{E_1}^{[f]}||_{H^1(\Omega_{ref})} \leq C||\Phi^{[f]}||_{H^{1/2}(\partial\Omega_{ref})}$$

The $H^{1/2}$-norm can be broken into a sum of $H^{1/2}$-norms on faces and weighted L^2-norms that represent the $H^{1/2}$-couplings between faces.

The $H^{1/2}$-couplings between F_3 and F_2, and between F_1 and F_4 are zero, for symmetry reasons. The $H^{1/2}$-coupling between F_3 and F_1, F_3 and F_4, F_2 and F_1, F_2 and F_4, are the weighted L^2-norm in the expression of $||U||_{H_{00}^{1/2}(T,I_2,I_3)}$, given in Theorem 4.3. The $H^{1/2}$-norm can be bounded from above as in Theorem 4.3. It follows that

$$||\Phi_{E_1}^{[f]}||_{H^{1/2}(\partial(\Omega_{ref})} \leq C||f||_{L^2(W)}.$$

In order to prove the theorem, we need to bound the $H_{00}^{1/2}(T,I_3)$-norm by the $H^{1/2}$-norm. We have

$$||U||_{H_{00}^{1/2}(T,I_3)} \leq C(1+\log p)||U||_{H^{1/2}(T)},$$

and, consequently,

$$||\Phi_{E_1}^{[f]}||_{H^{1/2}(\partial(\Omega_{ref})} \leq C(1+\log p)||f||_{L^2(I_1)}.$$

□

4.3 The wire basket interpolant and the face functions

In this section, we use some ideas and results given in Pavarino and Widlund [55, 56]. We adopt a different approach only in those parts of the proofs where the extension theorems given in the previous sections are involved.

We denote the vertex, edge, and face components of $u \in \tilde{V}^p$ by $\tilde{u}_{V_i},\ i = 1 \dots 4$, $\tilde{u}_{E_j},\ j = 1 \dots 6$, and $\tilde{u}_{F_k},\ k = 1 \dots 4$, respectively.

We first define a preliminary interpolation operator $\tilde{I}^W : \tilde{V}^p \to \tilde{V}^p_W$, by

$$\tilde{I}^W u = \sum_{i=1}^{4} \tilde{u}_{V_i} + \sum_{j=1}^{6} \tilde{u}_{E_j}.$$

We use Theorem 4.3 or 4.6 to prove the following bound on the energy of this interpolation operator. We denote by $C(p)$ either $C(1 + \log p)^2$ or $CN(p)^2$.

Lemma 4.13

$$|\tilde{I}^W u|^2_{H^1(\Omega_{ref})} \le C(p)||u||^2_{L^2(W)}.$$

Proof. From Theorem 4.3 or 4.6, it follows that

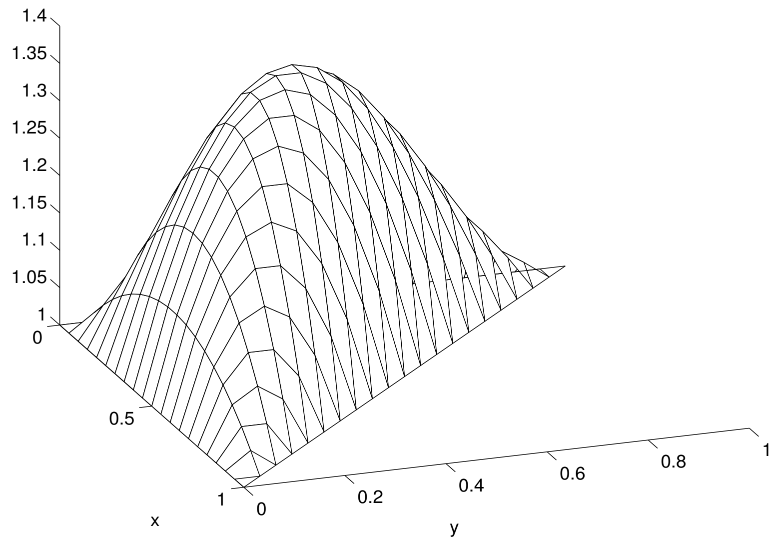
$$\begin{aligned}
|\tilde{I}^W u|^2_{H^1(\Omega_{ref})} &\le C(p)\left(\sum_{i=1}^{4} |\tilde{u}_{V_i}|^2_{H^1(\Omega_{ref})} + \sum_{j=1}^{6} |\tilde{u}_{E_j}|^2_{H^1(\Omega_{ref})}\right) \\
&\le C(p)\left(\sum_{i=1}^{4} ||\tilde{u}_{V_i}||^2_{L^2(W)} + \sum_{j=1}^{6} ||\tilde{u}_{E_j}||^2_{L^2(W)}\right) \\
&\le C(p)\Bigg(||\sum_{i=1}^{4} \tilde{u}_{V_i} + \sum_{j=1}^{6} \tilde{u}_{E_j}||^2_{L^2(W)} - 2\sum_{i_1,i_2} (\tilde{u}_{V_{i_1}}, \tilde{u}_{V_{i_2}})_{L^2(W)} \\
&- 2\sum_{j_1,k_1} (\tilde{u}_{V_{j_1}}, \tilde{u}_{E_{k_1}})_{L^2(W)} - 2\sum_{k_1,k_2} (\tilde{u}_{E_{k_1}}, \tilde{u}_{E_{k_2}})_{L^2(W)}\Bigg).
\end{aligned}$$

The L^2-scalar products on the last two lines vanish with the exception of $(\tilde{u}_{V_{j_1}}, \tilde{u}_{V_{j_2}})_{L^2(W)}$, which is negligible; cf. (4.20). Thus,

$$|\tilde{I}^W u|^2_{H^1(\Omega_{ref})} \le C(p)||\sum_{i=1}^{4} \tilde{u}_{V_i} + \sum_{j=1}^{6} \tilde{u}_{E_j}||^2_{L^2(W)} = C(p)||u||^2_{L^2(W)}.$$

$\square$

Figure 4.2: Extension of constant wire basket values



The range of this interpolation operator does not contain the constants, see Fig. 4.2. Our next goal is to define new vertex and edge function as well as a new interpolation operator which reproduces the constants.

We proceed as in [25, 55, 56]. We consider the expansion of the constant function 1 in the preliminary basis $\{\tilde{v}_{l,i} \;\; \tilde{e}_{l,j} \;\; \tilde{f}_{l,k}\}$. We define

$$\mathcal{F}_k = \sum_{l=1}^{p(p-1)/2} \alpha_{l,F_k} \tilde{f}_{l,k}; \tag{4.25}$$

and we define an interpolation operator I^W by

$$I^W u = \tilde{I}^W u + \sum_{k=1}^{4} \mathcal{F}_k \frac{\int_{\partial F_k} u}{\int_{\partial F_k} 1},$$

which maps a constant function into itself. The new vertex and edge basis functions $v_i,\ i = 1 \ldots 4$, and $e_j,\ j = 1 \ldots 6$, are the images of the preliminary ones under I^W. The face basis functions $f_k,\ k = 1 \ldots 4$, are the same as before. We denote the components of $u \in \tilde{V}^p$ in the new basis by $u_{V_i},\ i = 1 \ldots 4,\ u_{E_j},\ j = 1 \ldots 6$, and $u_{F_k},\ k = 1 \ldots 4$.

We now prove bounds on the energy of the face functions and the new I^W. The following is a stronger version of Lemma 4.7. It is an adaptation of Lemma 5.7 in Pavarino and Widlund [55] to our context.

Lemma 4.14 *Let $u \in P^p$ and let $\tilde{u}_{F_3}$ be its component associated with the face F_3 in the preliminary basis. Then,*

$$||\tilde{u}_{F_3}||^2_{H^{1/2}_{00}(F_3)} \le C(p)(1+\log p)^2 ||u||^2_{H^1(\Omega_{ref})}.$$

Proof. On the face F_3, $u = \tilde{u}_{F_3} + \tilde{u}_W$. By using Theorems 4.3 or 4.6.

$$|\tilde{u}_W|^2_{H^{1/2}(F_3)} \le C(p)||\tilde{u}_W||_{L^2(\partial F_3)} \le C(p)(1+\log p)||u||^2_{H^1(\Omega_{ref})}.$$

It follows:

$$|\tilde{u}_{F_3}|^2_{H^{1/2}(F_3)} \le C(p)(1+\log p)||u||^2_{H^1(\Omega_{ref})}.$$

The bound on the weighted L^2 norms of $\tilde{u}_{F_3}$ is obtained exactly like in Lemma 4.7, except the part where we use Lemma 4.5, which we now replace by Lemma 1.2.

$$\int_0^1 \frac{1}{y}\int_0^{1-y} \tilde{u}^2_{F_3} dxdy = \int_0^{1/p^2} \frac{1}{y}\int_0^{1-y} \tilde{u}^2_{F_3} dxdy + \int_{1/p^2}^1 \frac{1}{y}\int_0^{1-y} \tilde{u}^2_{F_3} dxdy. \tag{4.26}$$

The second integral is bounded by

$$\begin{aligned} & \int_{1/p^2}^1 \frac{1}{y}\int_0^{1-y} u^2 dxdy + \int_{1/p^2}^1 \frac{1}{y}\int_0^{1-y} \tilde{u}^2_W dxdy \\ \le\ & C(1+\log p)\Big(||u||^2_{H^1(\Omega_{ref})} + ||\tilde{u}_W||^2_{H^1(\Omega_{ref})}\Big) \\ \le\ & C(p)(1+\log p)||u||^2_{H^1(\Omega_{ref})}. \end{aligned}$$

We now bound the first integral in (4.26).

$$\begin{aligned} & \int_0^{1/p^2} \frac{1}{y}\int_0^{1-y} \tilde{u}_{F_3}(x,y)^2 dxdy \\ =\ & \int_0^{1/p^2} \frac{1}{y}\left(\int_0^{1-y} \tilde{u}_{F_3}(x,y)^2 dx - \int_0^1 \tilde{u}_{F_3} u(x,0)^2 dx\right) dy \\ =\ & \int_0^{1/p^2} y\frac{d}{dy}\int_0^{1-\tilde{y}} \tilde{u}_{F_3}(x,\tilde{y})^2 dxdy \\ \le\ & Cp^2 \max_y \left(\int_0^{1-y} \tilde{u}_{F_3}(x,y)^2 dx\right) \int_0^{1/p^2} dy \\ \le\ & C \max_y \left(\int_0^{1-y} \tilde{u}_{F_3}(x,y)^2 dx\right). \end{aligned}$$

The last member of the sequence of inequalities can be bounded as above. □

We provide now bounds on the energy of the special face functions and, consequently, on the energy of the corrected interpolant I^W.

Lemma 4.15

$$||\mathcal{F}_k||^2_{H^{1/2}_{00}(F_3)} \leq C(p)(1+\log p) \tag{4.27}$$

$$\tag{4.28}$$

Proof. This is an immediate consequence of Lemma 4.4. □

Lemma 4.16

$$|I^W u|^2_{H^1(\Omega_{ref})} \leq C(p)(1+\log p)||u||^2_{L^2(W)}. \tag{4.29}$$

Lemma 4.17 *Let $u \in P^p$ and let u_{F_3} be its component associated with the face F_3 in the corrected basis. Then,*

$$||u_{F_3}||^2_{H^{1/2}_{00}(F_3)} \leq C(p)(1+\log p)^2|u|^2_{H^1(\Omega_{ref})}.$$

Proof. We use Lemmas 4.14 and 4.16. The H^1-norm can be replaced by the H^1-seminorm because, by adding a constant to u, the face component u_{F_3} does not change. □

4.4 Theoretical bounds on the condition number

We now prove upper bounds on the condition numbers of the relevant operators for the algorithms defined in Chapter 3. We follow the proofs of analogous theorems given in Dryja, Smith, and Widlund [25], Pavarino and Widlund [55, 56], and Smith, Bjørstad, and Gropp[62]. The proofs consist in putting together all the technical results of Sections 4.1- 4.3 with the abstract Schwarz theory of Chapter 2.

Theorem 4.7 *For the additive wire basket-based algorithm defined in Section 3.1,*

$$\kappa(T) \leq C(p)(1 + \log p)^2.$$

The constant in the bound is independent of the jumps of the coefficient ρ_i in (1.11). The global bound is the largest of the local bounds.

Proof. We will find an upper as well as a lower bound for the eigenvalues of T. To fix the ideas, we consider the case of an inexact solver on the wire basket. It is enough to prove the local lower and upper bound

$$\hat{a}^{(j)}(u_W, u_W) + \sum_k a^{(j)}(u_{F_k}, u_{F_k}) \leq C(1 + \log p)^2 a^{(j)}(u, u) \quad \forall u \in \tilde{V}^p, \tag{4.30}$$

and

$$a^{(j)}(u, u) \leq C\hat{a}_W^{(j)}(u_W, u_W) + \sum_k a^{(j)}(u_{F_k}, u_{F_k}) \quad \forall u \in \tilde{V}^p, \tag{4.31}$$

respectively. Here $a^{(j)}(u, u)$ and $\hat{a}_W^{(j)}(u, u)$ are the contributions from the substructure Ω_j to the bilinear forms $a(u, u)$ and $\hat{a}_W(u, u)$, respectively. The global bounds are the same and are obtained by adding over the substructures. We prove the bounds on the reference tetrahedron Ω_{ref}.

(i) The lower bound. From Lemmas 4.4 and 4.17, we have

$$(1 + \log p)||u - \bar{u}_W||^2_{L^2(W)} + \sum_{k=1}^{4} |u_{F_k}|^2_{H^1(\Omega_{ref})} \leq C(p)(1 + \log p)^2 |u|^2_{H^1(\Omega_{ref})} \tag{4.32}$$

(ii) The upper bound. From Lemmas 4.15 and 4.13, we have

$$|I^W u|^2_{H^1(\Omega_{ref})} \quad \leq \quad 5\Big(|\tilde{I}^W u|^2_{H^1(\Omega_{ref})} + \sum_{k=1}^{4} \bar{u}^2_{\partial F_k} |\mathcal{F}_k|^2_{H^1(\Omega_{ref})}\Big)$$

$$\begin{aligned} &\leq C(p)\Big(||u||^2_{L^2(W)} + \sum_{k=1}^{4} \bar{u}^2_{\partial F_k}(1+\log p)\Big) \\ &\leq C(p)(1+\log p)||u||^2_{L^2(W)}. \end{aligned}$$

We shift u by a constant such that $\bar{u}_W = 0$. Then

$$\begin{aligned} a(u,u) &= \Big|I^W u + \sum_{k=1}^{4} u_{F_k}\Big|^2_{H^1(\Omega_{ref})} \\ &\leq C(p)\Big((1+\log p)||u||^2_{L^2(W)} + \sum_{k=1}^{4} |u_{F_k}|^2_{H^1(\Omega_{ref})}\Big), \end{aligned}$$

which proves (4.31). □

Theorem 4.8 *For the additive vertex-based algorithm defined in Section 3.2,*

$$\kappa(T) \leq C(p)(1+\log p)^2.$$

The constant in the bound might depend on the jumps of the coefficient ρ_i in (1.11).

