



TNF

Technisch-Naturwissenschaftliche  
Fakultät

# A Non-standard Finite Element Method using Boundary Integral Operators

DISSERTATION

zur Erlangung des akademischen Grades

Doktor

im Doktoratsstudium der

Technischen Wissenschaften

Eingereicht von:

DI Clemens Hofreither

Angefertigt am:

Doktoratskolleg Computational Mathematics

Beurteilung:

O.Univ.-Prof. Dipl.-Ing. Dr. Ulrich Langer (Betreuung)

Prof. Dr. Sergej Rjasanow

Linz, Oktober, 2012



# Acknowledgments

This thesis was created during my employment at the Doctoral Program “Computational Mathematics” (W1214) at the Johannes Kepler University in Linz.

First of all I would like to thank my advisor, Prof. Ulrich Langer, for interesting me in the research topic treated in this thesis, for giving me the chance to work on it within the framework of the Doctoral Program, and for his guidance and advice during my work on the thesis. Even before, during my graduate studies and supervising my diploma thesis, he has taught me a lot and shaped my academic career in a significant way.

Furthermore, I am very grateful to Clemens Pechstein for the countless hours he has invested in the past four years into discussions which have deepened my insight into many topics. Without him, many results of this thesis would not have been possible.

I want to thank all the other past and present students at the Doctoral Program for creating a very pleasant environment in which to discuss and work, as well as for the friendships which have developed during my time at the Program, making this time not only a productive, but also a highly enjoyable one. Furthermore, the director of the Doctoral Program, Prof. Peter Paule, must not go unmentioned for fostering such a positive atmosphere to work in.

I want to thank my hosts at foreign research institutions who have enabled me to spend, all in all, more than half a year in fruitful research stays abroad. In particular, I extend my thanks to Prof. Sergej Rjasanow and his whole group at the Universität des Saarlandes in Saarbrücken, Germany, for providing me with a great working environment and making my two prolonged stays very pleasant. Furthermore, I thank my coauthors and friends Irina Georgieva and Rumén Uluchev for inviting me to their respective institutions in Sofia, as well as for interesting me in problems of approximation theory which have become a fascinating and productive second research interest for me. In this context, I also want to thank Veronika Pillwein and Christoph Koutschan from the Research Institute for Symbolic Computation, Linz, for taking an immediate interest in the questions that we posed to them and being such productive forces in the interdisciplinary cooperation that has since developed.

I am deeply grateful for the constant support by my family and friends. Without them, I could never have hoped to succeed in the work I have done in the past four years. Thank you.

Finally, the financial support by the Austrian Science Fund (FWF) under grant W1214-N15, project DK04, is gratefully acknowledged.



## KURZFASSUNG

Das Thema dieser Dissertation ist die Analyse eines relativ neuen Diskretisierungsverfahrens für Randwertprobleme elliptischer partieller Differentialgleichungen zweiter Ordnung. Das Verfahren zerlegt das Rechengebiet in Elemente und verwendet Ansatzfunktionen mit lokalem Träger. Im Gegensatz zu herkömmlichen Finite Elemente-Methoden sind Gitter bestehend aus polygonalen oder polyhedralen Elementen zulässig, und die Ansatzfunktionen sind nicht lokale Polynome, sondern erfüllen lokal die partielle Differentialgleichung (PDE-harmonisch, von engl. *partial differential equation*). Das Verfahren steht somit in gewisser Weise in der Tradition der Trefftz-Methoden. Die PDE-harmonischen Ansatzfunktionen werden über die Lösung von element-lokalen Randwertproblemen eingeführt. Ein charakteristisches Merkmal der Methode ist, dass diese lokalen Probleme mittels Randintegraloperatoren behandelt werden. Genauer kommen Randelementmethoden (BEM, von engl. *boundary element method*) zum Einsatz. Die Methode wird daher als BEM-basierte FEM bezeichnet. Tatsächlich kann sie als Finite Elemente-Methode aufgefasst werden, in der die Berechnung der Element-Steifigkeitsmatrizen über die BEM geschieht.

Für die Konstruktion der Randintegraloperatoren ist die explizite Kenntnis einer Fundamentallösung des partiellen Differentialoperators Voraussetzung. Durch die lokale Konstruktion sind jedoch, im Gegensatz zu herkömmlichen Randelementmethoden, nur Fundamentallösungen für lokale Element-Probleme erforderlich. Wir betrachten daher partielle Differentialgleichungen mit stückweise konstanten Koeffizienten, da eine Fundamentallösung für Operatoren mit konstanten Koeffizienten in der Literatur zu finden ist.

Statt als Trefftz-Methode kann die BEM-basierte FEM auch als Variante eines Gebietszerlegungsverfahrens mit Randintegraloperatoren aufgefasst werden. Der Hauptunterschied besteht hier in der Diskretisierungsstrategie: in Gebietszerlegungsverfahren sind die Teilgebiete üblicherweise von moderatem bis großem Ausmaß, um effiziente parallele Behandlung zu ermöglichen, während wir in der BEM-basierten FEM diese Strukturen wie Elemente in der Finite Elemente-Methode mit relativ geringer Anzahl an Freiheitsgraden betrachten. Dies hat auch Konsequenzen für die Analyse der Methode, da wir neue technische Hilfsmittel für allgemeine polygonale oder polyhedrale Gitter entwickeln müssen, deren Maschengröße gleichmäßig gegen null geht. Solche Abschätzungen werden in der FEM-Literatur üblicherweise über das Abbildungsprinzip bewiesen, ein Zugang, der für heterogene Gitter von Vielfächern scheitert. Neue Ideen sind daher vonnöten.

Nach der Herleitung dieser Werkzeuge ist das erste Hauptresultat der Dissertation der Beweis von Fehlerabschätzungen für die BEM-basierte FEM für ein Modellproblem. Wir beweisen sowohl  $H^1$ - als auch  $L_2$ -Fehlerabschätzungen, wobei letztere den Übergang zu einer äquivalenten gemischten Formulierung nötig machen. Die Abschätzungen sind quasi-optimal in Bezug auf die Approximationseigenschaften des zugrunde liegenden diskreten Skelettraumes. Weitere Resultate umfassen die Herleitung und Konvergenzanalyse eines effizienten parallelen Löser für die resultierenden linearen Gleichungssysteme, welcher auf den Ideen des FETI-Gebietszerlegungsverfahrens basiert. Weiters wenden wir die BEM-basierte FEM auf Konvektions-Diffusionsprobleme an und beobachten, dass die Verwendung PDE-harmonischer Ansatzfunktionen einen Stabilitätsvorteil gegenüber herkömmlichen FEM-Diskretisierungen mit sich bringt. Wir zeigen, dass die BEM-basierte FEM für solche Probleme in enger Verwandtschaft mit der Methode der *residual-free bubbles* und somit auch mit SUPG steht. Im letzten Kapitel präsentieren wir numerische Beispiele, um die theoretischen Resultate der Dissertation zu untermauern.



## ABSTRACT

This thesis is concerned with the analysis of a relatively novel discretization scheme for boundary value problems of second-order elliptic partial differential equations. The method decomposes the computational domain into elements and uses trial functions with local support. In contrast to standard finite element method (FEM) discretizations, the mesh may consist of arbitrary polygonal or polyhedral elements, and the trial functions are not locally polynomial, but locally PDE-harmonic, by which we mean that they satisfy the partial differential equation on every element. The method may thus be considered to be in the tradition of Trefftz methods. The PDE-harmonic trial functions are constructed via the solution of element-local boundary value problems. A characteristic feature of the scheme discussed in this thesis is that these local problems are tackled using boundary integral operators, and in particular a boundary element method (BEM) discretization. For this reason, we refer to the method as a BEM-based FEM. Indeed, it can be regarded as a finite element method where the element stiffness matrices are computed using boundary element techniques.

For the construction of the involved boundary integral operators, explicit knowledge of a fundamental solution of the partial differential operator is required. However, due to the local construction, we only need local fundamental solutions for the element problems, in contrast to standard BEM approaches. Therefore, we study the setting of elementwise constant coefficients of the partial differential operator, since a fundamental solution is readily available in the literature for operators with constant coefficients.

Alternatively to the interpretation as a Trefftz method, the method may also be viewed as a variant of a domain decomposition technique using boundary integral operators. The main difference to this approach lies in the discretization strategy: in contrast to domain decomposition methods, where the subdomains are typically of moderate to large size in order to enable efficient parallel processing, we consider the substructures in the BEM-based FEM as elements with only a small number of degrees of freedom. This has ramifications for the analysis as well since we need to develop analytical tools for arbitrary polygonal or polyhedral meshes with mesh sizes which uniformly tend to zero. These estimates are typically proven using the mapping principle in the FEM literature, an approach which fails for heterogeneous polytopal meshes. Thus, new techniques have to be developed for deriving these results.

With these analytical tools at hand, the first major result of the thesis is the derivation of rigorous error estimates for the BEM-based FEM on heterogeneous polyhedral meshes for a model problem. In particular, we prove both  $H^1$ - and, by passing to an equivalent mixed formulation,  $L_2$ -error estimates which are quasi-optimal with respect to the approximation properties of the underlying skeletal space. Further results include the derivation and convergence analysis of an efficient parallel solver for the resulting system of linear equations which is based on the ideas of the one-level finite element tearing/interconnecting (FETI) substructuring technique. Furthermore, we consider the application of the method to convection-diffusion problems, where the use of PDE-harmonic trial functions confers a stability advantage over a standard FEM discretization. In fact, we show that the BEM-based FEM is closely related to the method of residual-free bubbles and thus also to the well-established SUPG scheme. In a final chapter, we present numerical examples in order to confirm some of the theoretical results of the thesis.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Numerical methods for elliptic partial differential equations . . . . .	1
1.2	Prior and related work . . . . .	5
1.3	Trefftz's methods . . . . .	7
1.4	Outline . . . . .	11
<b>2</b>	<b>Preliminaries</b>	<b>13</b>
2.1	Function spaces . . . . .	13
2.2	Variational problems . . . . .	17
2.3	Green's identities . . . . .	21
2.4	Boundary value problems . . . . .	21
2.5	The finite element method . . . . .	26
<b>3</b>	<b>The boundary element method</b>	<b>29</b>
3.1	Boundary integral operators . . . . .	29
3.2	Boundary integral equations and the Steklov-Poincaré operator . . . . .	34
3.3	Approximation of the Steklov-Poincaré operator . . . . .	36
3.4	Explicit constants for boundary integral operators . . . . .	38
<b>4</b>	<b>Derivation of the BEM-based FEM</b>	<b>45</b>
4.1	A skeletal variational formulation . . . . .	45
4.2	A mixed variational formulation . . . . .	50
4.3	A BEM-based finite element method . . . . .	51
4.3.1	Discrete function spaces . . . . .	51
4.3.2	Discretization of the primal formulation . . . . .	51
4.3.3	Discretization of the mixed formulation . . . . .	53
<b>5</b>	<b>Properties of polytopal elements</b>	<b>55</b>
5.1	Norms . . . . .	56
5.1.1	The energy norm . . . . .	56
5.1.2	The Sobolev-Slobodeckii norm . . . . .	57
5.2	Transformation properties . . . . .	57
5.3	Trace inequalities . . . . .	60
5.4	An auxiliary harmonic extension norm . . . . .	61

5.5	Approximation properties . . . . .	63
5.6	Explicit extension results . . . . .	65
5.6.1	The Jones extension operator . . . . .	65
5.6.2	An extension operator in the spirit of Stein . . . . .	67
<b>6</b>	<b>Error analysis</b>	<b>71</b>
6.1	Error of the inexact Galerkin scheme . . . . .	71
6.2	Geometric assumptions on the mesh . . . . .	74
6.3	Approximation error in the Dirichlet data . . . . .	76
6.4	Approximation error in the Neumann data . . . . .	78
6.5	$H^1$ -error estimate . . . . .	78
6.6	Analysis of the mixed formulation . . . . .	80
6.6.1	Convergence on the skeleton . . . . .	81
6.6.2	Convergence in the $H^1$ -norm . . . . .	82
6.6.3	Convergence in the $L_2$ -norm . . . . .	82
<b>7</b>	<b>Fast solution methods</b>	<b>85</b>
7.1	A FETI-type solver . . . . .	86
7.1.1	Derivation . . . . .	86
7.1.2	Preconditioning . . . . .	91
7.2	Convergence analysis . . . . .	92
7.2.1	The non-preconditioned case . . . . .	92
7.2.2	The preconditioned case . . . . .	97
<b>8</b>	<b>Convection-diffusion problems</b>	<b>99</b>
8.1	Stabilized methods: an overview . . . . .	99
8.2	Residual-free bubbles and their relation to the BEM-based FEM . . . . .	102
<b>9</b>	<b>Implementation and numerical examples</b>	<b>107</b>
9.1	Implementation details . . . . .	107
9.2	Numerical experiments . . . . .	108
9.2.1	Example 1: The Laplace equation . . . . .	108
9.2.2	Example 2: A convection-diffusion problem . . . . .	110
9.2.3	Example 3: Domain decomposition . . . . .	112
<b>10</b>	<b>Conclusion and outlook</b>	<b>119</b>
10.1	Conclusion . . . . .	119
10.2	Outlook and further work . . . . .	120
	<b>Bibliography</b>	<b>123</b>

# Chapter 1

## Introduction

### 1.1 Numerical methods for elliptic partial differential equations

Partial differential equations (PDEs) are one of the most important tools of mathematical modeling available today. They are used to describe such diverse physical phenomena like heat transfer, diffusion, mechanics of elastic and plastic materials, fluid mechanics, electrostatics and -dynamics, and many more. However, only few partial differential equations in certain simple settings permit an analytic solution. It is therefore no surprise that the numerical solution, or to be more precise, the approximation of solutions to partial differential equations by numerical methods, has become one of the main areas of research in computational mathematics ever since the development of computers has made such computations feasible for problems on large scales.

Many different numerical methods for PDEs, often adapted to particular problems and settings, have been devised. Finite difference methods (see, e.g., [55]) are certainly among the most traditional discretization schemes for PDEs. They operate on the principle of subdividing the computational domain into a regular grid and approximating derivatives by difference quotients. While they are easy to understand and implement, they suffer from a series of drawbacks, such as inflexibility with respect to geometry, difficulty of generalizing to higher orders of approximation, and inability to perform true adaptive local refinement. It should be mentioned, however, that some modern schemes like the finite volume method [34] often used in computation fluid dynamics or the finite integration method for Maxwell equations [122, 120] can be considered as more sophisticated offspring of the finite difference method which manage to overcome some of these problems.

In the forties of the twentieth century, the ideas of the finite element method (FEM) were developed. Richard Courant introduced the method in the form that we essentially still know today in a two-page appendix entitled “Numerical Treatment of the Plane Torsion Problem for Multiply-Connected Domains” which he added to the published version of an address he gave in 1942 to the American Mathematical Society [27]. Therein, he introduced the use of piecewise linear functions on a family of triangles for the solution of partial differential equations. Indeed, he calls his approach a “generalized method of finite differences on triangular nets.” As pioneers in the early development of the FEM, we

also mention Alexander Hrennikov, John Argyris, and Olgierd Zienkiewicz. Its invention was an evolutionary step based on many previous results for PDEs by, among many others, Lord Rayleigh, Walther Ritz, Ivan Bubnov, and Boris Galerkin. The method was further developed in the following decades and gained traction as computers became more widespread. It was put on a solid mathematical foundation and rigorously analyzed in the 1970s, where we refer in particular to the works of Strang and Fix [115] and Ciarlet [22]. From the Russian school in St. Petersburg, one must mention the contributions of Oganessian, Rivkind, Rukhovets, Korneev, and others. It is an interesting historical note that a construction resembling that of the finite element method was already given by Karl Schellbach in an article on variational calculus in 1852 [107]. For an overview of the history of the finite element method, we refer to historical surveys by Oden [90], Babuška [4], Oganessian and Rivkind [91], and Gander and Wanner [44], as well as to the web page of Martin J. Gander<sup>1</sup>.

Instead of treating the partial differential equation in its classical formulation, the core idea behind the FEM in its modern formulation is to pass to a so-called weak formulation, which takes the form of a variational equation in Hilbert or Banach spaces. The functional spaces are then discretized by replacing them with finite-dimensional subspaces of functions which are usually piecewise polynomials, or maps of piecewise polynomials, with respect to the “elements” of a predetermined mesh which typically consists of simplices, quadrilaterals, or hexahedra. In this way, a discretized formulation of the partial differential equation is obtained. It can easily be viewed as a system of equations which, for linear partial differential equations, is again linear. Methods from numerical linear algebra are then used to solve these linear and sparse systems. For many classes of problems, rigorous error estimates for the resulting approximation to the exact solution are known. The techniques used to derive these error estimates stem from functional analysis. We will give the most fundamental results in this area, as far as they are relevant to the present work, in Chapter 2.

The finite element method is today one of the most established approaches for the numerical solution of PDEs and has both a broad foundation in mathematical theory, where it is still an active area of research, as well as an excellent track record in practical applications in science and industry.

Another established method for the solution of partial differential equations is the boundary element method (BEM). Its development has its roots in the potential theory, Green’s identities, and the theory of singular integrals [66, 87]. The core idea is the reduction of the space dimension by using a Green’s identity to reduce a boundary value problem for a PDE to the boundary of the computational domain. Using a so-called fundamental solution for the partial differential operator, which has to be explicitly known, the solution is essentially represented in terms of potentials of its boundary data, in particular its Dirichlet data (i.e., the boundary traces of the functions) and its

---

<sup>1</sup><http://www.unige.ch/~gander/historicalreferences.php>

Neumann data (i.e., the normal derivative, or more generally conormal derivative, of the function at the boundary). Collectively, these pieces of information are sometimes referred to as the Cauchy data of the solution. By taking suitable traces of the potential representations, one obtains boundary integral equations which link these pieces of data. In a standard well-posed boundary value problem for a second-order PDE, only certain parts of the Cauchy data are given. By proper manipulation of the boundary integral equations, one obtains a formulation for recovering the unknown components of the Cauchy data. These integral equations are then discretized by introducing a finite element-type mesh which is now defined only on the boundary of the computational domain, referred to as a boundary element mesh. Boundary elements are typically line segments for boundaries of two-dimensional domains and triangles or quadrilaterals for boundaries of three-dimensional domains [64, 112, 102, 106]. Variations with curved boundary elements have been developed as well [106].

Due to the inherent dimensional reduction, a BEM formulation typically has less unknowns than a comparable FEM formulation for the same problem, but owing to the fact that the involved boundary integral operators are non-local, the resulting matrices are densely populated. This makes the solution of the resulting linear systems more computationally expensive. However, by the use of suitable data-sparse approximation techniques like multipole expansions or hierarchical matrices with adaptive cross approximation of subblocks (cf. [9, 56, 7, 102]), methods of quasi-optimal computational complexity and memory requirements can be obtained. By solving the discretized integral equations, one obtains approximations to the complete Cauchy data of the solution. Should an approximation to the solution itself be required, representation formulae provide a way to evaluate it and its derivatives at any point of the computational domain with high accuracy.

While the BEM is overall less popular than the FEM, it has been applied with great success to specialized problems in many fields where it is valued for its advantages, which include easy treatment of unbounded domains, the need to triangulate only the boundary as opposed to the entire computational domain, reduction of the number of unknowns due to the reduced dimensionality, and flexibility with respect to domains which change over time. Furthermore, it has become apparent that often a coupling of the FEM and the BEM can result in a method which yields the “best of both worlds” ([26, 112, 80]). We give an introduction to the boundary element method in Chapter 3.

The topic of this dissertation is a relatively new numerical method for the solution of elliptic partial differential equations which marries certain concepts from the finite and the boundary element methods. For this reason, it has sometimes been referred to as a BEM-based FEM, and we will use this name in the remainder of the work. To our knowledge, the method has first been studied and implemented by Copeland, Langer, and Pusch [25]. The method employs local boundary integral operators and can be viewed as a variation of the symmetric boundary element domain decomposition method proposed by Hsiao and Wendland [65].

The BEM-based FEM shares with the FEM the characteristic property that the computational domain is subdivided into a mesh of elements, and that the global stiffness matrix is assembled from contributions of element stiffness matrices. The resulting linear system is therefore sparse as in the FEM. In a certain sense, the method is the result of asking the question: what if, instead of the piecewise polynomial functions which are typically used as trial functions in finite element methods, we use trial functions which satisfy the partial differential equation locally in each element? In essence, this idea goes back to a historical method introduced by E. Trefftz [118] which we will summarize in Section 1.3.

In the BEM-based FEM, the piecewise PDE-harmonic trial functions (as we will call them in the sequel) are constructed in a particular way, namely as PDE-harmonic extensions of Dirichlet data which is prescribed on the boundary of each individual element. This leads to the use of so-called *skeletal* function spaces whose members live on the skeleton of the mesh, i.e., the union of all element boundaries. Indeed, the first step in the derivation of the BEM-based FEM in Section 4.1 will consist in obtaining a skeletal variational formulation which is equivalent in a certain sense to the standard domain variational formulation used in the FEM. We will see there that this skeletal variational formulation involves both the Dirichlet and the Neumann data of the trial functions on each element boundary. While the Dirichlet data is prescribed, the Neumann data is not easily accessible as it essentially requires the solution of a boundary value problem on each individual element. The exact solution of this problem is in general just as infeasible as the solution of the global problem we originally set out to solve. Therefore, the local Neumann data have to be approximated.

The approach we use for this approximation in this method, and which gives it the name BEM-based FEM, is to treat each element as a computational domain for a boundary element method. It is well known that the mapping from the Dirichlet to the Neumann data has a concise representation in terms of the boundary integral operators which appear in the standard boundary integral equations. After discretizing these operators, and discretizing the skeletal spaces by introducing piecewise linear and piecewise constant skeletal functions, we obtain a system of linear equations which shares many properties with that of the finite element method. In particular, the stiffness matrix we obtain in the BEM-based FEM for a symmetric PDE is symmetric, positive definite, and sparse. Even more: for the special case of the Laplace equation on a purely simplicial (i.e., triangular or tetrahedral) mesh, the resulting stiffness matrix is, modulo quadrature errors, *identical* to the one appearing in the standard Courant FEM.

An interesting side effect of constructing trial functions in the way described above is that one obtains a lot of freedom as far as the mesh geometry is concerned. Piecewise polynomial functions have a predetermined number of degrees of freedom in every element which are typically assigned to vertices or other geometric features of the element. This means that, for instance, a conforming FEM with piecewise linear trial functions usually requires a mesh of triangles in 2D or tetrahedra in 3D. Since our trial functions are PDE-

harmonic extensions of some given boundary data, we have no such restrictions on the number of degrees of freedom, which allows us to use very general polygonal or polyhedral element shapes. Even “polygonal” or “polyhedral” elements with curved edges or faces can easily be incorporated. The latter case can be realized by a suitable parameterization of the edges or faces, or by the ideas of isogeometric analysis [71]. Furthermore, we can mix any number of different element shapes within a single mesh without problems. Polyhedral elements are a useful tool, for instance, in reservoir simulation, and appear naturally in the modeling of the *stratum corneum*, the outermost layer of the skin [40]. Their use also gives great freedom in automatic mesh manipulation: elements can be split, joined and manipulated freely without the need to maintain a particular element topology. As an example, this feature is advantageous in adaptive mesh refinement: straightforward subdivision of individual elements usually results in so-called hanging nodes that are vertices of one element, but not a neighboring one. In our approach, such nodes can be artificially introduced in the latter element since we do not have to respect a particular number of degrees of freedom per element. We can thus retain the conformity of the discretization even in such cases.

Another interesting feature of the method stems from the use of PDE-harmonic trial functions itself. While for simple problems without stability problems like the Laplace equation this does not typically confer observable advantages, there is some indication that this property is advantageous for more difficult problems like convection-dominated equations. Indeed, in Chapter 8, we show that the BEM-based FEM for convection-diffusion problems is closely related to some established stabilized finite element methods, namely the residual-free bubbles method and the streamline-upwind/Petrov-Galerkin (SUPG) scheme. We also give some numerical results which serve to illustrate the improved stability of the BEM-based FEM in such situations in Chapter 9.

## 1.2 Prior and related work

The BEM-based FEM originated from the work of the BEM community, in particular, certain domain decomposition approaches using boundary integral equations. Specifically, in a work by Hsiao and Wendland [65], they proposed a domain decomposition technique using the symmetric Steklov-Poincaré approximation for coupling which already matches the formulation of the BEM-based FEM closely. Even more, while the authors focused on the domain decomposition case, they added a remark about the possibility to let the sizes of the domains become small compared to the size of the computational domain. We also point to later works by Langer [78] as well as Hsiao, Steinbach, and Wendland [67]. These references have in common that the analysis was performed therein in a fashion which is not explicit with respect to the diameters or shapes of the BEM domains. This is perfectly reasonable in certain domain decomposition settings, but is not applicable in the present more FEM-like setting where one typically considers families of meshes

whose element diameters tend to zero. It is therefore a major aim of the present work to derive FEM-like error estimates which are explicit in the mesh size and depend only on certain mesh regularity parameters.