Proof. The difficult part is to obtain the lower bound. In order to do so, we prove that every $u \in V$ can be decomposed as $u = u_0 + \sum_i u_v^i + \sum_j u_e^j + \sum_k u_f^k$, such that:

$$\begin{aligned} & a(u_0,u_0) + \sum_i a(u_{V_i},u_{V_i}) + \sum_j a(u_{E_j},u_{E_j}) + \sum_k a(u_{F_k},u_{F_k}) \\ \leq\ & C(p)(1+\log p)^2 a(u,u). \end{aligned} \tag{4.33}$$

Let $u_0 = Qu$, where Q is the L^2-projector on the space of piecewise linear functions defined on the triangulation of Ω. Let $w = u - u_0$ By using Lemma 4.8, the problem reduces to showing that

$$\sum_i a(w_{V_i},w_{V_i}) + \sum_j a(w_{E_j},w_{E_j}) + \sum_k a(w_{F_k},w_{F_k}) \leq C(1+\log p)^2||w||_{H^1(\Omega)}. \tag{4.34}$$

From now on, everything follows exactly as in Theorem 4.7. There are two differences: in Theorem 4.7, we use the second inequality in Lemma 4.4, but now it is the first inequality of the same lemma that is needed. We also use Lemma 4.14 instead of Lemma 4.17 since the wire basket space may not contain the constants. Lemma 4.14 is sufficient for our purposes, since we need a bound in terms of $||w||_{H^1(\Omega)}$, not in terms of $|w|_{H^1(\Omega)}$. □

The following theorem is given in Dryja and Widlund [29, Theorem 3], for the $h-$version. See also [62, Theorem 4, Section 5.3.3]. As in the previous two proofs, the proof for the $h-$version carries over to our case by making the appropriate changes, so we do not give it here.

Theorem 4.9 *For the additive Neumann-Neumann algorithm defined in Section 3.4,*

$$\kappa(T) \leq C(p)\frac{(1+\log p)^2}{H^2}.$$

The constant in the bound is independent of the jumps of the coefficient ρ_i in (1.11).

The factor H^2 appears because there is no coarse space. If we add a coarse component to the preconditioner, it can be dropped; see [29], [62], [54].

Chapter 5

Numerical experiments

In this chapter, we report on some numerical experiments with the methods introduced in Chapter 3. Our goal is to obtain numerical results that support the theory given in Chapter 4 and to get a sense of the size of the constants that appear in the theoretical bounds. We describe numerical experiments on a single tetrahedron as well as on a polyhedral region triangulated with relatively many tetrahedra.

5.1 Local experiments

In this section, we compute the local condition number of the preconditioned Schur complement, on the reference tetrahedron. We compare the condition numbers obtained in the case of low energy basis functions versus standard basis functions. The computation of the local condition numbers is particularly important for the wire basket-based algorithm, since the global condition number is bounded from above by the largest of the local condition numbers; see Theorem 4.7. We also look at how strong the couplings between particular elements are, with the goal of duplicating, in overlapping methods, those vertex or edge degrees of freedom that have the strongest couplings with the face degrees of freedom.

Next, we look at the bounds on these condition numbers, given by the theory. To this end, we compute all the constants in the inequalities used in the proof of the main results. The asymptotic logarithmic growth of these bounds is more visible than that of the actual condition numbers.

Finally, we present numerical experiments that support the conjectures related to

Lemma 4.3 and Lemma 4.6.

5.1.1 Local condition numbers

In Tables 5.1 and 5.2 we give the local condition numbers of the preconditioned local stiffness matrix for the wire basket algorithm, described in Section 3.1. We make the experiments for several choices of basis functions, described below.
(I) The low energy basis functions constructed in Subsection 4.2.3, by using the extensions (4.14-4.16), and the standard vertex functions (1.13-1.19).
(II) The basis obtained from (I) by adding the corrections (3.5).
(III) The basis obtained from (I) by applying the partial orthogonalization (3.2 - 3.3).
(IV) The basis obtained from (III) by adding the corrections (3.5).
If we start with the standard vertex functions, the initial wire basket space already contains the constants, so the bases (I) and (II) coincide, so we do not show column (II) in Table 5.2.

The Schur complement S is singular. If the preconditioner S_{prec} is nonsingular, then we define λ_{min} as the smallest eigenvalue of $S_{prec}^{-1}S$. If S_{prec} is singular, then, by elementary linear algebra, the two matrices have a common null eigenspace. We then define λ_{min} as the smallest positive eigenvalue, on the complement of the null space, of the generalized eigenvalue problem $Sx = \lambda S_{prec}x$. We already mentioned, in Section 3.1 and Theorem 4.7 that, for the wire basket algorithm, the global λ_{min} is the smallest of the local λ_{min}.

We remark that the condition numbers obtained by using low energy vertex and edge functions are substantially smaller than those obtained by using standard vertex and edge functions.

By comparing the numbers in columns I and III of Table 5.1, we conclude that the partial orthogonalization process, described in Section 3.1, reduces the condition numbers by at about a half. Things do not change much after we apply the corrections (3.5); the numbers in column IV are smaller than those in column II.

Turning now to the vertex-based algorithm, we can draw similar conclusions by studying Table 5.3. We show only columns I and III since, in this case, the wire basket space does not have to contain the constants, so we do not have to apply the corrections (3.5).

We next consider the couplings between the wire basket and the faces, and between

Table 5.1: Local condition numbers, wire basket algorithm, low energy vertex and edge functions

p	I			II		
	λ_{min}	λ_{max}	κ	λ_{min}	λ_{max}	κ
4	0.1166	2.4715	21.1886	0.1012	2.4403	24.1153
5	0.0852	2.5726	30.1823	0.0874	2.5347	28.9883
6	0.0730	2.6060	35.6864	0.0727	2.5650	35.2916
7	0.0581	2.6248	45.1407	0.0581	2.6248	45.1407
8	0.0511	2.6486	51.8690	0.0511	2.6486	51.8690
9	0.0443	2.6584	59.9418	0.0443	2.6584	59.9418
10	0.0407	2.6731	65.7369	0.0419	2.6334	62.8346
p	III			IV		
	λ_{min}	λ_{max}	κ	λ_{min}	λ_{max}	κ
4	0.1921	1.8000	9.3691	0.1331	2.2549	16.9416
5	0.1358	1.7788	13.1022	0.1063	2.3890	22.4775
6	0.1033	1.8203	17.6186	0.0842	2.4503	29.1136
7	0.0864	1.8205	21.0818	0.0753	2.4996	33.2026
8	0.0740	1.8407	24.8854	0.0655	2.5374	38.7335
9	0.0656	1.8476	28.1508	0.0601	2.5668	42.6989
10	0.0590	1.8588	31.4892	0.0541	2.5911	47.9346

the faces, to determine which decoupling causes more ill-conditioning. Table 5.4 contains the extreme eigenvalues of

$$\begin{pmatrix} S_{WW} & 0 \\ 0 & S_{FF} \end{pmatrix}^{-1} S;$$

and

$$\begin{pmatrix} S_{F_1F_1} & 0 \\ 0 & \ddots \end{pmatrix}^{-1} S_{FF},$$

respectively.

We remark that the condition numbers in the columns I, II, and IV of the upper part of Table 5.4, are larger than those in the lower part of the same table. This conclusion is important because we can only reduce the former, by using an appropriate extension in the construction of our vertex and edge functions. The latter depend solely on the geometry.

Table 5.2: Local condition numbers, wire basket algorithm, standard vertex and edge functions

p	I					
	λ_{min}	λ_{max}	κ			
4	0.0398	2.4723	62.0469			
5	0.0262	2.5643	98.0488			
6	0.0188	2.6010	1.3843e+02			
7	0.0140	2.6104	1.8589e+02			
8	0.0109	2.6265	2.4006e+02			
9	0.0088	2.6300	2.9996e+02			
10	0.0072	2.6398	3.6723e+02			
p	III			IV		
	λ_{min}	λ_{max}	κ	λ_{min}	λ_{max}	κ
4	0.0805	1.7716	22.0113	0.0128	2.2540	1.7582e+02
5	0.0389	1.8752	48.1564	0.0064	2.4210	3.7760e+02
6	0.0208	1.8516	89.0675	0.0040	2.4902	6.2770e+02
7	0.0134	1.9020	1.4184e+02	0.0026	2.5479	9.8516e+02
8	0.0090	1.8811	2.0937e+02	0.0018	2.5743	1.3964e+03
9	0.0066	1.9069	2.9056e+02	0.0014	2.6057	1.9262e+03
10	0.0049	1.892	3.8639e+02	0.0011	2.6324	2.3931e+03

Table 5.3: Local condition numbers, vertex-based algorithm

Low energy vertex and edge functions						
p	I			III		
	λ_{min}	λ_{max}	κ	λ_{min}	λ_{max}	κ
4	0.0472	3.9954	84.5857	0.0667	2.3817	35.6813
5	0.0349	4.0631	1.1629e+02	0.0511	2.1795	42.6378
6	0.0307	4.0788	1.3299e+02	0.0501	2.3263	46.4560
7	0.0262	4.0903	1.5642e+02	0.0427	2.1290	49.8143
8	0.0242	4.0951	1.6914e+02	0.0428	2.2920	53.5686
9	0.0219	4.0985	1.8742e+02	0.0383	2.0970	54.7887
10	0.0207	4.1002	1.9797e+02	0.0384	2.2594	58.8362

Standard vertex and edge functions						
p	I			III		
	λ_{min}	λ_{max}	κ	λ_{min}	λ_{max}	κ
4	0.0127	5.5692	4.3682e+02	0.0232	2.6511	114.2572
5	0.0076	5.6015	7.3695e+02	0.0149	2.4447	1.6403e+02
6	0.0052	5.6087	1.0845e+03	0.0107	2.4542	2.2985e+02
7	0.0037	5.6117	1.5102e+03	0.0078	2.3479	3.0292e+02
8	0.0028	5.6129	1.9862e+03	0.0060	2.3590	3.9296e+02
9	0.0022	5.6135	2.5361e+03	0.0046	2.2960	4.9399e+02
10	0.0018	5.6138	3.1380e+03	0.0036	2.3048	6.3943e+02

Table 5.4: Local condition numbers, etc.

Decoupling the wire basket from the faces						
p	I			II		
	λ_{min}	λ_{max}	κ	λ_{min}	λ_{max}	κ
4	0.1047	2.0000	19.1089	0.1236	1.8764	15.1827
5	0.0803	2.0000	24.9046	0.1007	1.8993	18.8605
6	0.0786	2.0000	25.4463	0.1005	1.8995	18.8953
7	0.0664	2.0000	30.1218	0.0869	1.9131	22.0084
8	0.0657	2.0000	30.4505	0.0852	1.9148	22.4781
9	0.0584	2.0000	34.2378	0.0774	1.9226	24.8429
10	0.0579	2.0000	34.5572	0.0730	1.9270	26.3947
p	III			IV		
	λ_{min}	λ_{max}	κ	λ_{min}	λ_{max}	κ
4	0.2554	2.0000	7.8316	0.1627	1.8373	11.2962
5	0.2553	2.0000	7.8336	0.1437	1.8563	12.9221
6	0.2022	2.0000	9.8890	0.1127	1.8873	16.7443
7	0.1962	2.0000	10.1916	0.1013	1.8987	18.7483
8	0.1762	2.0000	11.3515	0.0892	1.9108	21.4119
9	0.1716	2.0000	11.6579	0.0824	1.9176	23.2672
10	0.1611	2.0000	12.4151	0.0761	1.9239	25.2860

Decoupling the faces from each other			
p	λ_{min}	λ_{max}	κ
4	0.2099	1.6109	7.6746
5	0.1899	1.6906	8.9026
6	0.1326	1.7533	13.2225
7	0.1272	1.7792	13.9874
8	0.1026	1.8041	17.5838
9	0.0996	1.8205	18.2781
10	0.0855	1.8343	21.4538

5.1.2 Bounds on the local condition numbers

We now compute the bounds on the local condition number, as given in Theorem 4.7. We consider the case of an exact solver on the wire basket and the vertex and edge functions (I). The smallest eigenvalue is given by

$$\lambda_{\min}^{-1} = \max_u \frac{a(u_w, u_w) + \sum a(u_{f^k}, u_{f^k})}{a(u,u)},$$

but in Theorem 4.7, we provide a bound for

$$C_0^2 = \max_u \frac{a(u_w, u_w)}{a(u,u)} + \max_u \frac{\sum a(u_{f^k}, u_{f^k})}{a(u,u)}. \tag{5.1}$$

In a first step, we compute the lower bound given above and collect them in Table 5.5. We remark that the lower bound is slightly smaller than the actual smallest eigenvalue, given in column I of Table 5.1. The numbers in the first column correspond to the $C(1+\log p)^2$ factor of Lemma 4.7. We compute this factor by considering the generalized eigenvalue problems

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & S_{F_1F_1} & 0 \\ 0 & 0 & \ddots \end{pmatrix} x = \lambda S x$$

and

$$\begin{pmatrix} S_{WW} & 0 \\ 0 & 0 \end{pmatrix} x = \lambda S x,$$

respectively.

The number of zero eigenvalues is equal to the number of wire basket or face functions, respectively. There is also an infinite eigenvalue. Our factor, $C(1+\log p)^2$, corresponds to the largest finite eigenvalue of the generalized eigenvalue problem.

However, the numbers in the third column of Table 5.5 are not yet the lower bounds provided by Theorem 4.7 because the upper bound on $\max_u a(u_w, u_w)/a(u,u)$ is not computed directly, but in two steps. The upper bound on $\max_u a(u_w, u_w)/a(u,u)$ is the product of the two factors also given below in Tables 5.6 and 5.7. Their product is given in Table 5.8. Fig. 5.1 shows how the condition numbers and their upper bounds grow with p.

The numbers in Table 5.6 correspond to the smallest number $C(p)$ such that

$$||u||^2_{H^1(K)} \leq C(p) ||u||^2_{L^2(W)} \quad \forall u \in \tilde{V}_W.$$

Table 5.5: Lower bound on the smallest positive eigenvalue (1)

p	$\max_u \sum a(u_{f^k}, u_{f^k})/a(u,u)$	$\max_u a(u_w, u_w)/a(u,u)$	C_0^{-2}
4	4.3461	5.0410	0.1065
5	9.0413	6.4866	0.0644
6	11.3375	6.6218	0.0557
7	14.8239	7.7890	0.0442
8	16.9317	7.8711	0.0403
9	19.8529	8.8170	0.0349
10	21.7060	8.8968	0.0327

Table 5.6: Norm of the extension $L^2(W) \rightarrow H^1(\Omega_{ref})$.