As mentioned above, the BEM-based FEM as considered herein has first been proposed and implemented by Copeland, Langer, and Pusch [25]. This work already gives several numerical experiments for different problems including the diffusion equation, the Helmholtz equation and the Maxwell equations in the frequency domain (see also Copeland [24]). A rigorous error analysis was however missing from these first works.

A significant contribution to the study of the BEM-based FEM was made by Steffen Weißer in his dissertation [124] as well as the prior publication [123]. In contrast to the present work, which is more concerned with a priori error estimates, the focus in his work is on a posteriori error estimates and adaptive mesh refinement. In particular, we mention that he succeeded in generalizing the residual error estimator known from adaptive finite element methods to the exact BEM-based FEM, proved reliability of this new estimator, and provided numerical experiments which demonstrate optimal convergence for problems with solutions possessing singularities. Furthermore, in a recent article by Rjasanow and Weißer [103], the authors successfully generalized the method to quadratic trial functions as a first step towards high-order approximation using the BEM-based FEM.

From another point of view, rather than as a method of coupled BEM domains, one could consider the topical method as a FEM with piecewise PDE-harmonic trial functions as remarked above. A similar approach was, to our knowledge, first suggested by Trefftz in a 1926 paper [118] which we summarize in Section 1.3. A modern approach using this idea can be found in the multiscale FEM method, where we refer to, e.g., Babuška and Osborn [3], the book by Efendiev and Hou [32], and the book edited by Graham, Hou, Lakkis, and Scheichl [54].

As remarked in the previous section, the BEM-based FEM is also notable for the fact that it can treat generalized polygonal or polyhedral meshes. We therefore point out that other methods with this feature exist. One well-studied approach for this kind of problems is the family of so-called mimetic finite difference (MFD) methods. They are based on the construction of discrete spaces and operators which mimic properties of the continuous problem. MFD schemes for polygonal or polyhedral meshes have been investigated by Kuznetsov, Lipnikov, and Shashkov [77], Brezzi, Lipnikov, and Simoncini [19], and others. A convergence analysis has been provided by Brezzi, Lipnikov, and Shashkov [18]. The realization of these methods requires the construction of a mesh-dependent inner product on a space of discrete velocities. It is interesting that in the analysis, the authors hit upon several similar technical difficulties as we did and which stem from the lack of analytical tools for polyhedral elements.

Another approach that allows general meshes is the class of discontinuous Galerkin (DG) methods which have been intensively developed during the last decade; see [2] for an excellent overview of the field and a unified analysis of the most popular formulations.

DG methods usually retain the piecewise polynomial trial functions known from standard finite element methods, but drop the requirement that the trial functions are continuous across elements, instead introducing penalty terms to coerce the numerical solution to be close to continuity. This approach allows polyhedral element shapes since the coupling between degrees of freedom and mesh vertices does not have to be respected. However, most authors in the field restrict themselves to the common element shapes, one reason being presumably to avoid complications with approximation properties on non-standard element shapes. As an example for a DG method on polyhedral meshes (albeit for nonlinear convection-diffusion problems), we refer to the work by Dolejší, Feistauer, and Sobotíková [30]. We also point out that a DG approach generally necessitates the duplication of degrees of freedom across neighboring elements and thus a considerable increase in the number of unknowns.

We also mention the plane wave discontinuous Galerkin approach for the Helmholtz equation which combines ideas of the DG and Trefftz methods and which was analyzed in detail by Hiptmair and coworkers [50, 58, 89].

Publications on the BEM-based FEM and related topics which the present writer has co-authored or authored are [61, 59, 62, 63, 60].

It should be mentioned that the author has pursued a second line of research not represented in this thesis, loosely related by the use of harmonic basis functions, but focusing on interpolation using Radon projections. This topic has resulted in an interdisciplinary cooperation with members of the Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences, Todor Kableshkov University Sofia, and the Research Institute for Symbolic Computation, Linz. This work has resulted in a number of publications ([49, 47, 46, 48, 45]).

### 1.3 Trefftz's methods

The roots of the present work can be traced back, in some sense, to the Trefftz method introduced by E. Trefftz in an article entitled “Ein Gegenstück zum Ritzschen Verfahren” (“A counterpart to Ritz's Method”) [118] published in 1926. Due to its historic importance, we will briefly summarize the ideas introduced by Trefftz in this work, while using more modern notation. We will see that certain key concepts are shared by Trefftz's method and the method studied in the present work.

Trefftz considers a pure Dirichlet boundary value problem for the Laplace equation,  $-\Delta u = 0$ , in a domain  $\Omega$  with boundary  $\Gamma$ , with given boundary values  $g$ . He first describes Ritz's method, which W. Ritz had introduced previously in 1909 [100]: given a family of basis functions  $p_k$ ,  $k = 1, \dots, n$  which vanish on  $\Gamma$ , one makes the ansatz

$$v(x) = g(x) + \sum_{k=1}^n b_k p_k(x)$$

for an approximate solution. Here we assume that the boundary data  $g$  was extended in a suitable way into  $\Omega$ . The coefficients  $b_k$  are chosen such that the Dirichlet integral or  $H^1$ -energy is minimized, i.e.,

$$J(v) := \int_{\Omega} |\nabla v|^2 dx = |v|_{H^1(\Omega)}^2 \longrightarrow \min.$$

This is motivated by the fact that the exact solution  $u$  minimizes this energy functional  $J(u)$  over all admissible functions (in a modern setting,  $H^1(\Omega)$ ) which satisfy the Dirichlet boundary condition  $g$ . By the use of Green's identity, one shows that

$$\int_{\Omega} |\nabla v|^2 dx = \int_{\Omega} |\nabla u|^2 dx + \int_{\Omega} |\nabla(v - u)|^2 dx,$$

i.e.,  $J(v) = J(u) + J(u - v)$ . Therefore it is clear that minimizing  $J(v)$  is equivalent to minimizing the Dirichlet integral of the error,  $J(u - v)$ . From the above formula, since all integrals are non-negative, we also see that  $J(v) \geq J(u)$ . Hence, the Ritz method provides an upper bound for the Dirichlet integral of the exact solution  $u$ .

From a modern point of view, the most successful realization of Ritz's method is of course the finite element method. Indeed, choosing nodal basis functions defined over a standard finite element mesh for the functions  $p_k$ , one immediately obtains a finite element method, although it is today more commonly written as an equivalent variational problem instead of a minimization problem.

Trefftz points out that Ritz's method is based on the idea of approximating  $u$  by a function which matches the boundary data  $g$  exactly, but satisfies Laplace's equation only in an approximate way. The quality of this approximation is described by the Dirichlet error integral  $J(u - v)$ . He then proceeds to propose a counterpart to Ritz's method based on the opposite principle: satisfy the partial differential equation exactly, while approximating the boundary data. To this end, he introduces a set of harmonic basis functions  $q_k$ ,  $k = 1, \dots, n$ , and makes the ansatz

$$w(x) = \sum_{k=1}^n c_k q_k(x).$$

Again, the coefficients  $c_k$  are chosen in such a way that the error integral  $J(w - u)$  is minimized. This leads to the optimality condition  $\frac{\partial J(w-u)}{\partial c_\ell} = 0 \forall \ell$ . Since

$$\frac{\partial J(w - u)}{\partial c_\ell} = 2 \int_{\Omega} \frac{\partial}{\partial c_\ell} \nabla(w - u) \cdot \nabla(w - u) dx = 2 \int_{\Omega} \nabla q_\ell \cdot \nabla(w - u) dx$$

and, by Green's identity,

$$\int_{\Omega} \nabla q_\ell \cdot \nabla(w - u) dx = \int_{\Gamma} \frac{\partial q_\ell}{\partial n} (w - u) ds,$$

we arrive at the condition

$$\sum_{k=1}^n c_k \int_{\Gamma} q_k \frac{\partial q_{\ell}}{\partial n} ds = \int_{\Gamma} g(x) \frac{\partial q_{\ell}}{\partial n} ds \quad \forall \ell = 1, \dots, n. \quad (1.1)$$

We observe that  $\frac{\partial q_{\ell}}{\partial n}$  is nothing but the Neumann data belonging to the harmonic function  $q_{\ell}$ . In modern notation, we could denote it by

$$\frac{\partial q_{\ell}}{\partial n} = Sq_{\ell}$$

with the linear *Dirichlet-to-Neumann map*  $S$ . This allows us to rewrite (1.1) as

$$\langle w, Sq_{\ell} \rangle = \langle g, Sq_{\ell} \rangle \quad \forall \ell = 1, \dots, n, \quad (1.2)$$

where the angle brackets  $\langle \cdot, \cdot \rangle$  denote the integral over  $\Gamma$ . In a modern setting, their meaning can be generalized to the  $H^{-1/2}(\Gamma) \times H^{1/2}(\Gamma)$ -duality product. Variational equations of this type will play a central role in our work and will serve as starting points for the discretization in the BEM-based FEM. We stress in particular that these formulations live only on the boundary  $\Gamma$ .

Looking at the error integral  $J(w - u)$ , we see that

$$J(u) = J(w) - 2 \int_{\Omega} \nabla w \cdot \nabla(w - u) dx + J(w - u),$$

and since, by Green's identity,

$$\int_{\Omega} \nabla w \cdot \nabla(w - u) dx = \int_{\Gamma} (w - g) \frac{\partial w}{\partial n} ds = \langle w - g, Sw \rangle,$$

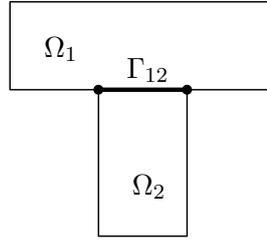
the middle term vanishes due to (1.2). We thus see that

$$J(w) \leq J(u) \leq J(v),$$

that is, Trefftz's method and Ritz's method provide lower and upper bounds for the Dirichlet integral  $J(u)$  of the exact solution, respectively.

It is notable that Trefftz already realized that his idea could be generalized to unions of domains in an early domain decomposition approach. This lays the foundation for the ideas of the BEM-based FEM which is the subject of the present work. Trefftz notes that "this generalization is only of practical significance for simple cases," but of course the computational resources at our disposal today allow us to treat significantly more complex situations.

Trefftz demonstrates his idea by the example of a T-shaped domain  $\Omega$  consisting of two rectangles  $\Omega_1$  and  $\Omega_2$  with the interface  $\Gamma_{12} = \overline{\Omega}_1 \cap \overline{\Omega}_2$ , see Figure 1.1. His premise is that, if the function values of  $u$  were known on  $\Gamma_{12}$ , then the solution  $u$  could be determined



**Figure 1.1:** The computational domain studied by Trefftz.

separately in  $\Omega_1$  and  $\Omega_2$  by solving Dirichlet boundary value problems using standard techniques. Since these values on the interface are not known, Trefftz makes the ansatz

$$v(x) = h_0 + \sum_{k=1}^n b_k \phi_k(x) \quad \forall x \in \Gamma_{12}$$

with suitable functions  $\{\phi_k\}_{k=1}^n$  defined on  $\Gamma_{12}$ . He then constructs piecewise harmonic functions  $\{q_k\}_{k=0}^n$  which satisfy the boundary conditions

- $q_0 = g$  on  $\Gamma$ ,  $q_0 = 0$  on  $\Gamma_{12}$ ,
- $q_k = 0$  on  $\Gamma$ ,  $q_k = \phi_k$  on  $\Gamma_{12}$  for  $k = 1, \dots, n$ .

We point out that Trefftz's choice of boundary conditions involves discontinuities and is thus prone to introducing artificial singularities in the approximate solution. With these basis functions, the ansatz for the solution is

$$v(x) = q_0(x) + \sum_{k=1}^n c_k q_k(x),$$

and again the coefficients  $c_k$  are chosen such that the Dirichlet integral  $J(v)$  is minimized. It is clear that again,  $J(v) \geq J(u)$ .

In the above approach, the basis functions  $q_k$  have continuous Dirichlet data, but discontinuous Neumann data, and we will use this principle in the construction of the BEM-based FEM in Chapter 4. However, Trefftz also describes a second approach where the opposite is true. Here, he observes that, if the normal derivative  $\frac{\partial u}{\partial n_1} = -\frac{\partial u}{\partial n_2}$  were known on  $\Gamma_{12}$ , where  $n_1$  and  $n_2$  denote the outward unit normal vector on  $\Gamma_1 = \partial\Omega_1$  and  $\Gamma_2 = \partial\Omega_2$ , respectively, then the solution  $u$  could be determined separately in  $\Omega_1$  and  $\Omega_2$  by solving a mixed boundary value problem with standard techniques. Since this normal derivative is not known, Trefftz makes the ansatz

$$\frac{\partial w}{\partial n_1}(x) = \sum_{k=1}^n c_k \psi_k(x) \quad \text{on } \Gamma_{12}$$

with suitable functions  $\psi_k$ ,  $k = 1, \dots, n$ , defined on  $\Gamma_{12}$ . The piecewise harmonic functions  $q_k$ ,  $k = 0, \dots, n$  are then defined by the mixed boundary value problems

- $q_0 = g$  on  $\Gamma$ ,  $\frac{\partial q_0}{\partial n_1} = 0$  on  $\Gamma_{12}$ ,
- $q_k = 0$  on  $\Gamma$ ,  $\frac{\partial q_k}{\partial n_1} = \psi_k$  on  $\Gamma_{12}$  for  $k = 1, \dots, n$ .