Low energy vertex and edge functions				
p	I	I I	III	IV
4	4.9293	4.9056	4.4232	4.4350
5	5.4829	5.4594	4.8942	4.9110
6	5.9562	5.9472	5.3827	5.3896
7	6.3320	6.3231	5.6860	5.6951
8	6.6874	6.6840	6.0260	6.0297
9	6.9591	6.9548	6.2373	6.2430
10	7.2396	7.2381	6.4914	6.4936

This factor can be viewed as the norm of an extension operator

$$E : L^2(W) \rightarrow H^1(\Omega_{ref}).$$

We have seen in Section 4.2 that the norm of this extension operator is bounded from above by $C(1 + \log p)^2$ if the extension is based on given in Theorem 4.3, and $CN(p)$ if based on Theorem 4.6. We conjectured that, in both cases, the norms are actually constant, independent of p. The experiments described in Table 5.6, support this conjecture in the second case. The factor $C(p)$ is the largest eigenvalue of the generalized eigenvalue problem

$$S_{WW}x = \lambda Mx.$$

The numbers in Table 5.7 correspond to the discrete Sobolev factor $C(1+\log p)$ that appear in Theorem 4.4. In order to compute this factor, we consider the generalized

Table 5.7: Best discrete Sobolev factors

p	$C(1+\log p)$
4	16.8552
5	18.2909
6	19.3894
7	20.3045
8	21.0353
9	21.6849
10	22.2217

Table 5.8: Lower bound on the smallest eigenvalue (2)

p	I	I I	III	IV
4	0.0114	0.0115	0.0127	0.0126
5	0.0091	0.0092	0.0101	0.0101
6	0.0079	0.0079	0.0086	0.0086
7	0.0070	0.0070	0.0077	0.0077
8	0.0063	0.0063	0.0070	0.0070
9	0.0059	0.0059	0.0064	0.0064
10	0.0055	0.0055	0.0060	0.0060

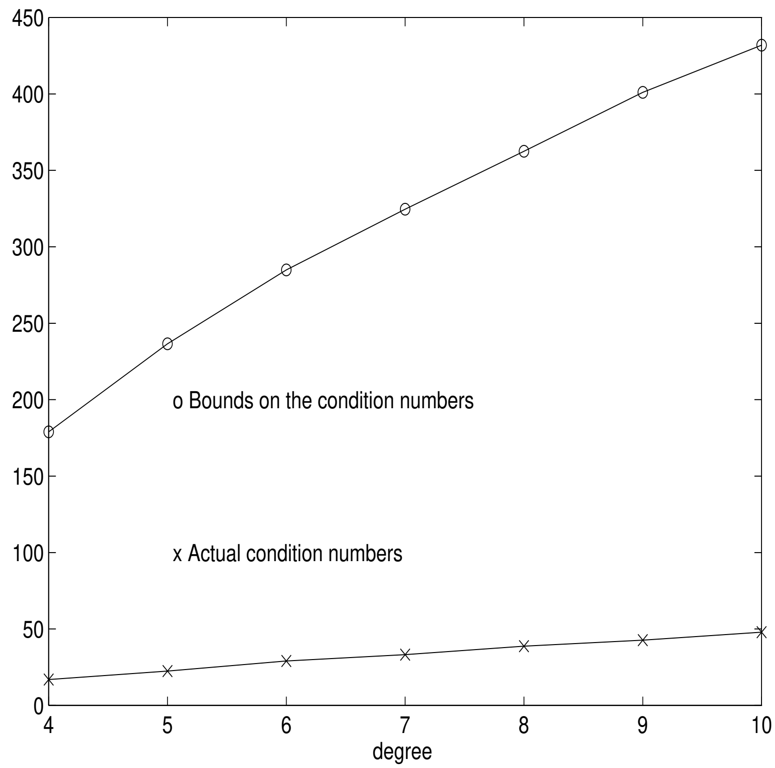
eigenvalue problem

$$\begin{pmatrix} M & 0 \\ 0 & 0 \end{pmatrix} x = \lambda S x.$$

where the null diagonal block correspond to the degrees of freedom associated with the faces of the reference tetrahedron. There is a number of zero eigenvalues, which is equal to the number of face functions. There is also an infinite eigenvalue. Our factor, $C(1+\log p)$, is largest finite eigenvalue of the generalized eigenvalue problem.

The factors in Table 5.6, which depend on the extension used to construct the vertex and edge functions, are much smaller than those in Table 5.7, which are independent of our construction of the extension operator.

Figure 5.1: Condition numbers and bounds on them, for the wire basket algorithm



5.2 Global experiments

In this section, we report on some numerical experiments with the methods introduced in Chapter 3. We solve the Poisson problem with mixed homogeneous boundary conditions, on regions that consists of many tetrahedral substructures.

As in the previous section, we obtain numerical results that support the theory given in Chapter 4 and we also get a sense of the size of the constants that appear in the theoretical bounds. We already have seen, in the case of one tetrahedron, see Fig 5.1, that the actual condition numbers may be much smaller than their upper bounds.

We look at the number of iterations required to reduce the error by a fixed factor as well as the condition number of the preconditioned system, which is directly related to the number of iterations. We do not consider the computing time since it depends heavily on the efficiency of the implementation. Our goal has been to develop an experimental code that allow us to compare the condition numbers of different algorithms; we have not had the time to do an efficient or parallel implementation.

In order to simplify the experiments, we have considered cubic regions Ω that consist of $N \times N \times N$ cubes, each cube being divided into 24 tetrahedra, as in Fig. 1.6. We set homogeneous Dirichlet boundary conditions on one face, and Neumann on the other five faces of the region.

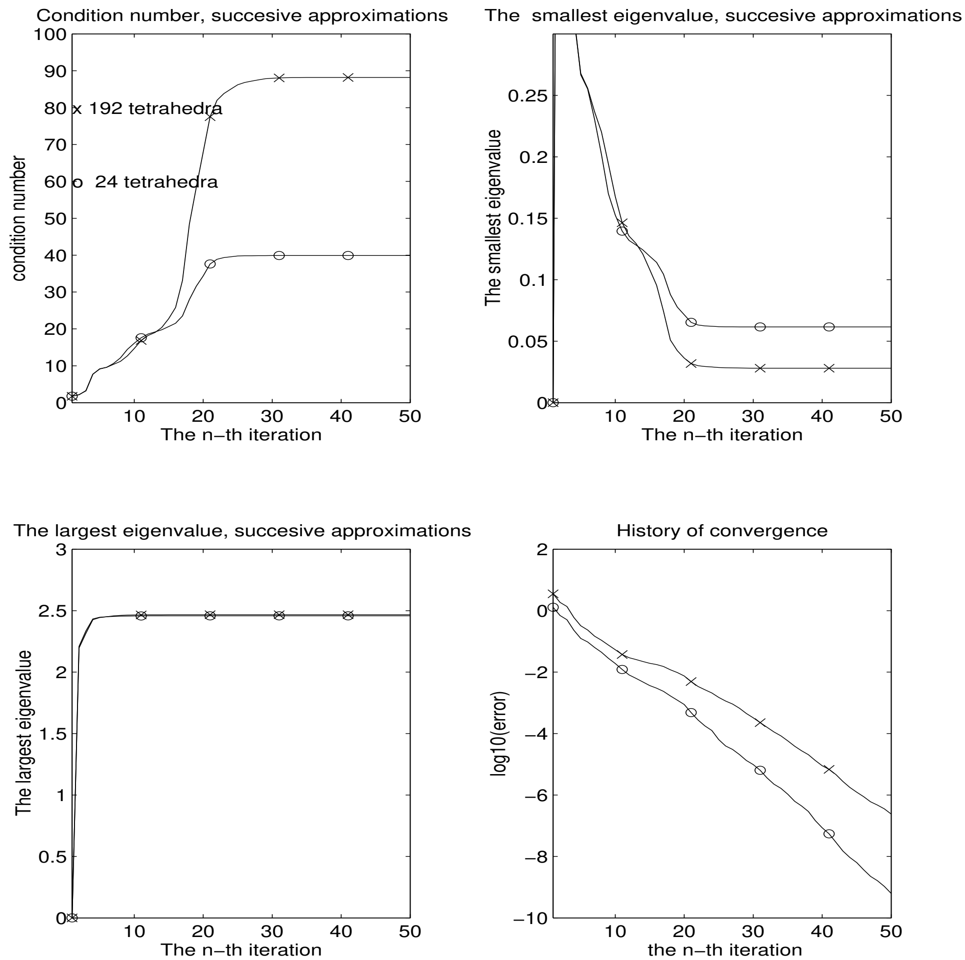
The system $Sx = b$ is solved by using the preconditioned conjugate gradient method. The number of iterations is fixed beforehand. We provide the exact solution x_{exact}, which is a random vector and compute the right hand side b. We compute the $a(\cdot, \cdot)$-norm of the error $x - x_{exact}$, where x is the solution computed by our iterative method.

The programs have been written in MATLAB and have been run on Sun workstations.

5.2.1 Coarse space

We have already mentioned the need of a coarse space that contains the constants. We now report on experiments for two regions, consisting of 24 and 192 tetrahedra, respectively, and study the condition number κ as well as the number of iterations required to reduce the norm of the error by a fixed factor, 10^{-5}. If there is no coarse space that contain the constants, the iteration count and condition number depend on the number of subregions, otherwise they do not. Fig. 5.2 contains the results of two

Figure 5.2: Coarse space that does not contain the constant functions, $p = 4$.



experiments, on regions made of $1 \times 1 \times 1 \times 24 = 24$, and $2 \times 2 \times 2 \times 24 = 192$ tetrahedra, respectively. We have used low energy vertex and edge functions (I), constructed by using the extensions (4.14-4.16) with the wire basket space as the coarse space, using the exact solver. The smallest eigenvalue and, consequently, the condition number as well as the rate of convergence deteriorate as the number of substructures increases. The largest eigenvalue remains the same, since it depends only on the local geometry; see (2.4).

Performing a similar experiment using the vertex and edge functions (III), for which the wire basket space contains the constants, the smallest eigenvalue, condition number, and rate of convergence are virtually the same in both cases; see Fig. 5.3.

Figure 5.3: Coarse space that contains the constant functions, $p = 4$.

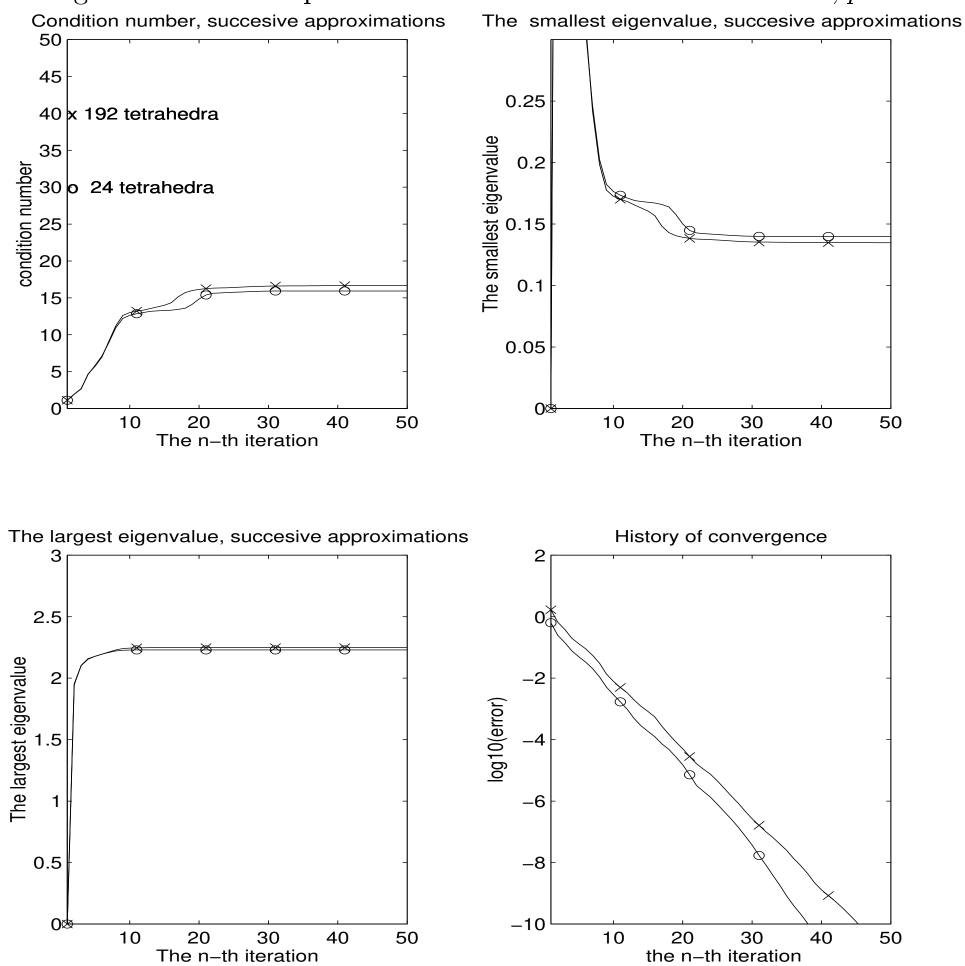


Figure 5.4: Comparison of basis functions, wire basket algorithm, $p = 6$.

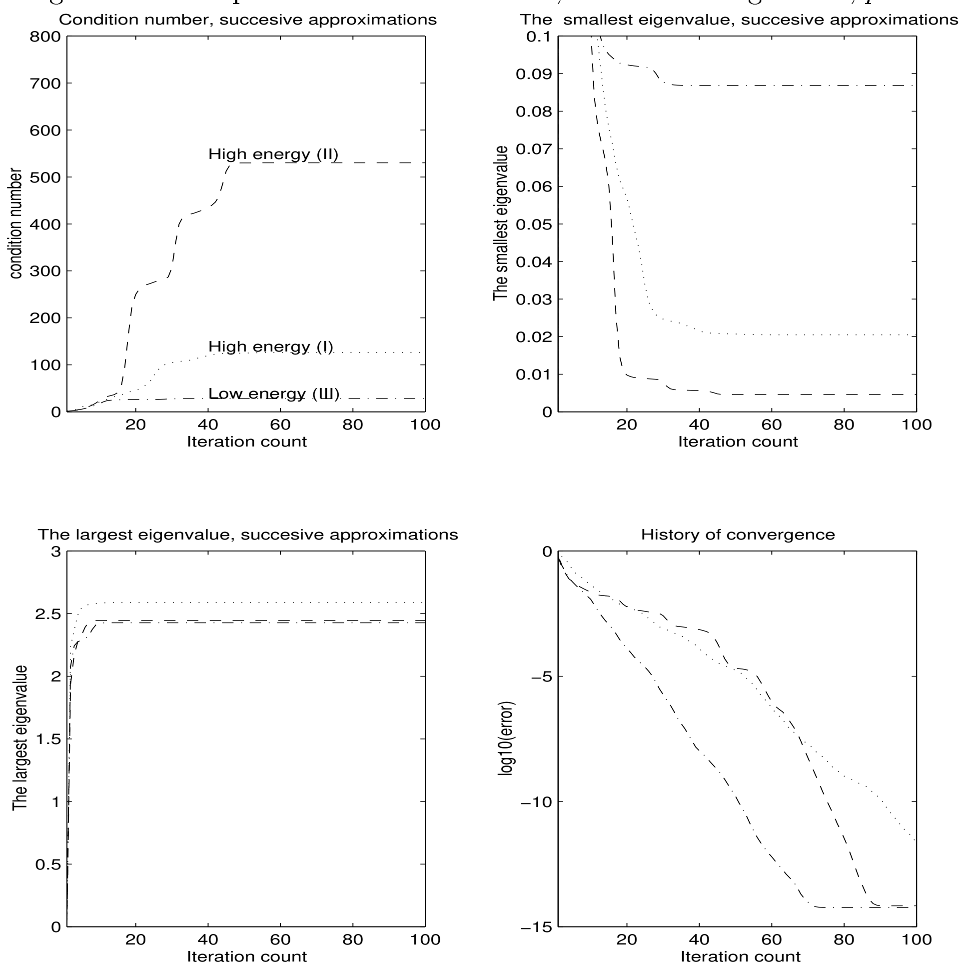


Figure 5.5: Comparison of basis functions, vertex based algorithm, $p = 6$.