They are discontinuous across the interface  $\Gamma_{12}$ , but have continuous normal derivative there. For the solution, we obtain the ansatz

$$w(x) = q_0(x) + \sum_{k=1}^n c_k q_k(x)$$

and determine the coefficients  $c_k$  such that the error integral  $J(w - u)$  is minimized. Similarly to the single-domain case, this leads to the condition

$$\int_{\Gamma_{12}} (w|_{\Omega_1} - u) \frac{\partial q_\ell}{\partial n_1} ds + \int_{\Gamma_{12}} (w|_{\Omega_2} - u) \frac{\partial q_\ell}{\partial n_2} ds = 0 \quad \forall \ell = 1, \dots, n.$$

Since  $\frac{\partial q_\ell}{\partial n_1} = \psi_\ell = -\frac{\partial q_\ell}{\partial n_2}$  and  $u$  is continuous at  $\Gamma_{12}$ , we arrive at the condition

$$\int_{\Gamma_{12}} (w|_{\Omega_1} - w|_{\Omega_2}) \psi_\ell ds = 0 \quad \forall \ell = 1, \dots, n$$

which is easily translated to a condition on the coefficients  $c_k$ . By an argument analogous to the single-domain case, Trefftz shows that again  $J(w) \leq J(u)$ .

## 1.4 Outline

Concluding the introduction, we give an overview of the structure of the remainder of the work.

In Chapter 2, we introduce preliminaries which include basic results from functional analysis, variational problems, partial differential equations, and the finite element method.

Chapter 3 is dedicated to boundary integral operators, the boundary element method, and the Steklov-Poincaré operator as well as its approximation, all of which play a crucial role in the derivation of the BEM-based FEM. Besides known results of the standard theory, we give some relatively recent results which will allow us to obtain explicit bounds for constants associated with boundary integral operators later on.

In Chapter 4, we derive the BEM-based FEM in several formulations. We begin by deriving a skeletal variational formulation, as already hinted at above, and proceed to discretize it by choosing discrete skeletal spaces and replacing local Dirichlet-to-Neumann maps with their BEM-approximated analogues. In addition to the primal variational

formulation, we will also derive and discretize a mixed formulation where both Dirichlet and Neumann unknowns appear explicitly.

Some analytical tools for polyhedral elements will be provided in Chapter 5. As mentioned above, many of the standard FEM estimates were thus far not available for such non-standard element shapes, and therefore a large part of the analysis is concerned with filling these gaps.

With these tools at hand, in Chapter 6, we proceed to analyze the BEM-based FEM for the Laplace equation and derive first skeletal error estimates and then domain error estimates in the  $H^1$ - and  $L_2$ -norms. For the latter, the mixed variational formulation will prove useful since it admits a Galerkin orthogonality, which is not true for the primal formulation.

In Chapter 7, we discuss efficient solution methods for the systems of linear equations arising in the BEM-based FEM. In particular, we show that a class of domain decomposition methods, the finite element tearing/interconnecting-approach (FETI), can be successfully generalized to this setting.

Chapter 8 is dedicated to the convection-diffusion equation, and here in particular the case of convection-dominated problems. We discuss stabilized methods and observe that the BEM-based FEM is closely related to the so-called residual-free bubbles method, and thus in turn also to the streamline-upwind/Petrov-Galerkin (SUPG) scheme which has proven to be a successful stabilization technique.

In Chapter 9, we discuss some practical issues of the numerical realization of the BEM-based FEM, present some numerical experiments, and discuss the results.

The thesis is concluded with Chapter 10, where we also discuss possible further work.

# Chapter 2

## Preliminaries

In this chapter, we provide a range of fundamental results which will be used in the remainder of the thesis. In particular, we recall Sobolev spaces on domains and manifolds, functional analytic tools for abstract variational problems, Green's identities, existence and uniqueness results for boundary value problems for second order partial differential equations, and finally some standard theory for the finite element method. We provide references to the literature, but also give proofs wherever they are short enough and may yield insight.

### 2.1 Function spaces

We give the most important definitions relating to Sobolev spaces on domains and manifolds, where we focus mostly on the Hilbert space case. For more details, we refer the reader to, e.g., [1, 33, 106].

Let  $\Omega \subset \mathbb{R}^d$  be a bounded, connected Lipschitz domain with  $d = 2$  or  $3$ . As usual, by  $L_p(\Omega)$  with  $1 \leq p < \infty$ , we denote the Banach space of Lebesgue-measurable functions  $u$  on  $\Omega$  which satisfy

$$\|u\|_{L_p(\Omega)} := \left( \int_{\Omega} |u(x)|^p dx \right)^{1/p} < \infty.$$

The definition can be extended to  $p = \infty$  with the norm

$$\|u\|_{L_{\infty}(\Omega)} := \inf \{C \geq 0 : |u(x)| \leq C \forall x \in \Omega \text{ a.e.}\}.$$

A special role is taken by the space  $L_2(\Omega)$  of square integrable functions which can be shown to be a Hilbert space equipped with the inner product

$$\langle u, v \rangle_{L_2(\Omega)} = \int_{\Omega} uv dx$$

which induces the norm  $\|\cdot\|_{L_2(\Omega)}$ . Here and in all following definitions of Sobolev spaces, functions which have the same values almost everywhere in  $\Omega$  with respect to the Lebesgue measure are identified with each other. These definitions are thus, in a strict reading, to be understood as equivalence classes of such functions.

Let  $C^\infty(\Omega)$  denote the space of infinitely differentiable functions in  $\Omega$  and  $\mathcal{D}(\Omega) = C_0^\infty(\Omega)$  the subset of functions with compact support in  $\Omega$ . Furthermore, let  $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$  denote a multi-index and  $|\alpha| = \sum_{i=1}^d \alpha_i$ . For  $x \in \mathbb{R}^d$ , we write  $x^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$ . We call

$$D^\alpha \phi = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_d}} \phi \quad \forall \phi \in C^\infty(\Omega)$$

the  $\alpha$ -th derivative of a smooth function  $\phi$ . Let now  $\mathcal{D}'(\Omega)$  denote the space of bounded linear functionals  $u : C_0^\infty(\Omega) \rightarrow \mathbb{R}$ , where  $u$  is considered to be bounded if there exist constants  $C_{\alpha,\beta} > 0$  and an integer  $m \geq 0$  such that

$$|\langle u, \phi \rangle| \leq \sum_{|\alpha|, |\beta| \leq m} C_{\alpha,\beta} \sup_{x \in \Omega} |x^\alpha D^\beta \phi(x)| \quad \forall \phi \in C_0^\infty(\Omega),$$

where  $\alpha, \beta \in \mathbb{N}_0^d$  are multi-indices. We call  $\mathcal{D}'(\Omega)$  the space of *distributions* on  $\Omega$ .

The  $\alpha$ -th distributional derivative  $D^\alpha u \in \mathcal{D}'(\Omega)$  of a functional  $u \in \mathcal{D}'(\Omega)$  is defined as the functional

$$\langle D^\alpha u, \phi \rangle = (-1)^{|\alpha|} \langle u, D^\alpha \phi \rangle \quad \forall \phi \in C_0^\infty(\Omega). \quad (2.1)$$

Furthermore, we will use the differential operators

$$\begin{aligned} \partial_i \phi &= \frac{\partial \phi}{\partial x_i}, & \nabla \phi &= (\partial_1 \phi, \dots, \partial_d \phi), \\ \operatorname{div}(\phi_1, \dots, \phi_d) &= \sum_{i=1}^d \partial_i \phi_i, & \Delta \phi &= \operatorname{div}(\nabla \phi) = \sum_{i=1}^d \frac{\partial^2 \phi}{\partial x_i^2} \end{aligned}$$

and their distributional analogues defined analogously to (2.1). We will not distinguish between classical and distributional derivatives in the sequel.

For  $1 \leq p < \infty$  and a nonnegative integer  $k \in \mathbb{N}_0$  called the Sobolev index, the Sobolev space  $W_p^k(\Omega)$  is defined as

$$W_p^k(\Omega) = \{u \in L_p(\Omega) : D^\alpha u \in L_p(\Omega) \text{ for } |\alpha| \leq k\}.$$

and is equipped with the seminorm and norm

$$|u|_{W_p^k(\Omega)}^p = \sum_{|\alpha|=k} \|D^\alpha u\|_{L_p(\Omega)}^p, \quad \|u\|_{W_p^k(\Omega)}^p = \sum_{|\alpha| \leq k} \|D^\alpha u\|_{L_p(\Omega)}^p,$$

respectively. The extension to  $p = \infty$  is again possible. Again, we are mostly interested in the Hilbert space  $H^k(\Omega) := W_2^k(\Omega)$  with the inner product

$$\langle u, v \rangle_{H^k(\Omega)} = \sum_{|\alpha| \leq k} \langle D^\alpha u, D^\alpha v \rangle_{L_2(\Omega)}.$$

For non-integer Sobolev index  $s > 0$ , we let  $\lambda \in (0, 1)$  such that  $s = \lfloor s \rfloor + \lambda$  (where  $\lfloor s \rfloor$  is the largest integer less than or equal to  $s$ ) and define the inner product

$$\begin{aligned} \langle u, v \rangle_{H^s(\Omega)} &= \sum_{|\alpha| \leq \lfloor s \rfloor} \langle D^\alpha u, D^\alpha v \rangle_{L_2(\Omega)} \\ &\quad + \sum_{|\alpha| \leq \lfloor s \rfloor} \int_{\Omega \times \Omega} \frac{(D^\alpha u(x) - D^\alpha u(y))(D^\alpha v(x) - D^\alpha v(y))}{|x - y|^{d+2\lambda}} dx dy \end{aligned}$$

and corresponding norm

$$\|u\|_{H^s(\Omega)} = \sqrt{\langle u, u \rangle_{H^s(\Omega)}}.$$

The Hilbert space  $H^s(\Omega)$  is then given by the closure

$$H^s(\Omega) = \text{clos}_{H^s(\Omega)} \left\{ u \in C^\infty(\Omega) : \|u\|_{H^s(\Omega)} < \infty \right\}$$

with respect to this norm. The same relationship holds for spaces with integer  $s$  as defined above. In addition, we define, for any  $s > 0$ ,

$$H_0^s(\Omega) := \text{clos}_{H^s(\Omega)}(C_0^\infty(\Omega)).$$

We mention that Sobolev spaces of fractional index can equivalently be defined as interpolation spaces of spaces of integer index and refer to the literature for further details [8, 119].

We will also frequently use Sobolev spaces on manifolds and subsets of manifolds. Let  $\Gamma \subset \mathbb{R}^d$  denote a  $(d - 1)$ -dimensional manifold, or part of a manifold, embedded in  $\mathbb{R}^d$ . Analogously to above, we define  $L_2(\Gamma)$  as the space of functions which are square integrable with respect to the surface measure  $ds$  and equip it with the inner product and norm

$$\langle u, v \rangle_{L_2(\Gamma)} = \int_{\Gamma} uv ds \quad \text{and} \quad \|u\|_{L_2(\Gamma)} = \left( \int_{\Gamma} u^2 ds \right)^{1/2},$$

respectively. The same remark concerning identifying functions which match almost everywhere applies, and thus  $L_2(\Gamma)$  is again, in a strict reading, a space of equivalence classes.

As on domains, one can introduce Sobolev spaces  $H^s(\Gamma)$  on manifolds. The definition is however more technical as one needs to introduce an ‘‘atlas’’ of mappings from a reference patch which produce an overlapping covering of the surface  $\Gamma$ . Furthermore, the smoothness of the surface itself has to be taken into account. For a Lipschitz surface, Sobolev spaces  $H^s(\Gamma)$  with indices  $0 \leq s \leq 1$  can be introduced, while surfaces which are  $C^k$ -smooth allow spaces with indices  $0 \leq s \leq k$ . As in the case of Sobolev spaces on domains, spaces with higher index are embedded in those with lower indices, and spaces with non-integer index can be characterized by interpolation. For more details, we refer to the literature, e.g., [1, 119, 106].

It is sometimes of interest how Sobolev norms behave under uniform scaling of the domain  $\Omega$ . The following lemma gives some basic results which will be used in dilation arguments.

**Lemma 2.1.1.** *Let  $H := \text{diam } \Omega$  and define the scaled domain*

$$\hat{\Omega} := \left\{ \frac{1}{H}x : x \in \Omega \right\}$$

*such that  $\text{diam } \hat{\Omega} = 1$ . For any function  $v$  defined on  $\Omega$ , we define a scaled version  $\hat{v}$  on  $\hat{\Omega}$  by*

$$\hat{v}(\xi) := v(H\xi) = v(x) \quad \forall \xi \in \hat{\Omega}.$$

*Then we have, for functions  $v$  in  $L_2(\Omega)$  and  $H^1(\Omega)$ , respectively,*

$$\|v\|_{L_2(\Omega)} = H^{\frac{d}{2}} \|\hat{v}\|_{L_2(\hat{\Omega})}, \quad |v|_{H^1(\Omega)} = H^{\frac{d}{2}-1} |\hat{v}|_{H^1(\hat{\Omega})}$$

*and for functions  $v$  in  $L_2(\partial\Omega)$  and  $H^{1/2}(\partial\Omega)$ , respectively,*

$$\|v\|_{L_2(\partial\Omega)} = H^{\frac{d}{2}-\frac{1}{2}} \|\hat{v}\|_{L_2(\partial\hat{\Omega})}, \quad |v|_{H^{1/2}(\partial\Omega)} = H^{\frac{d}{2}-1} |\hat{v}|_{H^{1/2}(\partial\hat{\Omega})}.$$

*Proof.* Easily obtained by direct calculation.

For a continuous function  $\phi \in C(\bar{\Omega})$ , let  $\gamma_\Omega \phi \in C(\partial\Omega)$  denote the restriction of  $\phi$  to the boundary, i.e.,  $(\gamma_\Omega \phi)(x) = \phi(x) \forall x \in \partial\Omega$ . While functions from Sobolev spaces do not, in general, have well-defined point values, it can be shown that for  $\frac{1}{2} < s < \frac{3}{2}$ ,  $\gamma_\Omega$  can be extended to a linear, bounded operator

$$\gamma_\Omega = \gamma_\Omega^0 : H^s(\Omega) \rightarrow H^{s-1/2}(\partial\Omega)$$

called the (Sobolev) trace operator, or also Dirichlet trace operator in the following. We will often omit the superscript 0 when there is no chance of confusion.

With the Dirichlet trace operator,  $H_0^1(\Omega)$  can be characterized as

$$H_0^1(\Omega) = \left\{ v \in H^1(\Omega) : \gamma_\Omega^0 v = 0 \right\}.$$

**Theorem 2.1.2** (Friedrichs inequality). *Let  $\Gamma_D \subseteq \partial\Omega$  be a subset of  $\partial\Omega$  with positive surface measure. Then, for any function  $u \in H^1(\Omega)$  with*

$$(\gamma_\Omega u)|_{\Gamma_D} = 0,$$

*we have*

$$\|u\|_{L_2(\Omega)} \leq C_F |u|_{H^1(\Omega)}$$

*with a constant  $C_F = C_F(\Omega, \Gamma_D) > 0$  which depends only on  $\Omega$  and  $\Gamma_D$ .*

*Proof.* See, e.g., [10].

In particular, the above theorem establishes that  $|\cdot|_{H^1(\Omega)}$  and  $\|\cdot\|_{H^1(\Omega)}$  are equivalent norms on  $H_0^1(\Omega)$ , in which case we have  $C_F(\Omega) = C_F(\Omega, \partial\Omega)$  as a constant which depends only on  $\Omega$ . Similarly, such a norm equivalence holds on larger subspaces of  $H^1(\Omega)$  whose members vanish on a non-trivial part of the boundary.

A similar result holds for functions with zero mean. We give it here in a form which is explicit in the diameter of the domain  $\Omega$ .

**Theorem 2.1.3** (Poincaré inequality). *For any function  $u \in H^1(\Omega)$  which satisfies  $\int_{\Omega} u \, dx = 0$ , we have*

$$\|u\|_{L_2(\Omega)} \leq C_P \operatorname{diam}(\Omega) |u|_{H^1(\Omega)}$$

with a constant  $C_P = C_P(\Omega) > 0$  which depends only on the shape, but not the diameter of  $\Omega$ .

*Proof.* The classical formulation of the theorem,  $\|u\|_{L_2(\Omega)} \leq \tilde{C}_P |u|_{H^1(\Omega)}$ , is well-known and a proof can be found in, e.g., [10]. Using the dilation results from Lemma 2.1.1, we find that  $\tilde{C}_P = C_P \operatorname{diam}(\Omega)$  with a constant  $C_P$  which does not depend on  $\operatorname{diam}(\Omega)$ .  $\square$

## 2.2 Variational problems

In this section we collect some basic results on unique solvability and error estimates for variational problems in Hilbert spaces. Many of the proofs are short and elementary and will be given for completeness. For further details, we refer the reader to, e.g., [22, 106].

We first give the most fundamental result for unique solvability of variational problems.

**Lemma 2.2.1** (Lax-Milgram). *Let  $X$  be a Hilbert space with the norm  $\|\cdot\|$ , and let the bilinear form  $a(\cdot, \cdot) : X \times X \rightarrow \mathbb{R}$  be coercive and bounded, i.e.,*

$$\begin{aligned} \gamma \|v\|^2 &\leq a(v, v), & \forall v \in X, \\ |a(u, v)| &\leq \beta \|u\| \|v\| & \forall u, v \in X, \end{aligned}$$

with constants  $\beta, \gamma > 0$ . Then, for any bounded linear functional  $F : X \rightarrow \mathbb{R}$ , there exists a unique solution  $u \in X$  to the variational problem

$$a(u, v) = \langle F, v \rangle \quad \forall v \in X$$

and it satisfies the estimate

$$\|u\| \leq \frac{1}{\gamma} \|F\|_{X^*}.$$

*Proof.* The mapping from  $X$  to its dual  $X^*$  given by

$$A : X \rightarrow X^*, \quad (Au)(v) := a(u, v) \quad \forall u, v \in X$$

is easily seen to be linear and bounded by the assumptions, with  $\|A\| \leq \beta$ . Let  $\mathcal{J} : X^* \rightarrow X$  denote the Riesz isomorphism and introduce the operator

$$T : X \rightarrow X, \quad v \mapsto v - \rho(\mathcal{J}Av - \mathcal{J}F)$$

with a parameter  $\rho > 0$ . We show that  $T$  is a contraction if  $\rho$  is suitably chosen. Let  $v_1, v_2 \in X$  and  $v = v_1 - v_2$ . Then

$$\begin{aligned} \|T(v_1) - T(v_2)\|^2 &= \|v - \rho\mathcal{J}Av\|^2 = \|v\|^2 - 2\rho\langle\mathcal{J}Av, v\rangle + \rho^2\|\mathcal{J}Av\|^2 \\ &= \|v\|^2 - 2\rho a(v, v) + \rho^2 a(v, \mathcal{J}Av) \leq (1 - 2\rho\gamma + \rho^2\beta^2)\|v\|^2, \end{aligned}$$

and with  $\rho \in (0, 2\gamma/\beta^2)$  we obtain a contraction. From the Banach fixed point theorem (cf., e.g., [75]), it follows the existence of a unique fixed point  $u^* \in X$  of  $T(u^*) = u^*$ , which is at the same time the unique solution of  $Au^* = F$  and thus also of the variational problem.

The stability estimate follows from

$$\gamma\|u^*\|^2 \leq a(u^*, u^*) = \langle F, u^* \rangle \leq \|F\|_{X^*}\|u^*\|. \quad \square$$

We note that, in the case of a symmetric bilinear form, the above variational problem may be rewritten as a minimization problem.

**Theorem 2.2.2** (Ritz). *Let  $X$  be a Hilbert space with the norm  $\|\cdot\|$ , and let the bilinear form  $a(\cdot, \cdot) : X \times X \rightarrow \mathbb{R}$  be coercive and bounded as in Lemma 2.2.1, and additionally symmetric. Furthermore, let  $F : X \rightarrow \mathbb{R}$  be a bounded linear functional. Then the variational problem to find  $u \in X$  such that*

$$a(u, v) = \langle F, v \rangle \quad \forall v \in X$$

*is equivalent to the minimization problem: find  $u \in X$  which minimizes the functional*

$$J(u) := \frac{1}{2}a(u, u) - \langle F, u \rangle \longrightarrow \min.$$

*Proof.* The minimization problem is clearly equivalent to the statement

$$J(u + tv) \geq J(u) \quad \forall v \in X \quad \forall t > 0,$$

and expanding the left-hand side of the above equation, using bilinearity and symmetry of  $a(\cdot, \cdot)$ , and dividing by  $t$ , we obtain the equivalent formula

$$a(u, v) - \langle F, v \rangle \geq -\frac{1}{2}ta(v, v) \quad \forall v \in X \quad \forall t > 0.$$

If  $u$  satisfies the variational equation, then this formula holds true since then the left-hand side is 0 while the right-hand side is always non-positive, and thus  $u$  also solves the minimization problem. Conversely, if  $u$  solves the minimization problem, then by letting  $t \rightarrow 0$  above, we obtain  $\forall v \in X : a(u, v) - \langle F, v \rangle \geq 0$ . Replacing  $v$  by  $-v$ , we obtain  $\forall v \in X : a(u, v) - \langle F, v \rangle \leq 0$ , and thus the variational equation must be fulfilled.  $\square$

The functional  $J(u)$  in the above theorem is sometimes referred to as the energy functional.

Once unique solvability of a variational problem is established, we aim to compute approximations to its solution in discrete spaces. In the conforming case, where the discrete space is contained within the original Hilbert space and the bilinear form is evaluated exactly, error estimates can be obtained from the following result which states that the approximation to the exact solution in the discrete space is quasi-optimal up to a constant which depends on the properties of the bilinear form.

**Lemma 2.2.3** (Céa). *Let  $X_h \subset X$  be Hilbert spaces with the norm  $\|\cdot\|$ . Let the bilinear form  $a(\cdot, \cdot) : X \times X \rightarrow \mathbb{R}$  be coercive and bounded as in Lemma 2.2.1 with constants  $\beta$  and  $\gamma$ . Furthermore, let the linear functional  $F : X \rightarrow \mathbb{R}$  be bounded. Then the unique solutions  $u \in X$  and  $u_h \in X_h$ , respectively, of the variational problems*

$$\begin{aligned} a(u, v) &= \langle F, v \rangle & \forall v \in X, \\ a(u_h, v_h) &= \langle F, v_h \rangle & \forall v_h \in X_h \end{aligned}$$

satisfy the estimate

$$\|u - u_h\| \leq \frac{\beta}{\gamma} \inf_{v_h \in X_h} \|u - v_h\|.$$

*Proof.* By subtracting the variational equations, we obtain the Galerkin orthogonality

$$a(u - u_h, v_h) = 0 \quad \forall v_h \in X_h.$$

Therefore, we can estimate with an arbitrary  $v_h \in X_h$

$$\gamma \|u - u_h\|^2 \leq a(u - u_h, u - u_h) = a(u - u_h, u - v_h) \leq \beta \|u - u_h\| \|u - v_h\|,$$

and the statement follows after dividing by  $\gamma \|u - u_h\|$  and taking the infimum over all  $v_h \in X_h$ .  $\square$

In practice, the bilinear form  $a(\cdot, \cdot)$  may be impossible to evaluate exactly in a numerical scheme. In such cases, it can be substituted with a computable bilinear form  $\tilde{a}(\cdot, \cdot)$  which approximates  $a(\cdot, \cdot)$  in a certain sense. Similarly, the right-hand side  $F$  may need to be approximated. The following variant of the lemma of Strang will allow us to obtain error estimates also in this case, where consistency errors of the approximated bilinear form and right-hand side play an important role.

**Lemma 2.2.4** (Strang). *Let  $X_h \subset X$  be Hilbert spaces with the norm  $\|\cdot\|$ . Assume that there are constants  $c_1, c_2, \tilde{c}_1, \tilde{c}_2 > 0$  such that the bilinear forms  $a(\cdot, \cdot), \tilde{a}(\cdot, \cdot) : X \times X \rightarrow \mathbb{R}$  satisfy*

$$\begin{aligned} c_1 \|v\|^2 &\leq a(v, v), & \tilde{c}_1 \|v\|^2 &\leq \tilde{a}(v, v) & \forall v \in X, \\ |a(v, w)| &\leq c_2 \|v\| \|w\|, & |\tilde{a}(v, w)| &\leq \tilde{c}_2 \|v\| \|w\| & \forall v, w \in X. \end{aligned}$$

Let  $F, \tilde{F} \in X^*$  be bounded linear functionals and assume that  $u \in X$  and  $u_h \in X_h$  solve

$$\begin{aligned} a(u, v) &= \langle F, v \rangle & \forall v \in X, \\ \tilde{a}(u_h, v_h) &= \langle \tilde{F}, v_h \rangle & \forall v_h \in X_h. \end{aligned}$$

Then

$$\|u - u_h\| \leq C \left( \inf_{v_h \in X_h} \|u - v_h\| + \sup_{w_h \in X_h} \frac{|\tilde{a}(u, w_h) - \langle \tilde{F}, w_h \rangle|}{\|w_h\|} \right),$$

where  $C = \max \left\{ 1 + \frac{\tilde{c}_2}{c_1}, \frac{1}{c_1} \right\}$ .