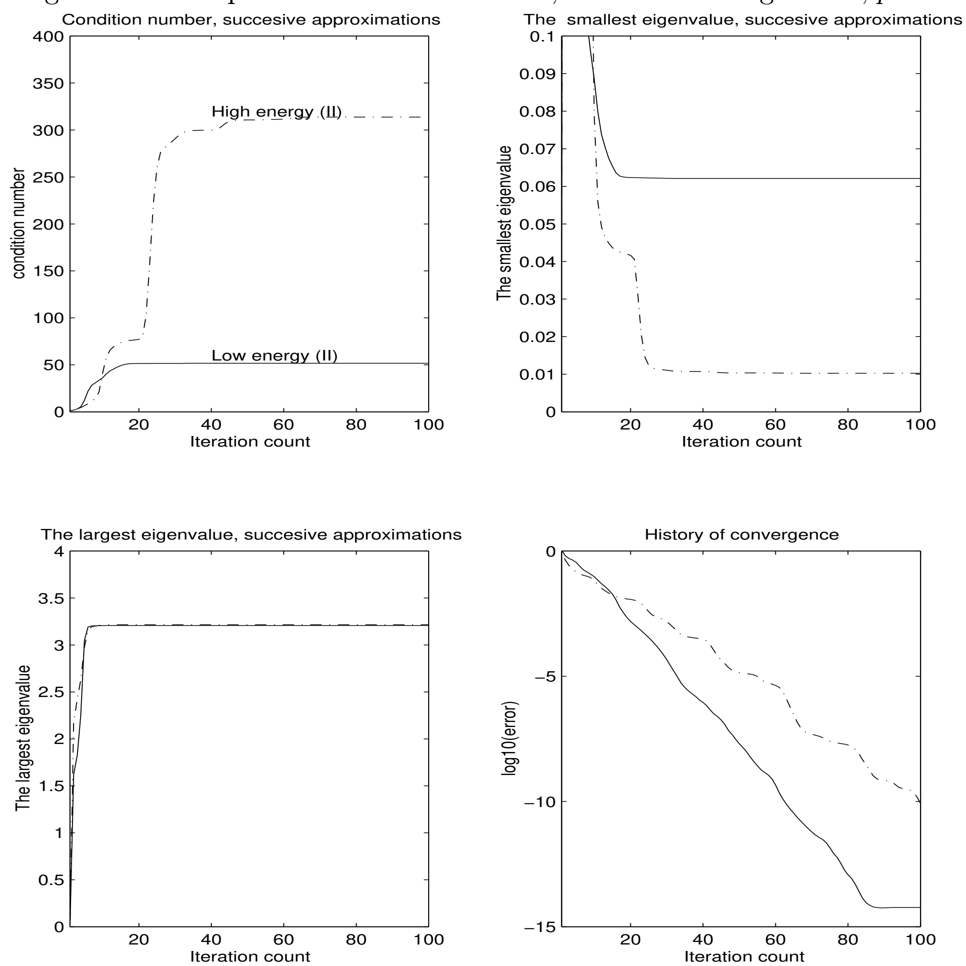


Table 5.9: Condition numbers and iteration count, the wire basket algorithm.

p	Exact solver				Inexact solver			
	λ_{min}	λ_{max}	κ	# iter	λ_{min}	λ_{max}	κ	# iter
4	0.1400	2.2291	15.9260	20	0.0566	4.2247	74.6729	37
5	0.1100	2.3658	21.5138	24	0.0470	4.8037	102.2064	41
6	0.0868	2.4268	27.9439	26	0.0434	5.2117	119.9645	44
7	0.0790	2.4743	31.3309	28	0.0411	5.5890	135.8435	47
8	0.0667	2.5152	37.6905	30	0.0368	5.8769	159.5147	50
9	0.0607	2.5452	41.9467	31	0.0342	6.1498	179.7665	56
10	0.0546	2.5699	47.0280	32	0.0327	6.3605	194.4376	59

5.2.2 Basis functions

In this section, we compare the condition number and number of iterations, to achieve a desired tolerance, using low energy and standard basis functions. We have already seen, in Section 5.1 that, by using low energy vertex and edge functions, we obtain condition numbers several times smaller than those obtained by using standard vertex and edge functions. We now consider the case of a cubic region made of 24 tetrahedra, with homogeneous Dirichlet boundary conditions on one face.

In Fig. 5.4, we give results for the wire basket algorithm. Since the wire basket space should contain the constants, we use the low energy functions (III) and standard basis functions (I) and (III).

In Fig. 5.5, we give results for the vertex-based algorithm. We use the vertex and edge functions (II) since they have the lowest energy among (I), (II), and (III) in both the low and high energy basis function case.

We conclude that the convergence of the algorithm deteriorates when using standard vertex and edge functions.

5.2.3 Additive, hybrid, and multiplicative methods

In this section, we start by considering the additive variant of the wire basket algorithm, with an exact and inexact solver on the wire basket. We study the condition number and number of iterations required to reduce the error by 10^{-5}. We collect the results in Table 5.9.

Even if the condition numbers corresponding to an inexact solver on the wire basket are much larger than the ones corresponding to the exact solver, the convergence does not deteriorate very much; it takes less than twice as many iterations to reduce the error by the same factor. We have already mentioned, in Chapter 3 that the algorithm of practical interest is the one that uses an inexact solver.

We next compare the performances of additive, hybrid, and multiplicative methods, for the wire basket and vertex based algorithms. We recall that the additive Schwarz method (ASM) is defined by the operator

$$T = T_W + T_{F_1} + \cdots T_{F_{n_F}}$$

for the wire basket based algorithm, and by

$$T = T_0 + T_{V_1} + \cdots T_{V_{n_V}} + T_{E_1} + \cdots T_{E_{n_E}} + T_{F_1} + \cdots T_{F_{n_F}}$$

for the vertex-based algorithm, where n_V, n_E, n_F is the number of vertices, edges, and faces, respectively.

The hybrid method HSM1 is given by the operator

$$T = I - (I - T_W)(I - T_F)(I - T_W)$$

for the wire basket based algorithm, and by

$$T = I - (I - T_0)(I - T_V)(I - T_E)(I - T_F)(I - T_E)(I - T_V)(I - T_0)$$

for the vertex-based algorithm. Here, $T_V = T_{V_1} + \cdots T_{V_{n_V}}$, $T_E = T_{E_1} + \cdots T_{E_{n_E}}$, and $T_F = T_{F_1} + \cdots T_{F_{n_F}}$.

The hybrid method HSM2 is defined by the operator

$$T = I - (I - T_F)(I - T_W)(I - T_F),$$

for the wire basket based algorithm, and by

$$T = I - (I - T_F)(I - T_E)(I - T_V)(I - T_0)(I - T_V)(I - T_E)(I - T_F),$$

for the vertex-based algorithm.

The symmetrized multiplicative method MSM1 is defined by the operator

$$T = I - (I - T_W)(I - T_{F_1}) \cdots (I - T_{F_{n_F}}) \cdots (I - T_{F_1})(I - T_W),$$

for the wire basket based algorithm, and by

$$\begin{aligned} T &= I - (I - T_0)(I - T_{V_1}) \cdots (I - T_{V_{n_V}})(I - T_{E_1}) \cdots (I - T_{E_{n_E}}) \\ &\quad (I - T_{F_1}) \cdots (I - T_{F_{n_F}}) \cdots (I - T_{F_1}) \\ &\quad (I - T_{E_{n_E}}) \cdots (I - T_{E_1})(I - T_{V_{n_V}}) \cdots (I - T_{V_1})(I - T_0). \end{aligned}$$

for the vertex-based algorithm.

The symmetrized multiplicative method MSM2 is defined by the operator

$$T = I - (I - T_{F_1}) \ldots (I - T_{F_M})(I - T_W) \ldots (I - T_{F_1}),$$

for the wire basket based algorithm, and by

$$\begin{aligned} T &= I - (I - T_{F_1}) \cdots (I - T_{F_{n_F}})(I - T_{E_1}) \cdots (I - T_{E_{n_E}}) \\ &\quad (I - T_{V_1}) \cdots (I - T_{V_{n_V}})(I - T_0)(I - T_{V_{n_V}}) \cdots (I - T_{V_1}) \\ &\quad (I - T_{E_{n_E}}) \cdots (I - T_{E_1})(I - T_{F_{n_F}}) \cdots (I - T_{F_1}). \end{aligned}$$

for the vertex-based algorithm.

In Fig. 5.6, we consider the wire basket based algorithm, using the exact solver on the wire basket. For the experiments of Fig. 5.7, we use the inexact solver defined in Section 3.1, obtained by multiplying the mass matrix on the wire basket by $(1 + \log p)$. Finally, we scale the inexact solver by a different factor, such that $\omega = 1$; see Chapter 2.2. The results are given in Fig. 5.8.

The results of similar experiments with the vertex-based algorithm are given in Fig. 5.9, for the basis functions (II).

As expected, the hybrid and multiplicative algorithms perform better than the additive. Also, MSM1 performs better than HSM1 and MSM2. On the other hand, MSM2 does not perform any better than HSM2.

Figure 5.6: Comparison of additive, hybrid, and multiplicative methods, $p = 6$, exact solver on the wire basket.

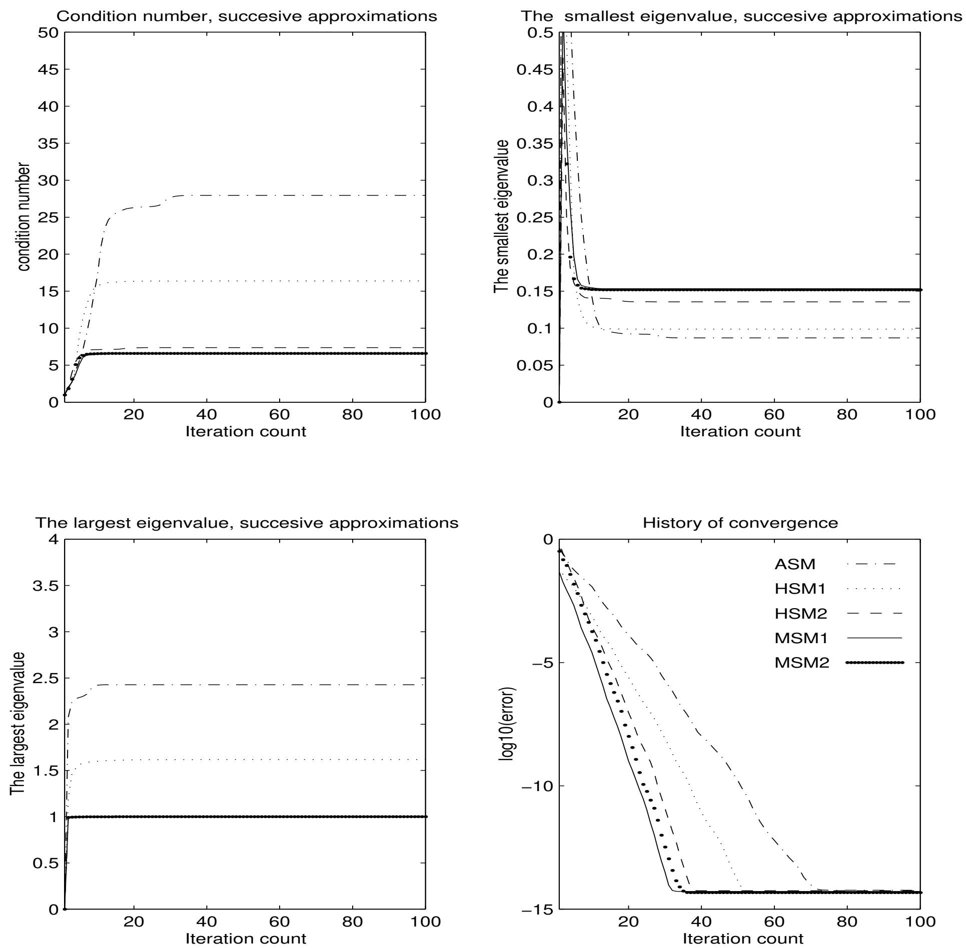


Figure 5.7: Comparison of additive, hybrid, and multiplicative methods, $p = 6$, inexact solver on the wire basket, with $(1 + \log p)$ scaling

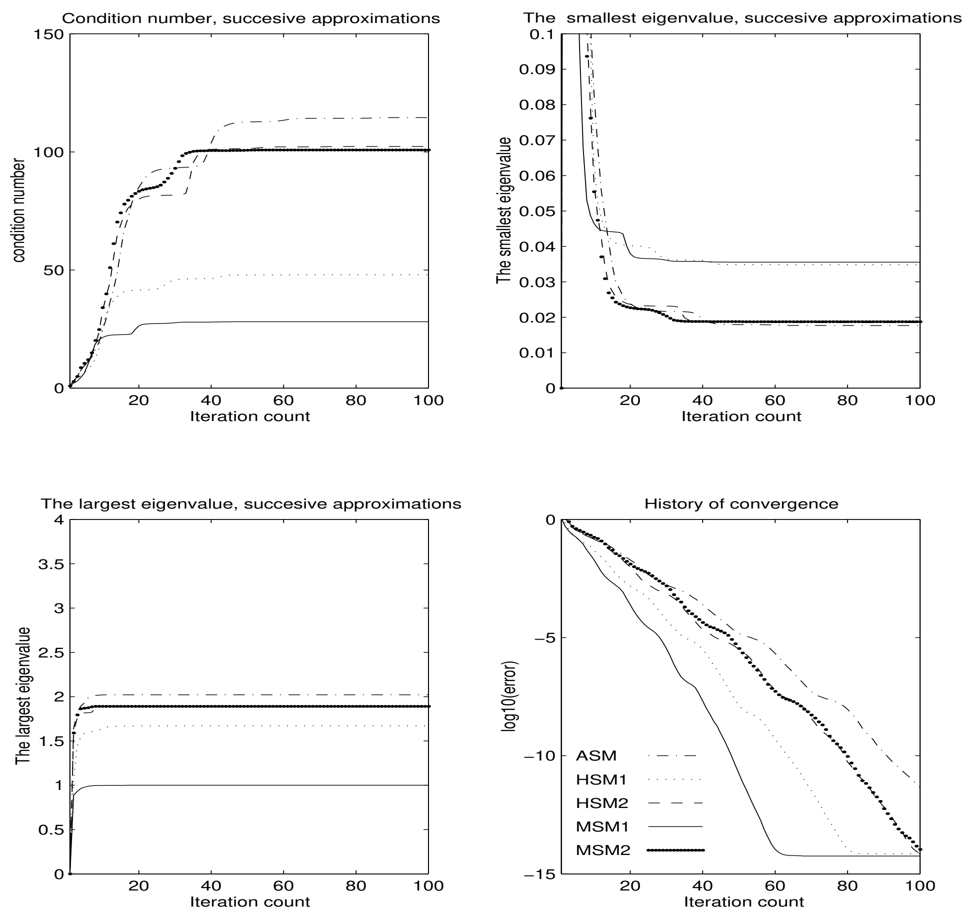


Figure 5.8: Comparison of additive, hybrid, and multiplicative methods, $p = 6$, inexact solver on the wire basket, with scaling such that $\omega = 1$.

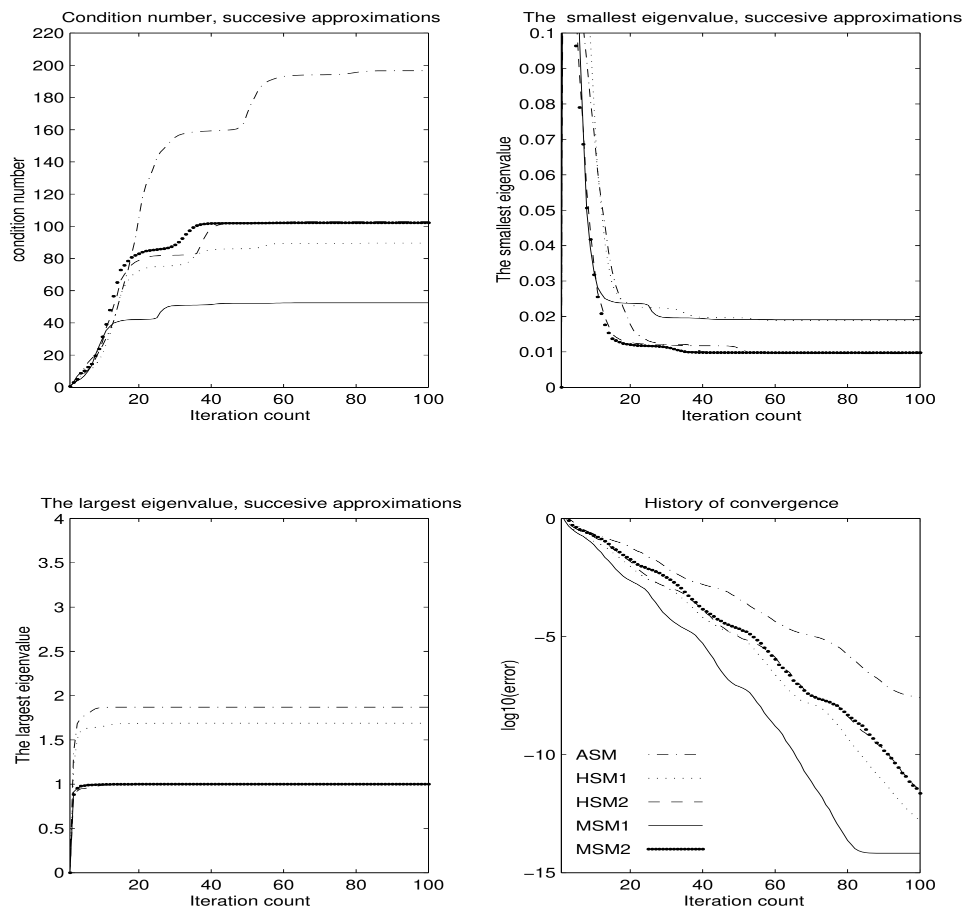
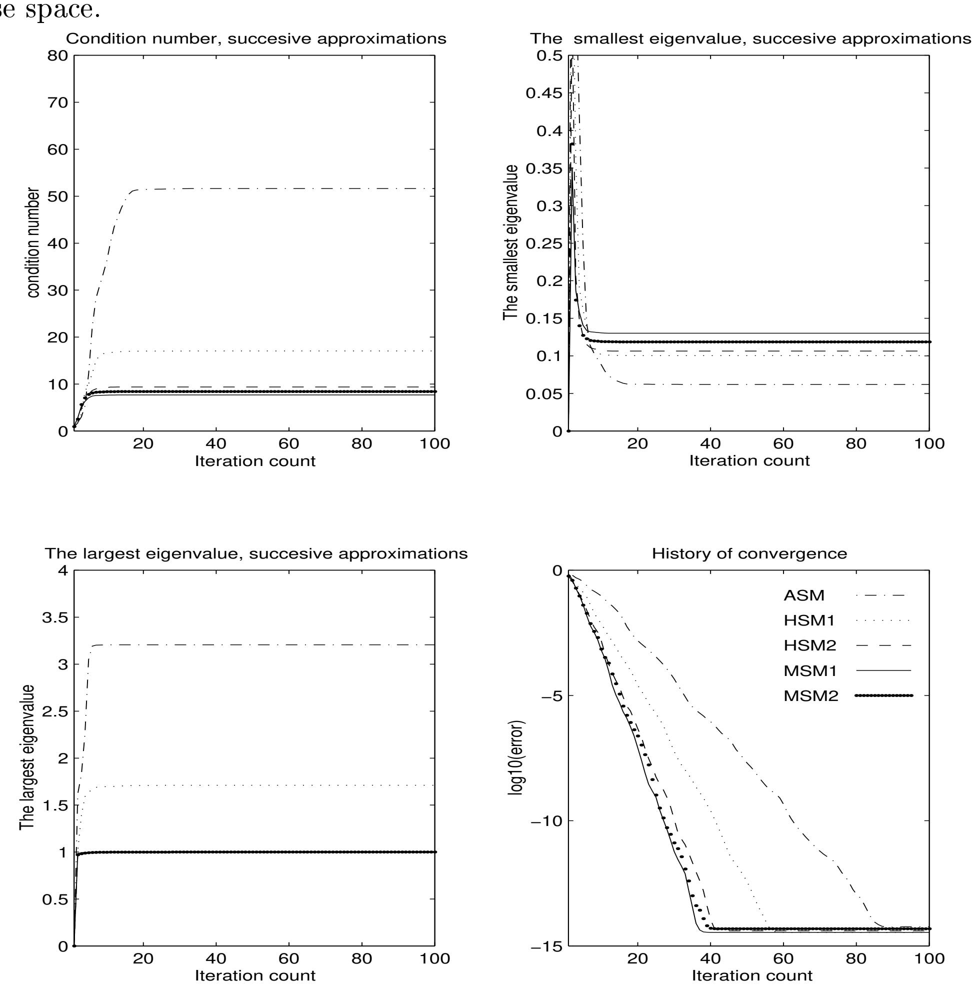


Figure 5.9: Comparison of additive, hybrid, and multiplicative methods, $p = 6$, standard coarse space.



5.2.4 Algorithms using overlap

In this section, we describe experiments with algorithms using overlap, as described in Section 3.3. To each face space, we add the entire edge subspaces that correspond to edges that are coupled to the face and are (1) or are not (2) part of the boundary of the face. An edge is coupled to a face if they belong to the same tetrahedron. The overlap is very generous but similar results can be obtained if, instead of adding the entire edge space to the face space, we add a subspace spanned only by those degrees of freedom that are coupled the strongest to the face, since they are the main cause of the ill-conditioning. Unfortunately, we did not find a regular pattern, like, e.g., that the first and the second edge functions are coupled stronger to the face than the other edge functions. One idea would be to scan the local Schur complements and determine which these degrees of freedom are in each case.

As we already noticed in Section 3.3, in order to reduce the condition number, we must obtain a lower bound which is large enough to compensate for the increase of the upper bound. This turns out to be unsuccessful for all the additive methods. Actually, the condition number and rate of convergence deteriorate. We do not see an improvement for the wire basket algorithm with the exact solver on the wire basket for either the additive, hybrid, or multiplicative methods. However, the improvement is clear for multiplicative variants of the wire basket based algorithm with inexact solver on the wire basket, and for vertex-based algorithms. In some cases, the hybrid methods break down; we do not show those results in the figures. There is no contradiction here, since we do not have an abstract Schwarz theory for the hybrid methods.

5.2.5 Neumann-Neumann algorithms

We describe experiments that correspond to several ways of treating the singularity of the subproblems as well as several choices of the coarse space, as described in Section 3.4.

Fig. 5.20 contains results of experiments for zero Dirichlet boundary conditions on the entire wire baskets. We use the exact solver on the wire basket as well as the inexact solver with the natural scaling $(1+\log p)$, indicated by "inexact solver (1)" in the picture, and the inexact solver scaled such that $\omega = 1$, indicated by "inexact solver (2)".

Figure 5.10: Wire basket algorithm with overlap (1), exact solver on the wire basket, $p = 6$

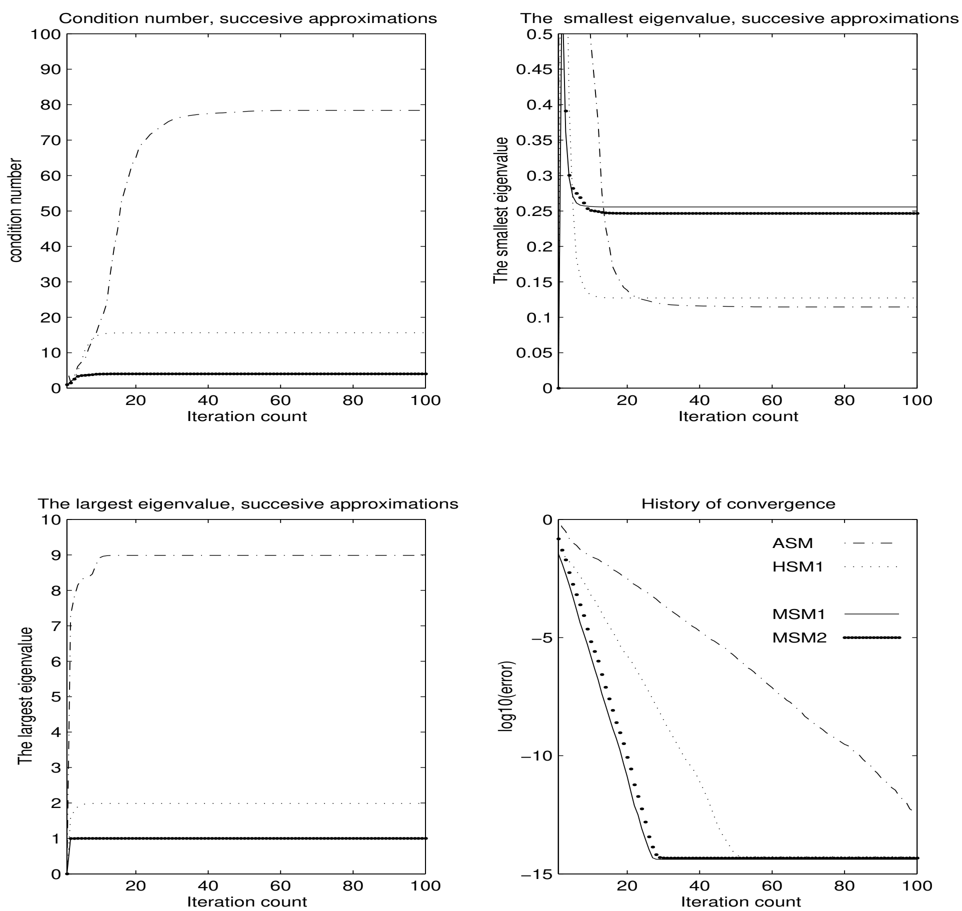


Figure 5.11: Wire basket algorithm with overlap (2); exact solver on the wire basket; $p = 6$

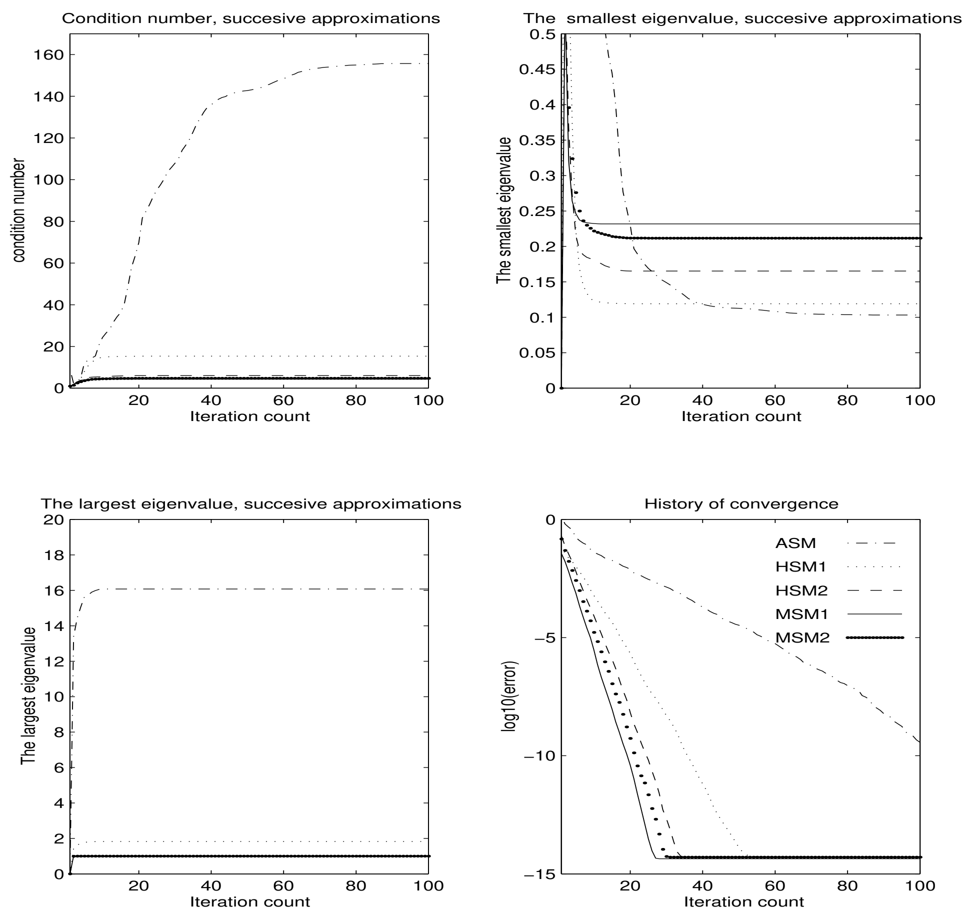


Figure 5.12: Wire basket algorithm with overlap (1); inexact solver on the wire basket; $(1 + \log p)$ scaling; $p = 6$

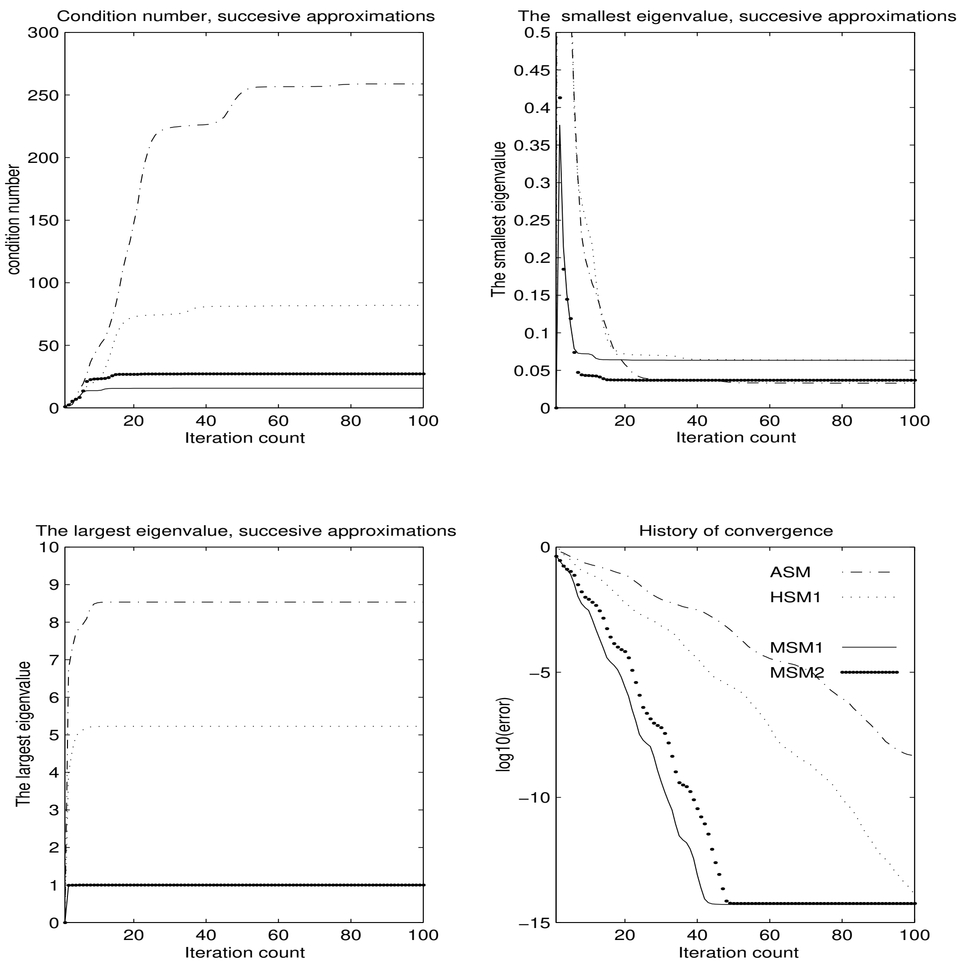


Figure 5.13: Wire basket algorithm with overlap (2); inexact solver on the wire basket; $(1 + \log p)$ scaling; $p = 6$