*Proof.* For an arbitrary  $v_h \in X_h$ , we have

$$\begin{aligned} \tilde{\gamma}_1 \|v_h - u_h\|^2 &\leq \tilde{a}(v_h - u_h, v_h - u_h) \\ &= \tilde{a}(v_h - u, v_h - u_h) + \tilde{a}(u, v_h - u_h) - \langle \tilde{F}, v_h - u_h \rangle \\ &\leq \tilde{\gamma}_2 \|v_h - u\| \|v_h - u_h\| + |\tilde{a}(u, v_h - u_h) - \langle \tilde{F}, v_h - u_h \rangle|. \end{aligned}$$

Dividing by  $\tilde{\gamma}_1 \|v_h - u_h\|$  yields

$$\begin{aligned} \|v_h - u_h\| &\leq \frac{\tilde{\gamma}_2}{\tilde{\gamma}_1} \|v_h - u\| + \frac{1}{\tilde{\gamma}_1} \frac{|\tilde{a}(u, v_h - u_h) - \langle \tilde{F}, v_h - u_h \rangle|}{\|v_h - u_h\|} \\ &\leq \frac{\tilde{\gamma}_2}{\tilde{\gamma}_1} \|v_h - u\| + \frac{1}{\tilde{\gamma}_1} \sup_{w_h \in X_h} \frac{|\tilde{a}(u, w_h) - \langle \tilde{F}, w_h \rangle|}{\|w_h\|}. \end{aligned}$$

Use of the triangle inequality and of the above estimate gives us

$$\begin{aligned} \|u - u_h\| &\leq \|u - v_h\| + \|v_h - u_h\| \\ &\leq \left( 1 + \frac{\tilde{\gamma}_2}{\tilde{\gamma}_1} \right) \|u - v_h\| + \frac{1}{\tilde{\gamma}_1} \sup_{w_h \in X_h} \frac{|\tilde{a}(u, w_h) - \langle \tilde{F}, w_h \rangle|}{\|w_h\|}. \end{aligned}$$

Taking the infimum over all  $v_h \in X_h$  finishes the proof.  $\square$

*Remark.* The above Strang lemma is adapted to the situation we will encounter in the analysis of the BEM-based FEM. It can be made more general, both by weakening the assumptions (for instance, the approximated bilinear form  $\tilde{a}(\cdot, \cdot)$  and linear functional  $\tilde{F}$  need only be defined on the discrete spaces) and by allowing the case of a non-conforming discrete space  $X_h \not\subset X$ . We refer to Ciarlet [22] for further details on such so-called “variational crimes.”

## 2.3 Green's identities

Green's identities are an important tool in the treatment of partial differential equations. They stem from the divergence theorem, also called Gauss theorem or Ostrogradsky's theorem.

**Theorem 2.3.1** (Divergence theorem). *Let  $\Omega \subset \mathbb{R}^d$  be a Lipschitz domain. For a vector function  $F \in [C^1(\Omega)]^d$  or  $F \in [H^1(\Omega)]^d$ , we have*

$$\int_{\Omega} \operatorname{div} F \, dx = \int_{\partial\Omega} F \cdot n \, ds,$$

where  $n(x)$  is the outward unit normal vector on  $\partial\Omega$ .

From this theorem and suitable closure properties, many Green's type identities can be derived. We give the following important theorem, often referred to as the first and second Green's identities, but will often invoke an argument of this type simply by mentioning "integration by parts."

**Theorem 2.3.2** (Green's identities [87]). *Let  $\Omega \subset \mathbb{R}^d$  be a Lipschitz domain, and let  $A : \Omega \rightarrow \mathbb{R}^{d \times d}$  a matrix-valued, Lipschitz-continuous function. Then, for  $u \in H^2(\Omega)$  and  $v \in H^1(\Omega)$ , we have*

$$\int_{\Omega} \operatorname{div}(A\nabla u)v \, dx = - \int_{\Omega} A\nabla u \cdot \nabla v \, dx + \int_{\partial\Omega} (A\nabla u \cdot n)v \, ds_x.$$

For  $u \in H^2(\Omega)$  and  $v \in H^2(\Omega)$ , we have

$$\begin{aligned} \int_{\Omega} \operatorname{div}(A\nabla u)v \, dx - \int_{\Omega} u \operatorname{div}(A\nabla v) \, dx &= \\ &= \int_{\partial\Omega} (A\nabla u \cdot n)v \, ds_x - \int_{\partial\Omega} (A\nabla v \cdot n)u \, ds_x. \end{aligned}$$

## 2.4 Boundary value problems

We consider partial differential operators of the form

$$Lu(x) = -\operatorname{div}(A(x)\nabla u(x)) + 2b(x) \cdot \nabla u(x) + c(x)u(x) \quad (2.2)$$

with bounded coefficient functions  $A(x) \in \mathbb{R}^{d \times d}$ ,  $b(x) \in \mathbb{R}^d$ , and  $c(x) \in \mathbb{R}$ . Generally, we will assume  $A(x)$  to be symmetric and uniformly positive definite. The associated Dirichlet boundary value problem is to determine a function  $u$  such that

$$\begin{aligned} Lu(x) &= f(x), & x &\in \Omega, \\ \gamma_{\Omega}u(x) &= g(x), & x &\in \partial\Omega, \end{aligned} \quad (2.3)$$

where  $g \in H^{1/2}(\partial\Omega)$  is the given Dirichlet data, and  $f \in L_2(\Omega)$  is a given right-hand side. As a notable special case, we will often consider the problem  $A = I$ ,  $b = 0$ ,  $c = 0$ , which results in  $L = -\Delta$  being the Laplace operator, and the problem (2.3) is then referred to as the Poisson equation or, if  $f = 0$ , the Laplace equation. Also, if  $A(x) = \alpha(x)I$ ,  $b = 0$ ,  $c = 0$ , we refer to the problem as the diffusion equation with diffusion coefficient  $\alpha(x) > 0$ .

In the modern treatment of such partial differential equations, one usually passes to a variational equation. The equation is multiplied by a test function  $v \in H_0^1(\Omega)$ , and after applying Theorem 2.3.2 (Green's identity), the standard variational formulation of the above boundary value problem reads as follows: find  $u \in H^1(\Omega)$  such that

$$\gamma_\Omega u = g, \quad \mathcal{L}(u, v) = \langle F, v \rangle \quad \forall v \in H_0^1(\Omega), \quad (2.4)$$

with the bilinear form  $\mathcal{L}$  and the linear functional  $F$  given by

$$\begin{aligned} \mathcal{L}(u, v) &= \int_\Omega (A \nabla u \cdot \nabla v + 2b \cdot \nabla u v + cvv) \, dx, \\ \langle F, v \rangle &= \int_\Omega f v \, dx, \end{aligned}$$

The relation of (2.4) to the classical boundary value problem (2.3) is not trivial. We do point out that the solution  $u$  of the variational problem (2.4) is equivalently the unique solution of the distributional equation

$$u \in H^1(\Omega) : \quad \gamma_\Omega u = g \quad \text{and} \quad Lu = f \text{ in } \mathcal{D}'(\Omega),$$

cf. [22]. Under regularity assumptions on the domain  $\Omega$  and the given data, smoothness of  $u$  can be shown, for instance  $u \in C^2(\Omega) \cap C^0(\bar{\Omega})$ , such that the partial differential equation (2.3) holds.

Using the tools from Section 2.2, one can show that the chosen functional spaces make sense in that they yield a uniquely solvable variational problem under certain assumptions imposed on the coefficients.

**Theorem 2.4.1.** *Let  $f \in L_2(\Omega)$ ,  $b \in [L_\infty(\Omega)]^d$ ,  $c \in L_\infty(\Omega)$ , and  $A(x) \in [L_\infty(\Omega)]^{d \times d}$  which satisfies, for some  $\alpha_0 > 0$  and  $\bar{\alpha} > 0$ ,*

$$\begin{aligned} \alpha_0 |\xi|^2 &\leq (A(x)\xi) \cdot \xi & \forall x \in \Omega \text{ a.e.} \quad \forall \xi \in \mathbb{R}^d, \\ \|A(x)\|_{\ell_2} &\leq \bar{\alpha} & \forall x \in \Omega \text{ a.e.} \end{aligned}$$

Furthermore, assume that  $\operatorname{div} b \in L_\infty(\Omega)$  and

$$-\operatorname{div} b(x) + c(x) \geq 0 \quad \forall x \in \Omega \text{ a.e.}$$

Then the variational problem (2.4) has a unique solution.

*Proof.* After homogenization with a suitable extension  $\tilde{g} \in H^1(\Omega)$  of  $g$ , (2.4) has the form: find  $u_0 \in H_0^1(\Omega)$  such that

$$\mathcal{L}(u_0, v) = \langle F, v \rangle - \mathcal{L}(\tilde{g}, v) \quad \forall v \in H_0^1(\Omega).$$

Using the Cauchy-Schwarz inequality, we obtain the upper bound

$$\begin{aligned} |\mathcal{L}(u, v)| &\leq \bar{\alpha} |u|_{H^1(\Omega)} |v|_{H^1(\Omega)} \\ &\quad + 2 \|b\|_\infty |u|_{H^1(\Omega)} \|v\|_{L_2(\Omega)} + \|c\|_\infty \|u\|_{L_2(\Omega)} \|v\|_{L_2(\Omega)} \\ &\leq (\bar{\alpha} + 2 \|b\|_\infty + \|c\|_\infty) \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)}. \end{aligned}$$

For the coercivity, we observe using the divergence theorem, Theorem 2.3.1, that

$$\int_{\Omega} (2b \cdot \nabla v) v \, dx = \int_{\Omega} b \cdot \nabla (v^2) \, dx = - \int_{\Omega} (\operatorname{div} b) v^2 \, dx.$$

Therefore, we have

$$\begin{aligned} \mathcal{L}(v, v) &= \int_{\Omega} \left[ A(x) \nabla v(x) \cdot \nabla v(x) + (-\operatorname{div} b(x) + c(x)) v(x)^2 \right] dx \\ &\geq \alpha_0 |v|_{H^1(\Omega)}^2 \geq \frac{\alpha_0}{C_F(\Omega)^2} \|v\|_{H^1(\Omega)}^2, \end{aligned}$$

where we used the Friedrichs inequality, Theorem 2.1.2, in the last estimate. The statement follows by applying Lax-Milgram (Lemma 2.2.1).  $\square$

*Remark.* We point out that we had to impose an additional regularity condition on  $b$ , namely that  $\operatorname{div} b$  exists, as well as the condition  $-\operatorname{div} b + c \geq 0$ , in order to preserve coercivity of the bilinear form in the above theorem. Indeed, it seems that conditions on either regularity or “smallness” of the convective term cannot be avoided in proofs of unique solvability which use the classic theory for coercive partial differential equations. We point to a relatively recent paper by Droniou [31] where these restrictions are removed and unique solvability of the variational problem (2.4) is shown under quite general assumptions. In particular, apart from standard coercivity assumptions on  $A$ , the convective term  $b$  is only assumed to lie in the Lebesgue space  $[L_{d_*}(\Omega)]^d$ , where  $d_* = 3$  if  $d \geq 3$  and  $d_* \in (2, \infty)$  if  $d = 2$ , and the reactive term  $c$  is assumed to lie in  $L_{d_*/2}(\Omega)$  and be nonnegative almost everywhere. These integrability conditions are, due to the Sobolev embeddings, the weakest possible to ensure that all terms in the bilinear form are well-defined and are thus not a real restriction. The proof technique abandons coercivity and the Lax-Milgram lemma and instead employs the Leray-Schauder topological degree.

For pure Dirichlet boundary value problems, the solution operator which maps given Dirichlet data  $g$  on the boundary  $\partial\Omega$  to the solution  $u$  of (2.4) with a zero right-hand side  $F \equiv 0$  is of particular importance.