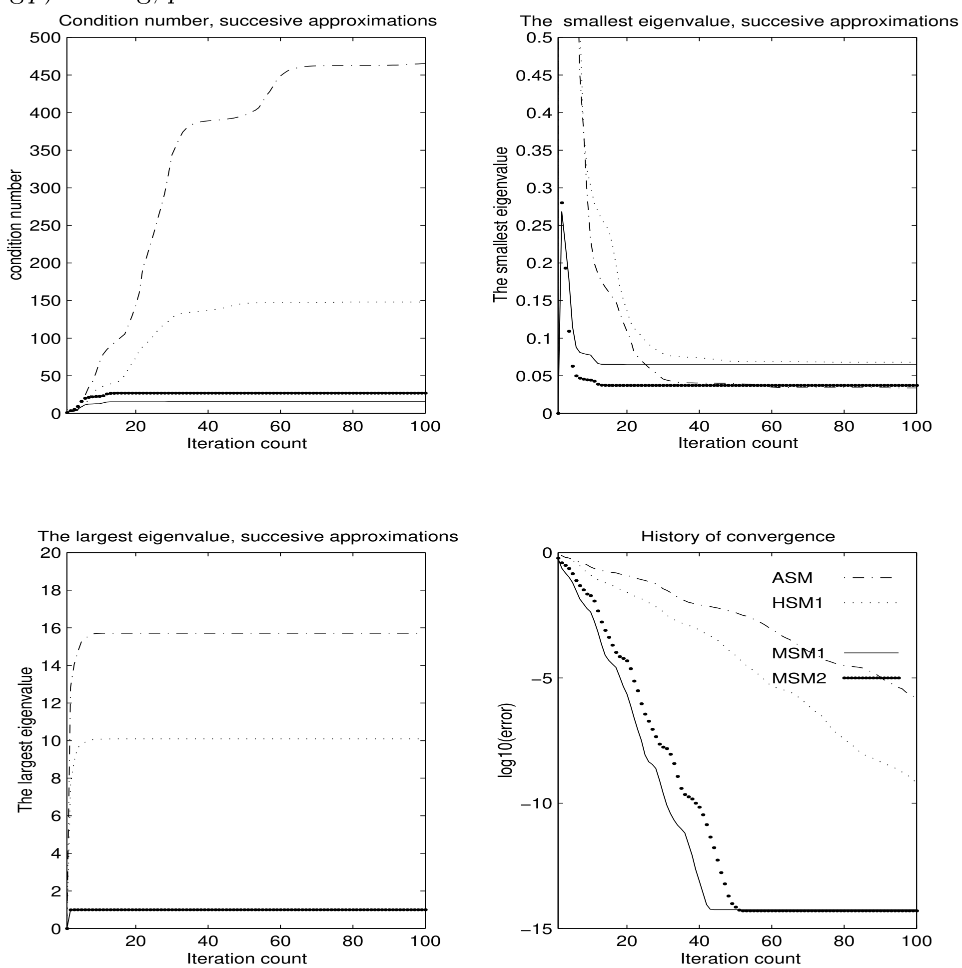


Figure 5.14: Wire basket algorithm with overlap (1); inexact solver on the wire basket; scaling such that $\omega = 1$; $p = 6$

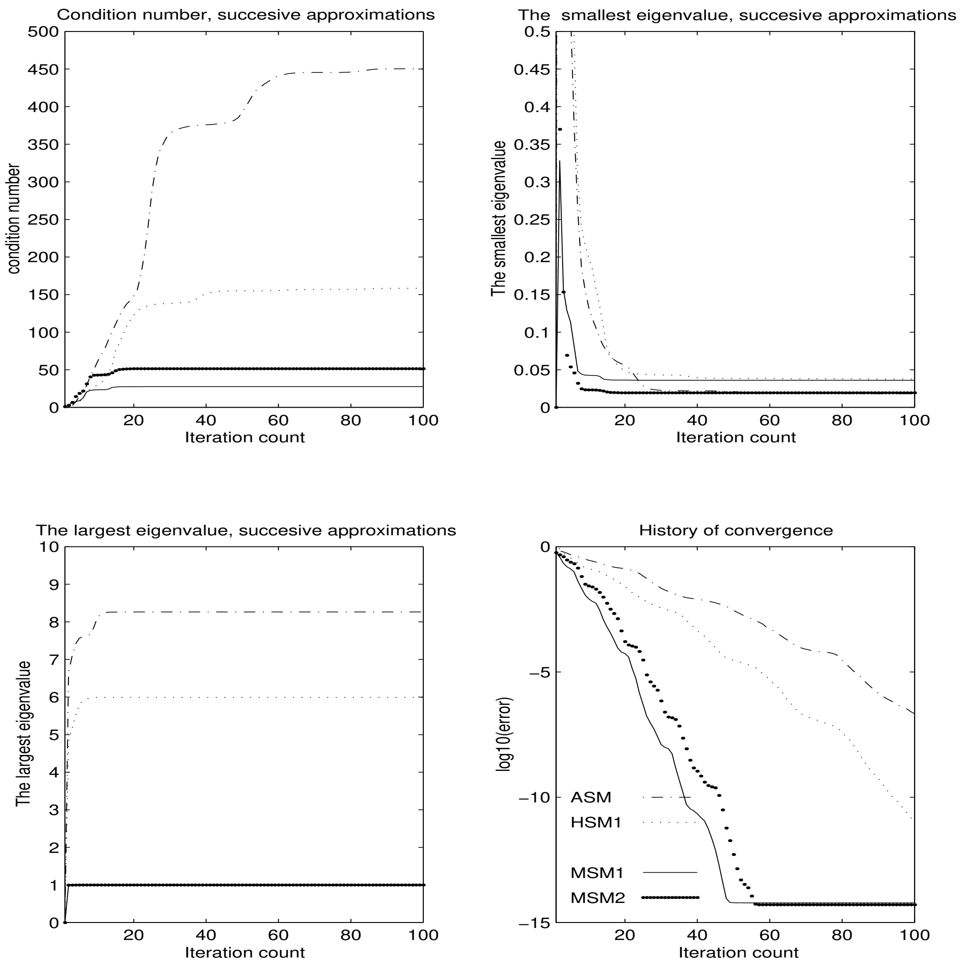


Figure 5.15: Wire basket algorithm with overlap (2); inexact solver on the wire basket; scaling such that $\omega = 1$; $p = 6$

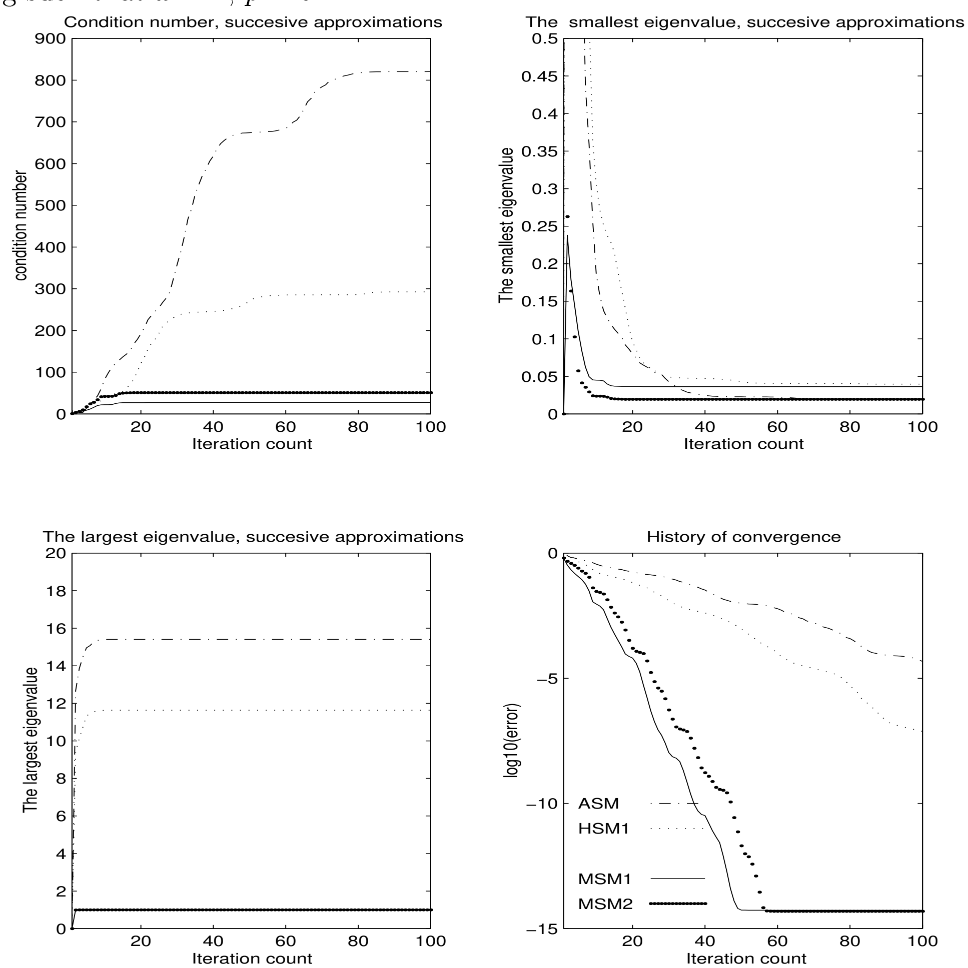


Figure 5.16: Vertex-based algorithm (I) with overlap (1); $p = 6$

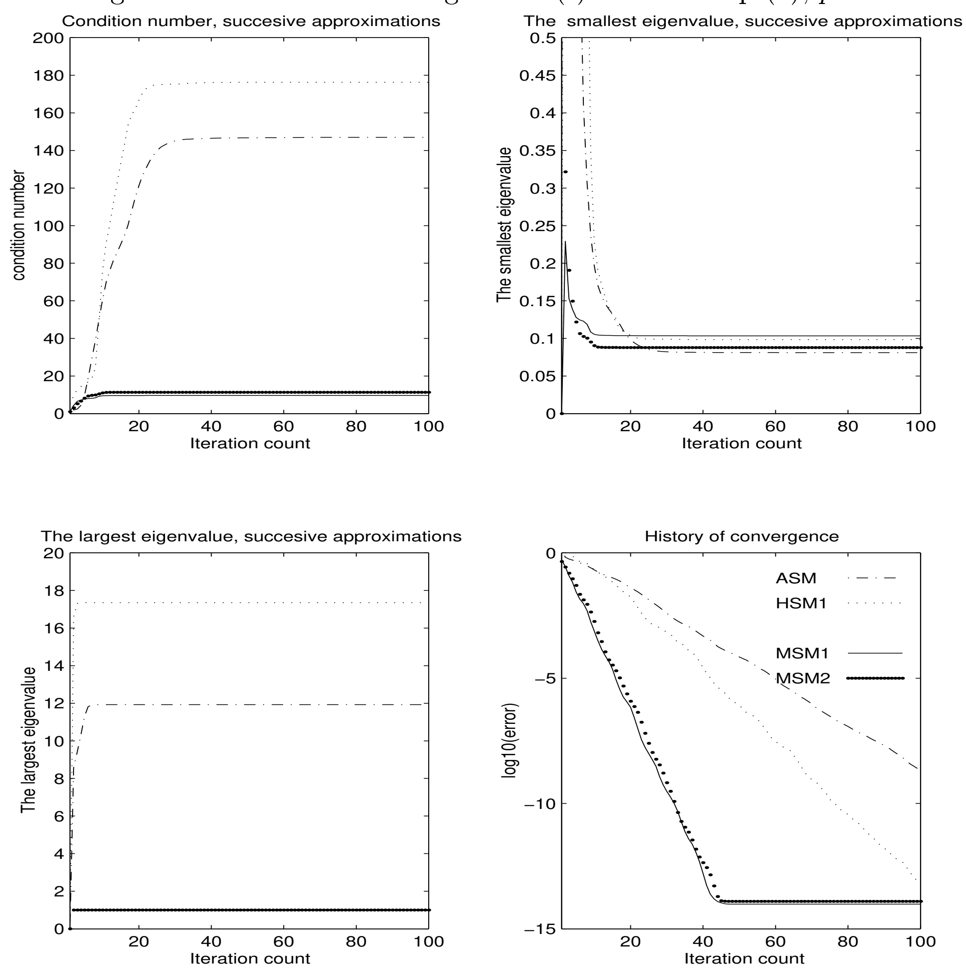


Figure 5.17: Vertex-based algorithm (II) with overlap (1); $p = 6$

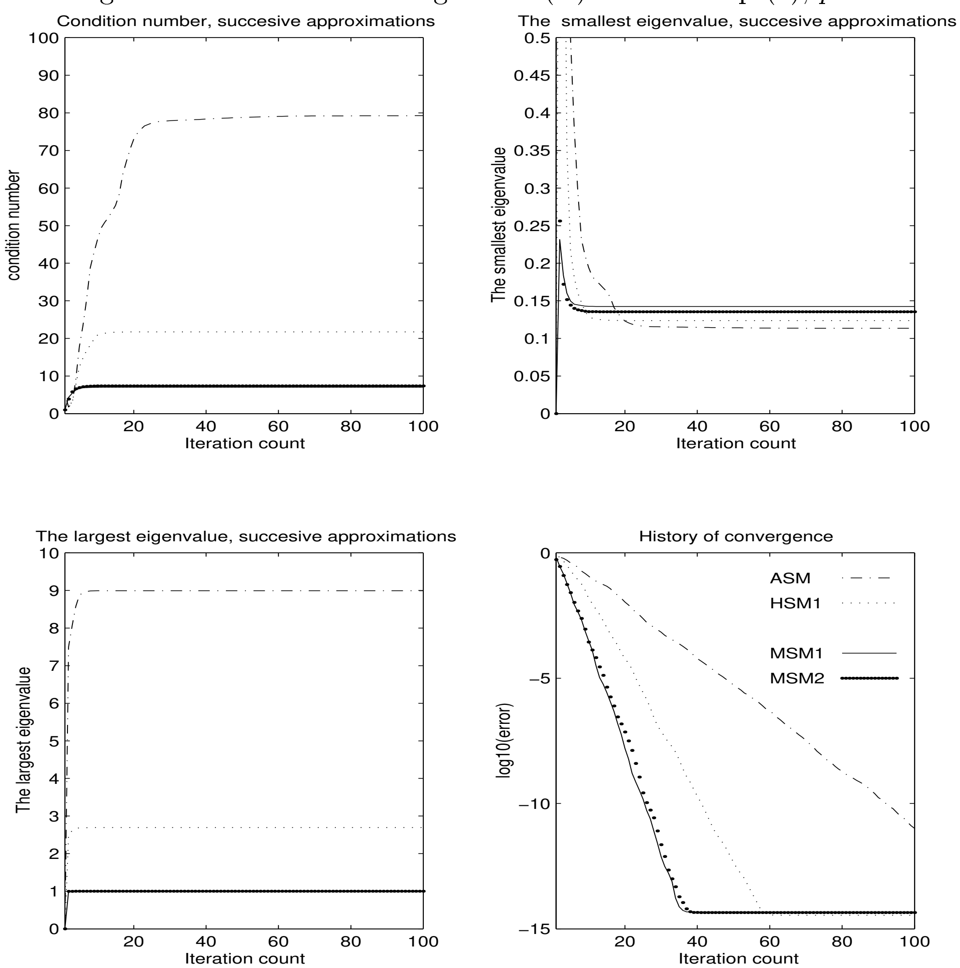


Figure 5.18: Vertex-based algorithm (I) with overlap (2); $p = 6$

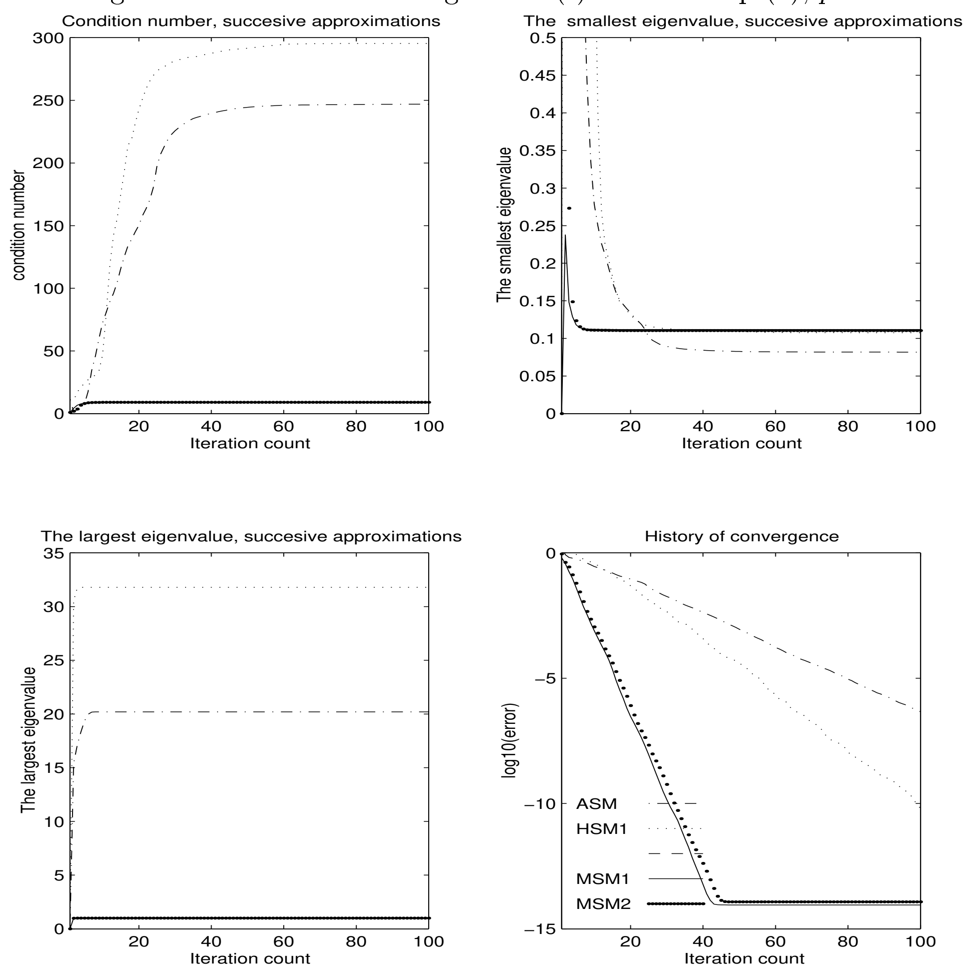
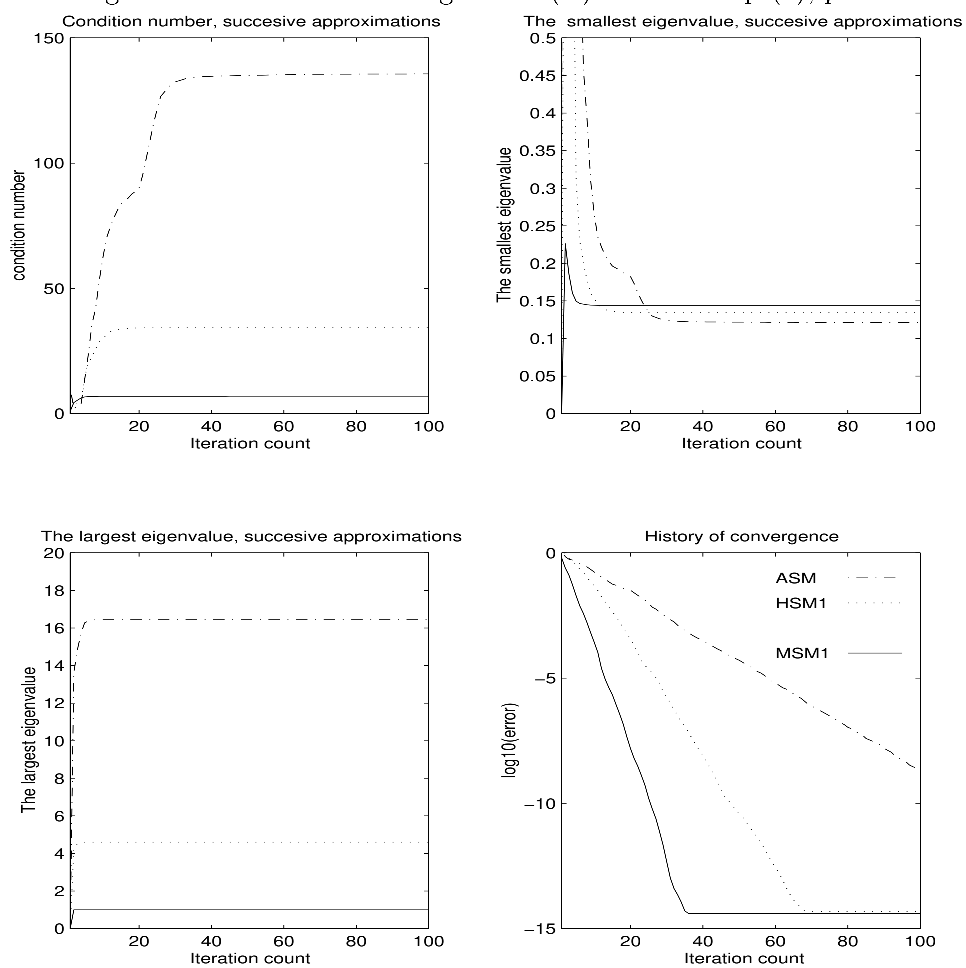


Figure 5.19: Vertex-based algorithm (II) with overlap (2); $p = 6$



For the experiments in Fig. 5.21, we have used the pseudoinverses of the local Schur to compute the preconditioner. We use three different coarse spaces, the standard coarse space and the wire basket space with the inexact solvers. This algorithm performs better than the previous one.

If we treat the singularities by imposing zero Dirichlet boundary conditions at the vertices only, or use no coarse space, then the convergence is very slow. Fig. 5.22 contains results of the experiments for the following choices of local and coarse spaces:

(1) Zero Dirichlet boundary conditions at the vertices of each substructure, and the standard vertex-based coarse space.

(2) No Dirichlet boundary conditions on the substructures, i.e. pseudoinverses are used, and no coarse space.

(3) Zero Dirichlet boundary conditions at the vertices of each substructure, and the wire basket space with inexact solver, with the natural scaling.

(4) Zero Dirichlet boundary conditions at the vertices of each substructure, and the wire basket space with inexact solver, with scaling such that $\omega = 1$.

Figure 5.20: Neumann-Neumann, zero Dirichlet boundary conditions on each local wire basket.

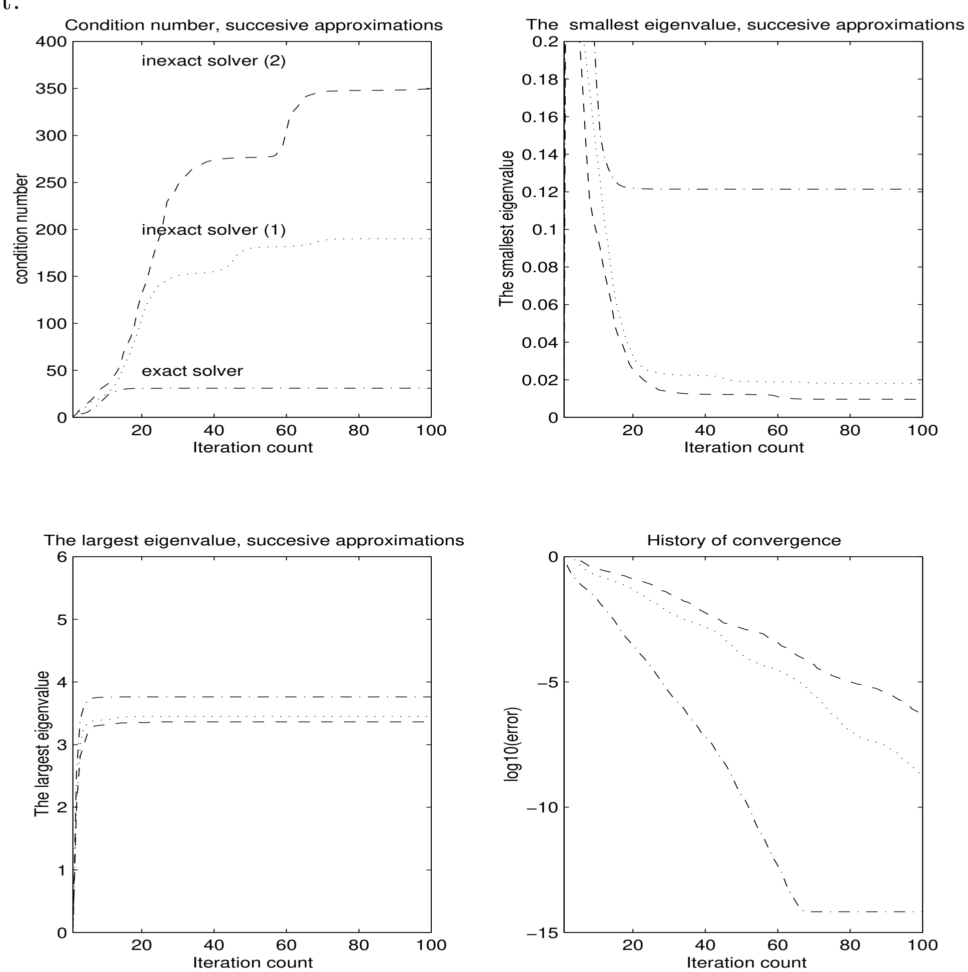


Figure 5.21: Neumann-Neumann, without Dirichlet boundary conditions on the local wire baskets.

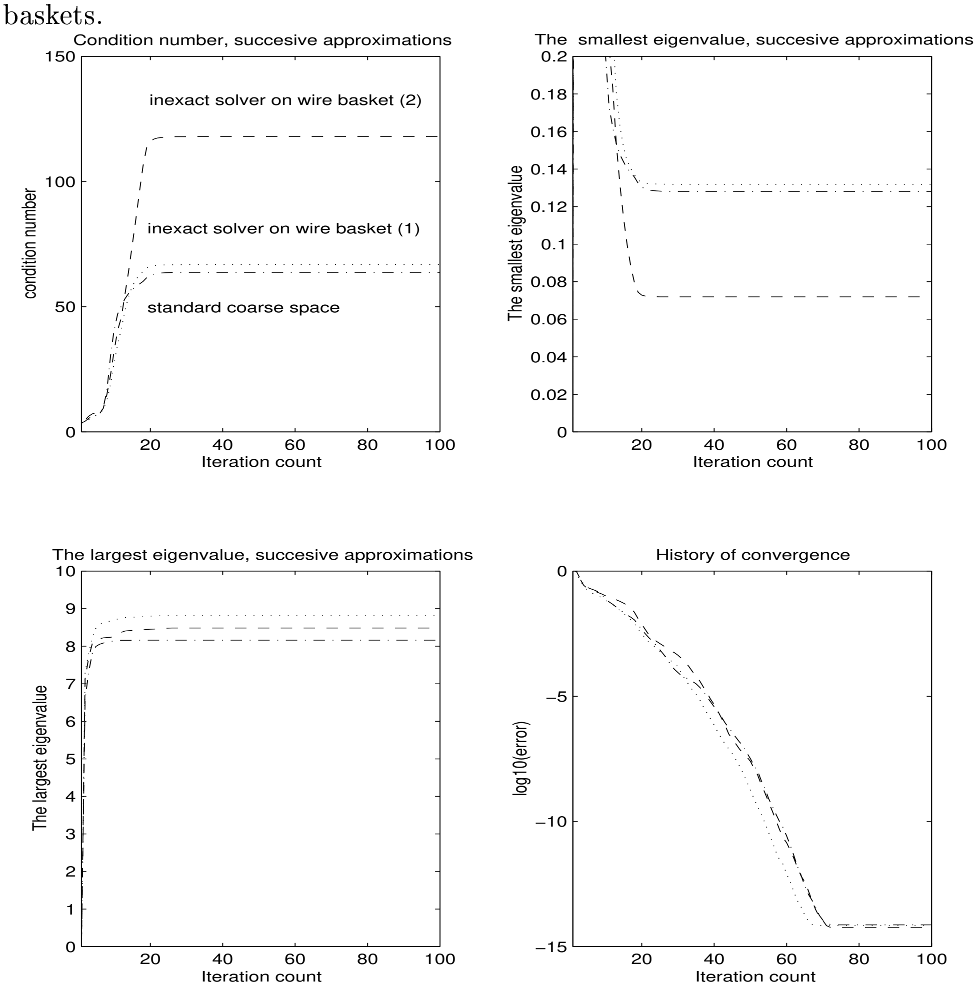
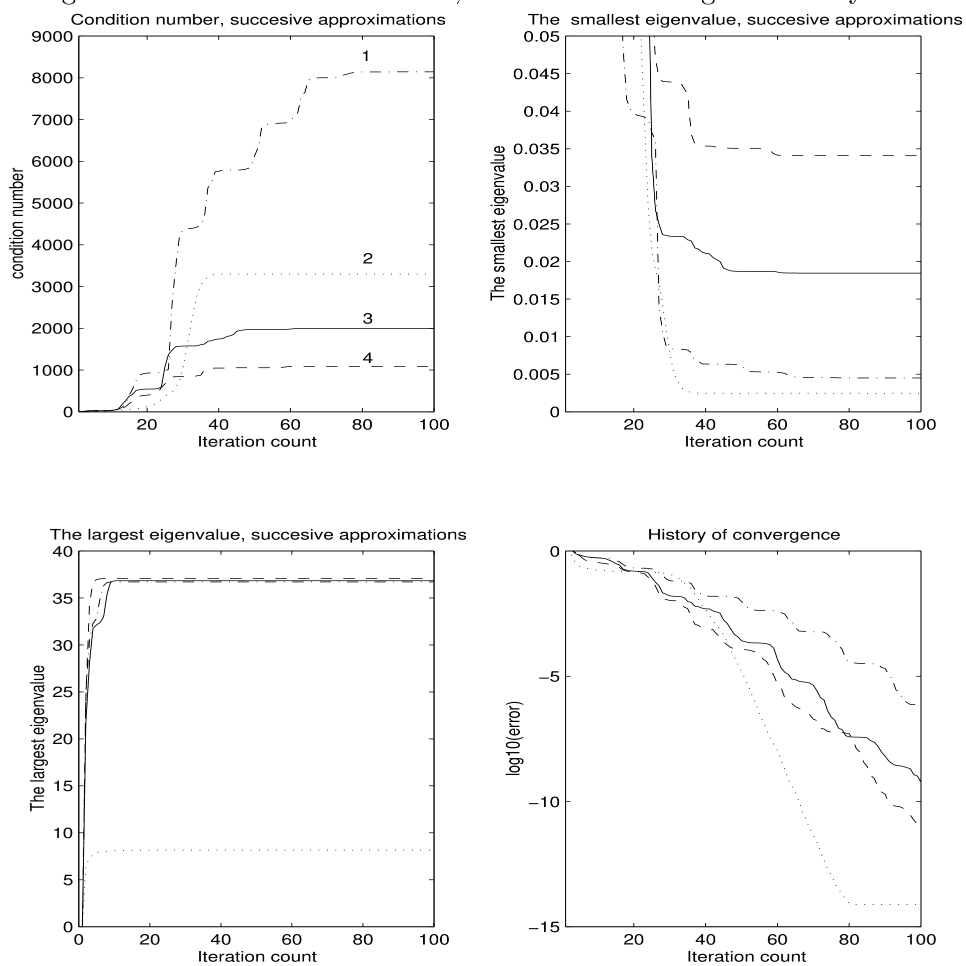


Figure 5.22: Neumann-Neumann, cases when convergence is very slow.