**Definition 2.4.1.** Assume that the variational problem (2.4) is uniquely solvable. Then we have the solution operator

$$\mathcal{H} = \mathcal{H}_\Omega : H^{1/2}(\partial\Omega) \rightarrow H^1(\Omega)$$

which maps a given  $g \in H^{1/2}(\partial\Omega)$  to the unique solution  $\mathcal{H}g \in H^1(\Omega)$  of the variational problem

$$\gamma_\Omega(\mathcal{H}g) = g, \quad \mathcal{L}(\mathcal{H}g, v) = 0 \quad \forall v \in H_0^1(\Omega).$$

We call this operator the  $\mathcal{L}$ -harmonic extension operator from  $H^{1/2}(\partial\Omega)$  to  $H^1(\Omega)$ . In particular, if  $L = -\Delta$  is the Laplace operator,  $\mathcal{H}$  is just the harmonic extension operator.

Clearly, the operator  $\mathcal{H}$  maps to the space  $\mathcal{H}(\Omega)$  of  $\mathcal{L}$ -harmonic functions on  $\Omega$  defined by

$$\mathcal{H}(\Omega) := \left\{ u \in H^1(\Omega) : \mathcal{L}(u, v) = 0 \quad \forall v \in H_0^1(\Omega) \right\} \subset H^1(\Omega), \quad (2.5)$$

and indeed it is a bijection between  $H^{1/2}(\partial\Omega)$  and  $\mathcal{H}(\Omega)$ , with its inverse given by the trace operator  $\gamma_\Omega$ .

Applying Theorem 2.2.2 to this variational problem in the special case of the Laplace operator,  $L = -\Delta$ , we obtain an important energy minimization result for the harmonic extension operator.

**Theorem 2.4.2.** Let  $L = -\Delta$ . Then, for all  $g \in H^{1/2}(\partial\Omega)$ , we have

$$|\mathcal{H}g|_{H^1(\Omega)} = \inf_{\substack{v \in H^1(\Omega) \\ \gamma_\Omega v = g}} |v|_{H^1(\Omega)}. \quad (2.6)$$

*Proof.* Follows directly from the definition of  $\mathcal{H}$  and Theorem 2.2.2 by noting that  $\mathcal{L}(v, v) = |v|_{H^1(\Omega)}^2$ .  $\square$

We now introduce a generalized normal derivative operator. Note that the straightforward definition  $A\gamma_\Omega(\nabla v) \cdot n$  applies only if  $v$  is sufficiently regular, that is,  $v \in H^s(\Omega)$  with  $s > \frac{3}{2}$  such that the Dirichlet trace of  $\nabla v$  is well-defined. If, however,  $v \in H^1(\Omega)$  is the weak solution of a partial differential equation, then a generalized conormal derivative, which is in general a functional, can still be defined as follows.

**Definition 2.4.2.** Following McLean [87, Lemma 4.3], we define the *Neumann trace operator*  $\gamma^1 = \gamma_\Omega^1 : \mathcal{H}(\Omega) \rightarrow H^{-1/2}(\partial\Omega)$  by the relation

$$\langle \gamma_\Omega^1 u, w \rangle_{\partial\Omega} = \mathcal{L}(u, \tilde{w}) \quad \forall w \in H^{1/2}(\partial\Omega),$$

where  $\tilde{w} \in H^1(\Omega)$  is an arbitrary extension of  $w$  into  $\Omega$  and  $\langle \cdot, \cdot \rangle_{\partial\Omega}$  denotes the duality product between  $H^{-1/2}(\partial\Omega)$  and  $H^{1/2}(\partial\Omega)$ .

It follows from the definition of  $\mathcal{H}(\Omega)$  that the Neumann trace  $\gamma_\Omega^1 u$  does not depend on the actual choice of  $\tilde{w}$  and is thus well-defined. In other words, we have for any  $u \in \mathcal{H}(\Omega)$

$$\langle \gamma_\Omega^1 u, \gamma_\Omega^0 v \rangle_{\partial\Omega} = \mathcal{L}(u, v) \quad \forall v \in H^1(\Omega). \quad (2.7)$$

This is, in essence, Green's first identity for  $\mathcal{L}$ -harmonic functions. In case of sufficient regularity, e.g.,  $u \in H^2(\Omega)$ , it follows by an application of Theorem 2.3.2 that  $\gamma_\Omega^1 u = A\nabla u \cdot n$  and thus the generalized Neumann trace operator coincides with the classical definition of the conormal derivative in this case.

Composing the  $\mathcal{H}$ -harmonic extension operator with the Neumann trace operator, we obtain the *Dirichlet-to-Neumann map*

$$\begin{aligned} S_\Omega = S : H^{1/2}(\partial\Omega) &\rightarrow H^{-1/2}(\partial\Omega) \\ v &\mapsto \gamma_\Omega^1(\mathcal{H}_\Omega v), \end{aligned} \quad (2.8)$$

which maps some given Dirichlet data  $v \in H^{1/2}(\partial\Omega)$  to the Neumann data of the solution of the corresponding homogeneous partial differential equation.

In the following, we derive the adjoint of the Dirichlet-to-Neumann map. For this, let  $\Phi \in \mathcal{H}(\Omega)$  be a  $\mathcal{L}$ -harmonic function with traces  $u = \gamma^0 \Phi$  and  $t = \gamma^1 \Phi$ . Furthermore, let  $\Psi \in H^1(\Omega)$  be a solution of the adjoint variational problem

$$\mathcal{L}(w, \Psi) = 0 \quad \forall w \in H_0^1(\Omega)$$

and set  $\phi := \gamma^0 \Psi \in H^{1/2}(\partial\Omega)$ . By integration by parts, it is easy to see that  $\Psi$  is the weak solution of the adjoint boundary value problem

$$\begin{aligned} L^* \Psi &:= -\operatorname{div}(A^\top \nabla \Psi) - \operatorname{div}(2b\Psi) + c\Psi = 0 && \text{in } \Omega, \\ \Psi &= \phi && \text{on } \partial\Omega. \end{aligned}$$

In analogy to the definition of the Neumann trace operator  $\gamma^1$ , the generalized Neumann trace  $\psi := \tilde{\gamma}^1 \Psi \in H^{-1/2}(\partial\Omega)$  of  $\Psi$  is given as the functional

$$\langle \tilde{\gamma}^1 \Psi, \xi \rangle := \mathcal{L}(\tilde{\xi}, \Psi) \quad \forall \xi \in H^{1/2}(\partial\Omega), \quad (2.9)$$

where again  $\tilde{\xi} \in H^1(\Omega)$  is an extension. For sufficiently regular  $\Psi$ , using integration by parts, one shows that

$$\psi = \tilde{\gamma}^1 \Psi = \gamma^0(A^\top \nabla \Psi) \cdot n + 2\langle b, n \rangle \gamma^0 \Psi. \quad (2.10)$$

By (2.7) and (2.9), we have

$$\langle Su, \phi \rangle = \langle t, \gamma^0 \Psi \rangle = \mathcal{L}(\Phi, \Psi) = \langle \psi, \gamma^0 \Phi \rangle = \langle \psi, u \rangle,$$

and thus the adjoint operator to the Dirichlet-to-Neumann map  $S$  is given by  $S^* \phi = \psi = \tilde{\gamma}^1 \Psi = \tilde{\gamma}^1 \mathcal{H}^* \phi$  with  $\mathcal{H}^*$  the solution operator for the adjoint boundary value problem. In particular, if  $A$  is symmetric and  $b = 0$ , the bilinear form  $\mathcal{L}(\cdot, \cdot)$  is symmetric and thus  $S$  is self-adjoint.

## 2.5 The finite element method

Let us consider a boundary value problem of the type (2.3) discussed in Section 2.4, with the associated variational formulation (2.4). The finite element method (FEM) is a well-established approach, both in theory and in practice, for approximating solutions of such problems numerically. In the following, we outline the standard theory for the simple case of the linear Courant element. More comprehensive treatment may be found in the literature, e.g., [22].

For simplicity, we assume that  $\Omega$  is a polygonal or polyhedral Lipschitz domain. Let  $\Xi$  be a collection of  $d$ -dimensional open simplices  $\tau \in \Xi$  such that

$$\bar{\Omega} = \bigcup_{\tau \in \Xi} \bar{\tau}, \quad \tau_1 \cap \tau_2 = \emptyset \quad \forall \tau_1 \neq \tau_2 \text{ from } \Xi.$$

Thus,  $\Xi$  consists of mutually disjoint triangles or tetrahedra in 2D or 3D, respectively, which cover the domain  $\Omega$ . We call  $\Xi$  a (*simplicial*) *mesh* or *triangulation* (regardless of the actual space dimension) of the computational domain  $\Omega$ . We assume the mesh  $\Xi$  to be *conforming*, that is, for any two elements  $\tau_1, \tau_2 \in \Xi$ , the intersection  $\bar{\tau}_1 \cap \bar{\tau}_2$  is either empty, an element vertex, an element edge, or (in 3D) a face of both elements.

We set up the discrete trial spaces

$$\begin{aligned} \mathcal{V}_h &:= \mathcal{V}_h(\Omega) := \{v \in C(\bar{\Omega}) : v|_{\tau} \in P^1(\tau) \quad \forall \tau \in \Xi\} \subset H^1(\Omega), \\ \mathcal{V}_{h,0} &:= \mathcal{V}_{h,0}(\Omega) := \mathcal{V}_h \cap H_0^1(\Omega) \end{aligned}$$

of continuous, piecewise linear functions. Here,  $P^1(\tau)$  refers to the space of affine linear functions on the simplex  $\tau$ , that is, of polynomials with total degree at most 1. The Galerkin discretization of (2.4) then takes the form: find  $u_h \in \mathcal{V}_h$  such that

$$\gamma_{\Omega} u_h = g, \quad \mathcal{L}(u_h, v_h) = \langle F, v_h \rangle \quad \forall v_h \in \mathcal{V}_{h,0}. \quad (2.11)$$

Clearly, in order to be solvable, this problem requires the given Dirichlet data  $g$  to be piecewise linear with respect to the boundary triangulation induced by  $\Xi$ . We write  $\mathcal{V}_h(\partial\Omega)$  for the space of such piecewise linear boundary functions and will for simplicity always assume  $g \in \mathcal{V}_h(\partial\Omega)$  in the following. More general cases can be handled by, for example, interpolating  $g$  in  $\mathcal{V}_h(\partial\Omega)$  if it is continuous, or projecting  $g$  to  $\mathcal{V}_h(\partial\Omega)$  if it is in  $L_2(\partial\Omega)$ .

Often, for the analysis, it is convenient to work with the homogenized formulation. For this, one takes an extension of  $g$  to the entire FEM space  $\mathcal{V}_{h,0}$ , which exists due to the assumption of  $g$  being piecewise linear. For simplicity, we denote the extension again by  $g$ . By setting  $u_h = u_{h,0} + g$ , one can solve for  $u_{h,0} \in \mathcal{V}_{h,0}$  with

$$\mathcal{L}(u_{h,0}, v_h) = \langle F, v_h \rangle - \mathcal{L}(g, v_h) =: \langle F_0, v_h \rangle \quad \forall v_h \in \mathcal{V}_{h,0}. \quad (2.12)$$

It is easy to see that this problem is equivalent to (2.11).

**Theorem 2.5.1.** *Under the assumptions of Theorem 2.4.1, the finite element equation (2.11) has a unique solution  $u_h$ , and it satisfies the quasi-optimality estimate*

$$\|u - u_h\|_{H^1(\Omega)} \leq C \inf_{\substack{v_h \in \mathcal{V}_h \\ v_h|_{\partial\Omega} = g}} \|u - v_h\|_{H^1(\Omega)},$$

where  $u \in H^1(\Omega)$  is the unique solution of (2.4) and the constant  $C$  depends on the coefficients of the partial differential operator  $L$  and on the domain  $\Omega$ .

*Proof.* The boundedness and coercivity of  $\mathcal{L}(\cdot, \cdot)$  from the proof of Theorem 2.4.1 carry over directly since  $\mathcal{V}_{h,0} \subset H_0^1(\Omega)$ , and thus Lemma 2.2.1 (Lax-Milgram) gives unique solvability of (2.12), and thus also (2.11). The error estimate follows from Lemma 2.2.3 (Céa).  $\square$

In order to obtain an explicit *a priori* error estimate, we thus need to know the approximation properties of the finite element space. To ensure good approximation, certain regularity assumptions on the mesh  $\Xi$  need to be made. Usually, this is either done by requiring bounds on the interior angles of the element  $\tau \in \Xi$ , or by viewing the elements as transformations of a reference element and imposing constraints on these transformations. Typically, the resulting definitions are equivalent. We take the latter route and write

$$\begin{aligned} \Delta_d &:= \{(x_1, \dots, x_d)^\top \in \mathbb{R}^d : x_i > 0, x_1 + \dots + x_d < 1\} \\ &= \text{conv}^\circ\{0, e_1, \dots, e_d\} \end{aligned}$$

for the unit simplex in  $\mathbb{R}^d$ , where  $\text{conv}^\circ$  denotes the interior of the convex hull of the given set and  $e_j \in \mathbb{R}^d$  is the unit vector with 1 at the  $j$ -th coordinate.