Chapter 6

Possible future work

We see several possible extensions of this work.

Extensions from the wire basket, based on quadrature points. This would be analogous, in the case of tetrahedral elements, to the work of Pavarino and Widlund [56]. We could use either the ideas of Sherwin and Karniadakis [59] or those of Babuška and Chen [7, 8].

Iterative substructuring methods for linear elasticity. The recent work of Pavarino and Widlund, on spectral elements [52, 53] seems to extend straightforwardly to our case. One technical difficulty, solved by these authors, is to make the extensions from the wire basket preserve the rigid body motions while also being low energy, which is not the case for our extension from the wire basket, and have low energy. We note that another approach has been explored by Mandel [42].

Iterative substructuring algorithms for the *hp* version finite element method. Domain decomposition algorithms for the *hp* version finite element method have been analysed by several authors in case of square, triangular, and cubic elements; see Chapter 1. There seems to be no major technical difficulty to extending our work to the *hp* version.

Large numerical experiments and parallel implementations. As we noted in the beginning of Chapter 5, our goal has been to develop an experimental code that allow us to compare the condition numbers of different algorithms. We have not found the time for not an efficient or parallel implementation. To more fully assess the practicality of our methods, it would be very interesting to study the performances of efficient, parallel implementations of these algorithms.

Bibliography

[1] R. A. Adams. *Sobolev Spaces.* Academic Press, New York, 1975.

[2] I. Babuška. The p and $hp-$version finite element method: the state of the art. In D. L. Dweyer, M. Hussaini, and R. G. Voigt, editors, *Finite Elements: theory and applications.* Springer-Verlag, 1988.

[3] I. Babuška and M. Suri. The $hp-$version finite element method on quasiuniform meshes. M^2AN, 21:199–238, 1987.

[4] I. Babuška and M. Suri. The optimal convergence rate of the $p-$version finite element method. *SIAM J. Numer Anal*, 24(4):750–776, 1987.

[5] I. Babuška and M. Suri. The p and $hp-$version finite element method, an overview. *Comp. Methods. Appl. Appl. Mech. Eng*, 80:5–26, 1990.

[6] I. Babuška, A. Szabó, and I. Katz. The $p-$version finite element method. *SIAM J. Numer Anal*, 18(4):515–545, 1981.

[7] Ivo Babuška and Qi Chen. Approximate optimal points for interpolation of real functions in an interval and in a triangle. *Comp. Methods Appl. Mech. Engrg.*, 128:405–417, 1995.

[8] Ivo Babuška and Qi Chen. The optimal symmetrical points for interpolation of real functions in a tetrahedron. *Comp. Methods Appl. Mech. Engrg.*, 137:89–94, 1996.

[9] Ivo Babuška, Alan Craig, Jan Mandel, and Juhani Pitkäranta. Efficient preconditioning for the p-version finite element method in two dimensions. *SIAM J. Numer. Anal.*, 28(3):624–661, 1991.

[10] Ivo Babuška, Michael Griebel, and Juhani Pitkäranta. The problem of selecting the shape functions for a p-type finite element. *Int. J. Numer. Meth. Eng.*, 28:1891 – 1908, 1989.

[11] C Bernardi, M Dauge, and Y Maday. Interpolation of nullspaces for polynomial approximation of a divergence-free function in a cube. Technical Report R93035, Universite Pierre and Marie Curie, 1994.

[12] Jean-François Bourgat, Roland Glowinski, Patrick Le Tallec, and Marina Vidrascu. Variational formulation and algorithm for trace operator in domain decomposition calculations. In Tony Chan, Roland Glowinski, Jacques Périaux, and Olof Widlund, editors, *Domain Decomposition Methods*, pages 3–16, Philadelphia, PA, 1989. SIAM.

[13] James H. Bramble, Joseph E. Pasciak, and Alfred H. Schatz. The construction of preconditioners for elliptic problems by substructuring, I. *Math. Comp.*, 47(175):103–134, 1986.

[14] James H. Bramble, Joseph E. Pasciak, and Alfred H. Schatz. The construction of preconditioners for elliptic problems by substructuring, IV. *Math. Comp.*, 53:1–24, 1989.

[15] James H. Bramble, Joseph E. Pasciak, Junping Wang, and Jinchao Xu. Convergence estimates for product iterative methods with applications to domain decomposition. *Math. Comp.*, 57(195):1–21, 1991.

[16] James H. Bramble and Jinchao Xu. Some estimates for a weighted L^2 projection. *Math. Comp.*, 56:463–476, 1991.

[17] Xiao-Chuan Cai. An optimal two-level overlapping domain decomposition method for elliptic problems in two and three dimensions. *SIAM J. Sci. Comp.*, 14:239–247, January 1993.

[18] Mario A. Casarin. Diagonal edge preconditioners in p-version and spectral element methods. Technical Report 704, Department of Computer Science, Courant Institute, September 1995. To appear in SIAM J. Sci. Comp.

[19] Mario A. Casarin. Quasi-optimal Schwarz methods for the conforming spectral element discretization. Technical Report 705, Department of Computer Science, Courant Institute, September 1995. To appear in SIAM J. Numer. Anal.

[20] Lawrence C. Cowsar, Jan Mandel, and Mary F. Wheeler. Balancing domain decomposition for mixed finite elements. *Math. Comp.*, 64(211):989–1015, July 1995.

[21] Yann-Hervé De Roeck and Patrick Le Tallec. Analysis and test of a local domain decomposition preconditioner. In Roland Glowinski, Yuri Kuznetsov, Gérard Meurant, Jacques Périaux, and Olof Widlund, editors, *Fourth International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 112–128. SIAM, Philadelphia, PA, 1991.

[22] M. R. Dorr. The approximaton theory for the $p-$version finite element method. *SIAM J. Numer. Anal.*, 21:1181–1207, 1984.

[23] M. R. Dorr. The approximaton of solutions of elliptic boundary value problems by the $p-$version finite element method. *SIAM J. Numer. Anal.*, 23, 1986.

[24] Maksymilian Dryja. A method of domain decomposition for 3-D finite element problems. In Roland Glowinski, Gene H. Golub, Gérard A. Meurant, and Jacques Périaux, editors, *First International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 43–61, Philadelphia, PA, 1988. SIAM.

[25] Maksymilian Dryja, Barry F. Smith, and Olof B. Widlund. Schwarz analysis of iterative substructuring algorithms for elliptic problems in three dimensions. *SIAM J. Numer. Anal.*, 31(6):1662–1694, December 1994.

[26] Maksymilian Dryja and Olof B. Widlund. An additive variant of the Schwarz alternating method for the case of many subregions. Technical Report 339, also Ultracomputer Note 131, Department of Computer Science, Courant Institute, 1987.

[27] Maksymilian Dryja and Olof B. Widlund. Towards a unified theory of domain decomposition algorithms for elliptic problems. In Tony Chan, Roland Glowinski, Jacques Périaux, and Olof Widlund, editors, *Third International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 3–21. SIAM, Philadelphia, PA, 1990.

[28] Maksymilian Dryja and Olof B. Widlund. Domain decomposition algorithms with small overlap. *SIAM J. Sci.Comput.*, 15(3):604–620, May 1994.

[29] Maksymilian Dryja and Olof B. Widlund. Schwarz methods of Neumann-Neumann type for three-dimensional elliptic finite element problems. *Comm. Pure Appl. Math.*, 48(2):121–155, February 1995.

[30] P. Grisvard. Caractérisation de quelques espaces d'interpolation. *Arch. Rat. Mech. Anal.*, 25:40–63, 1967.

[31] Benqi Guo and Weiming Cao. A preconditioner for the h-p version of the finite element method in two dimensions. *Numer. Math.*, 75:59–77, 1996.

[32] Benqi Guo and Weiming Cao. An additive schwarz method for the $hp-$version finite element method in three dimensions. *SIAM Sci. Comp.*, To appear.

[33] Benqi Guo and Weiming Cao. Preconditioning for the h-p version of the finite element method in two dimensions. *SIAM Numer. Analysis.*, To appear.

[34] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities.* Cambridge University Press, Cambridge, 1934.

[35] V. G. Korneev and S. Jensen. Preconditioning of the $p-$version of the finite element method. Technical report, University of Maryland Baltimore county, 1996.

[36] Yuri Kuznetsov, Petri Manninen, and Yuri Vassilevski. On numerical experiments with Neumann-Neumann and Neumann-Dirichlet domain decomposition preconditioners. Technical report, University of Jyväskylä, 1993.

[37] Cornelius Lanczos. An iterative method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Nat. Bur. Standards, Sect. B.*, 45:225–280, 1950.

[38] M. Le Tallec, Y. H. De Roeck, and M. Vidrascu. Domain decomposition methods for large linearly elliptic three dimensional problems. *J. Comput. Appl. Math., 34, 1991.*, 34, 1991.

[39] Jacques-Louis Lions and Enrico Magenes. *Nonhomogeneous Boundary Value Problems and Applications*, volume I. Springer, New York, Heidelberg, Berlin, 1972.

[40] Pierre-Louis Lions. On the Schwarz alternating method. I. In Roland Glowinski, Gene H. Golub, Gérard A. Meurant, and Jacques Périaux, editors, *First International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 1–42, Philadelphia, PA, 1988. SIAM.

[41] Yvon Maday. Relèvement de traces polynomiales et interpolations hilbertiennes entre espaces de polynômes. *C. R. Acad. Sci. Paris*, 309, Série I:463–468, 1989.

[42] Jan Mandel. Two-level domain decomposition preconditioning for the p-version finite element version in three dimensions. *Int. J. Numer. Meth. Eng.*, 29:1095–1108, 1990.

[43] Jan Mandel. Balancing domain decomposition. *Comm. Numer. Meth. Engrg.*, 9:233–241, 1993.

[44] Jan Mandel and Marian Brezina. Balancing domain decomposition for problems with large jumps in coefficients. Math. Comp., to appear.

[45] Jan Mandel and Marian Brezina. Balancing domain decomposition: Theory and computations in two and three dimensions. Technical Report UCD/CCM 2, Center for Computational Mathematics, University of Colorado at Denver, 1993.

[46] Sergey V. Nepomnyaschikh. *Domain Decomposition and Schwarz Methods in a Subspace for the Approximate Solution of Elliptic Boundary Value Problems.* PhD thesis, Computing Center of the Siberian Branch of the USSR Academy of Sciences, Novosibirsk, USSR, 1986.

[47] Jindřich Nečas. *Les méthodes directes en théorie des équations elliptiques.* Academia, Prague, 1967.

[48] Rafael Mu noz Sola. Polynomial liftings on a tetrahedron and applications to the $h-p$ version of the finite element method in three dimensions. *SIAM J. Numer. Anal.*, 34(1):282–314, February 1997.

[49] J. T. Oden, L. Demkowicz, W. Rachowicz, and T. A. Wastermann. Toward a universal $hp-$ adaptive finite element strategy. part 2. a posteriori estimates. *Comp. Methods. Appl. Appl. Mech. Eng*, 77:113–180, 1989.

[50] J. T. Oden, Abani Patra, and Yusheng Feng. Parallel domain decomposition solver for adaptive hp finite element methods. Technical report, TICAM, University of Texas at Austin, 1994.

[51] D. Pathria and G. E. Karniadakis. Spectral element methods for elliptic problems in nonsmooth domains. *J. Comp. Physics*, 123:83–95, 1995.

[52] Luca Pavarino and Olof Widlund. Iterative substructuring methods for spectral element discretization of elliptic systems. i: Compressible linear elasticity. Technical report, New York University, To appear.

[53] Luca Pavarino and Olof Widlund. Iterative substructuring methods for spectral element discretization of elliptic systems. ii: Mixed methods for linear elasticity and stokes flow. Technical report, New York University, To appear.

[54] Luca F. Pavarino. Neumann-Neumann algorithms for spectral elements in three dimensions. Technical Report I.A.N.-CNR 980, Istituto di Analisi Numerica del CNR, Pavia, Italy, 1995. To appear in RAIRO Mathematical Modelling and Numerical Analysis.

[55] Luca F. Pavarino and Olof B. Widlund. A polylogarithmic bound for an iterative substructuring method for spectral elements in three dimensions. *SIAM J. Numer. Anal.*, 33(4):1303–1335, August 1996.

[56] Luca F. Pavarino and Olof B. Widlund. Iterative substructuring methods for spectral elements: Problems in three dimensions based on numerical quadrature. *Computers Math. Applic.*, 33(1/2):193–209, January 1997.

[57] Theodore J. Rivlin. *The Chebyshev Polynomials.* Wiley Interscience, 1990.

[58] H. A. Schwarz. *Gesammelte Mathematische Abhandlungen*, volume 2, pages 133–143. Springer, Berlin, 1890. First published in Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich, volume 15, 1870, pp. 272–286.

[59] S. J. Sherwin and G. E. Karniadakis. A new triangular and tetrahedral basis for high order (*hp*) finite element methods. *IJNM*, pages 3775–3802, 1995.

[60] S. J. Sherwin and G. E. Karniadakis. A trianguler spectral element method; applications to the incopresible navier-stokes equations. *omp. Meth. Appl. Mech. Eng.*, 123:189–229, 1995.

[61] S. J. Sherwin and G. E. Karniadakis. Tetrahedral *hp* finite elements: algorithms and flow simulations. *J. Comp. Physics*, 124:14–45, 1996.

[62] Barry F. Smith, Petter Bjørstad, and William Gropp. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations.* Cambridge University Press, 1996.

[63] B. A. Szabó. Some recent developments in finite element analysis. *Comp. Methods. Appl. Appl. Mech. Eng*, 5:99–115, 1979.

[64] B. A. Szabó and A. K. Mehta. P-convergent finite element approximation in fracture mechanics. *Int. J. Numer. Meth. Eng.*, 12:551–560, 1978.

[65] Barna Szabó and Ivo Babuška. *Finite Element Analysis.* John Wiley & Sons, New York, 1991.

[66] Jinchao Xu. Iterative methods by space decomposition and subspace correction. *SIAM Review*, 34(4):581–613, December 1992.

[67] Xuejun Zhang. *Studies in Domain Decomposition: Multilevel Methods and the Biharmonic Dirichlet Problem.* PhD thesis, Courant Institute, New York University, September 1991.

[68] Gerhard Zumbusch. *Simultaneous h-p Adaptation in Multilevel Finite Elements.* PhD thesis, Freie Universität Berlin, April 1995.