For any element  $\tau \in \Xi$ , we fix an affine linear mapping  $F_\tau : \mathbb{R}^d \rightarrow \mathbb{R}^d$  such that  $F_\tau(\Delta_d) = \tau$  and denote its Jacobian by  $J_\tau = \nabla F_\tau \in \mathbb{R}^{d \times d}$ .

**Definition 2.5.1.** A simplicial mesh  $\Xi$  is called *shape-regular* if and only if there exist positive constants  $\underline{c}_1$ ,  $\bar{c}_1$ ,  $\underline{c}_2$ , and  $\bar{c}_2$  such that for all elements  $\tau \in \Xi$ , we have

$$\underline{c}_1 (\text{diam } \tau)^d \leq |\det J_\tau| \leq \bar{c}_1 (\text{diam } \tau)^d, \quad (2.13)$$

$$\|J_\tau\|_{\ell_2} \leq \bar{c}_2 \text{diam } \tau, \quad (2.14)$$

$$\|J_\tau^{-1}\|_{\ell_2} \leq (\underline{c}_2 \text{diam } \tau)^{-1}, \quad (2.15)$$

where  $J_\tau$  is the Jacobian of the affine linear mapping  $F_\tau$  from the unit simplex  $\Delta_d$  to  $\tau$ , and  $\|A\|_{\ell_2} = \sqrt{\lambda_{\max}(A^\top A)}$  denotes the spectral matrix norm.

In the following, let  $h_\tau := \text{diam } \tau$  denote the diameter of an element  $\tau \in \Xi$  and  $h := \max_{\tau \in \Xi} h_\tau$  the mesh size of  $\Xi$ , that is, the maximum element diameter.

**Theorem 2.5.2** (Scott and Zhang [109]; cf. [10]). *Let  $\Xi$  be a shape-regular, conforming mesh of  $\Omega$ . Then there exists a quasi-interpolation operator  $\Pi : H^1(\Omega) \rightarrow \mathcal{V}_h$  which preserves piecewise linear boundary data,*

$$(\Pi u)|_{\partial\Omega} = u|_{\partial\Omega} \quad \text{for all } u \in H^1(\Omega) \text{ with } u|_{\partial\Omega} \in \mathcal{V}_h(\partial\Omega),$$

and satisfies the estimates

$$\begin{aligned} |\Pi u|_{H^1(\Omega)} &\leq c_\Pi |u|_{H^1(\Omega)} & \forall u \in H^1(\Omega), \\ \left( \sum_{\tau \in \Xi} h_\tau^{2(\ell-k)} \|u - \Pi u\|_{H^\ell(\tau)}^2 \right)^{1/2} &\leq c_\Pi |u|_{H^k(\Omega)} & \forall u \in H^k(\Omega), \end{aligned}$$

where  $0 \leq \ell \leq k \leq 2$  and the constant  $c_\Pi > 0$  depends only on the shape regularity of  $\Xi$ .

In particular, the second estimate in the above theorem implies

$$\|u - \Pi u\|_{H^\ell(\Omega)} \leq c_\Pi h^{k-\ell} |u|_{H^k(\Omega)} \quad \forall u \in H^k(\Omega)$$

for  $\ell \in \{0, 1\}$ .

**Corollary 2.5.3.** *Assume that the mesh  $\Xi$  of  $\Omega$  is shape-regular and conforming. Then for any function  $u \in H^{1+s}(\Omega)$ ,  $s \in [0, 1]$ , with piecewise linear boundary data, we have*

$$\inf_{\substack{v_h \in \mathcal{V}_h \\ v_h|_{\partial\Omega} = u|_{\partial\Omega}}} \|u - v_h\|_{H^1(\Omega)} \leq Ch^s |u|_{H^{1+s}(\Omega)},$$

where the constant  $C$  depends only on the regularity parameters of  $\Xi$ .

*Proof.* Follows immediately with the choice  $v_h = \Pi u$  and Theorem 2.5.2.  $\square$

From Theorem 2.5.1 and the above result follows the standard  $H^1$ -error estimate for the Courant finite element method. In particular, the following theorem states that for a sequence  $\{\Xi_h\}$  of uniformly shape-regular meshes with  $h \rightarrow 0$ , the finite element approximation  $u_h$  converges to  $u$ , and for a fully regular solution it does so with a rate of  $\mathcal{O}(h)$  in the  $H^1$ -norm.

**Theorem 2.5.4.** *Assume that the exact solution  $u$  of (2.4) satisfies  $u \in H^{1+s}(\Omega)$ ,  $s \in [0, 1]$ , and that the mesh  $\Xi$  satisfies the regularity assumptions of Theorem 2.5.2. Then the FEM approximation  $u_h$  from (2.11) satisfies*

$$\|u - u_h\|_{H^1(\Omega)} \leq Ch^s |u|_{H^{1+s}(\Omega)},$$

where the constant  $C$  depends on the coefficients of the partial differential operator  $L$ , the domain  $\Omega$ , and the regularity parameters of the mesh  $\Xi$ .

*Proof.* Follows directly from Theorem 2.5.1 and Corollary 2.5.3.  $\square$

# Chapter 3

## The boundary element method

In the following, we consider a bounded Lipschitz domain  $T \subset \mathbb{R}^d$  with boundary  $\Gamma = \partial T$  and (unbounded) exterior domain  $T_{\text{ext}} = \mathbb{R}^d \setminus \bar{T}$ . By  $n(x)$ ,  $x \in \Gamma$ , we denote the unit outward normal vector on the surface  $\Gamma$ . We have  $n \in [L_\infty(\Gamma)]^3$ , and  $n$  is piecewise Lipschitz-continuous.

We consider a boundary value problem of the form (2.3), now posed on  $\Omega = T$ , and with constant coefficients  $A$ ,  $b$ , and  $c$ .

We can only give a brief summary of some standard results on boundary integral operators here and refer the reader to [66, 87, 106, 113, 102] for further details.

### 3.1 Boundary integral operators

Solutions of the partial differential equation (2.3) can be expressed in terms of its so-called fundamental solution, that is, a scalar function  $G(x, y)$  of the arguments  $x, y \in \mathbb{R}^d$  which satisfies the equation

$$L_x G(x, \cdot) = \delta_x \quad \forall x \in \mathbb{R}^d$$

in a distributional sense, where  $L_x$  means the differential operator  $L$  with respect to the variable  $x$ , and  $\delta_x$  is the Dirac delta distribution concentrated in  $x$ .

Partial differential operators with constant coefficients have translation-invariant fundamental solutions. We can thus write  $G$  as a function of a single vector argument

$$G(x, y) = G(x - y)$$

in this case. The importance of the fundamental solution for solving partial differential equations becomes clear by the argument that, for a suitable right-hand side  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ , the convolution  $G * f$  is a solution to the partial differential equation  $Lu = f$  on  $\mathbb{R}^d$ , since

$$L(G * f) = (LG) * f = \delta * f = f.$$

For the general elliptic partial differential operator  $L$  from (2.2) with constant coefficients  $A$ ,  $b$ , and  $c$ , Sauter and Schwab [106] give the fundamental solution  $G(x, y) =$

$G(x - y) = G(z)$  as

$$G(z) = \begin{cases} \frac{\exp(\langle b, z \rangle_{A^{-1}})}{2\pi\sqrt{\det A}} \log \frac{1}{\|z\|_{A^{-1}}}, & \text{for } d = 2 \text{ and } \lambda = 0, \\ \frac{\exp(\langle b, z \rangle_{A^{-1}})}{4\sqrt{\det A}} iH_0^{(1)}(i\lambda\|z\|_{A^{-1}}), & \text{for } d = 2 \text{ and } \lambda \neq 0, \\ \frac{1}{4\pi\sqrt{\det A}} \frac{\exp(\langle b, z \rangle_{A^{-1}} - \lambda\|z\|_{A^{-1}})}{\|z\|_{A^{-1}}}, & \text{for } d = 3, \end{cases}$$

where  $\theta = c + \|b\|_{A^{-1}}^2$ ,  $\lambda = \sqrt{\theta}$  for  $\theta \geq 0$  and  $\lambda = -i\sqrt{|\theta|}$  otherwise, and  $H_\alpha^{(1)}(x)$  is the Hankel function, or Bessel function of the third type. For  $A = I$ ,  $b = 0$ ,  $c = 0$ , we get the Laplace operator  $L = -\Delta$ , and the fundamental solution simplifies to

$$G(z) = \begin{cases} -\frac{1}{2\pi} \log |z| & \text{if } d = 2, \\ \frac{1}{4\pi} |z|^{-1} & \text{if } d = 3. \end{cases}$$

Following [87, 113, 106], we introduce the potentials

$$\begin{aligned} (\tilde{V}v)(x) &:= \int_{\Gamma} G(x - y)v(y) ds_y & \forall x \in \mathbb{R}^d \setminus \Gamma, \\ (\tilde{W}v)(x) &:= \int_{\Gamma} \tilde{\gamma}_y^1 G(x - y)v(y) ds_y & \forall x \in \mathbb{R}^d \setminus \Gamma \end{aligned}$$

as the boundary convolution of the fundamental solution or its modified conormal derivative with a given function  $v$ . Here  $\tilde{\gamma}_y^1$  denotes the conormal derivative  $\tilde{\gamma}^1$  associated with the adjoint problem introduced in Section 2.4 with respect to the variable  $y$ . This operator is also called the modified conormal derivative. We recall that, for sufficiently smooth functions, it is given by

$$\tilde{\gamma}^1 v = n \cdot \gamma^0(A^\top \nabla v + 2bv) = \gamma^0(A^\top \nabla v) \cdot n + 2\langle b, n \rangle \gamma^0 v,$$

where  $\gamma^0$  and  $\gamma^1$  are the Dirichlet and Neumann trace operators on  $T$  as introduced in Chapter 2. If  $v$  is less smooth,  $\tilde{\gamma}^1 v$  is given by the functional (2.9).

For suitable  $v$ , both of these potentials satisfy the partial differential equation in the classical sense away from the boundary.

**Lemma 3.1.1** ([106]). *For  $v \in L^1(\Gamma)$ , we have*

$$(L\tilde{V}v)(x) = (L\tilde{W}v)(x) = 0 \quad \forall x \in \mathbb{R}^d \setminus \Gamma.$$

They however differ in their behavior at the boundary  $\Gamma$ . The following lemma describes their jump properties at  $\Gamma$ .

**Lemma 3.1.2** ([106]). *For  $\psi \in H^{-1/2}(\Gamma)$  and  $\phi \in H^{1/2}(\Gamma)$ , we have the jump relations*

$$\begin{aligned} \llbracket \gamma^0 \tilde{V} \psi \rrbracket &= 0, & \llbracket \gamma^0 \tilde{W} \phi \rrbracket &= \phi & \text{in } H^{1/2}(\Gamma), \\ \llbracket \gamma^1 \tilde{V} \psi \rrbracket &= -\psi, & \llbracket \gamma^1 \tilde{W} \phi \rrbracket &= 0 & \text{in } H^{-1/2}(\Gamma), \end{aligned}$$

where we use the notation

$$\llbracket \gamma^0 w \rrbracket = \gamma_{\text{ext}}^0 w - \gamma_{\text{int}}^0 w, \quad \llbracket \gamma^1 w \rrbracket = \gamma_{\text{ext}}^1 w - \gamma_{\text{int}}^1 w,$$

$\gamma_{\text{int}}^0 = \gamma^0$  and  $\gamma_{\text{int}}^1 = \gamma^1$  are the interior trace operators on  $T$  as before, and  $\gamma_{\text{ext}}^0$  and  $\gamma_{\text{ext}}^1$  are the analogous (exterior) trace operators on  $T_{\text{ext}}$ .

The boundary integral operators associated with the interior boundary value problem (2.3) can now be introduced as Dirichlet and Neumann traces of these potentials.

**Definition 3.1.1.** We define the boundary integral operators

$$\begin{aligned} Vw &:= \gamma^0 \tilde{V}w, & \tilde{K}v &:= \gamma_{\text{int}}^0 \tilde{W}v, \\ \tilde{K}'w &:= \gamma_{\text{int}}^1 \tilde{V}w, & Dv &:= -\gamma^1 \tilde{W}v, \end{aligned}$$

as traces of the potentials  $\tilde{V}$  and  $\tilde{W}$ . The operators  $V$ ,  $\tilde{K}$ ,  $\tilde{K}'$ , and  $D$  are called, in turn, the *single layer potential*, *double layer potential*, *adjoint double layer potential*, and *hypersingular operators*.

The following theorem describes the mapping properties of these operators.

**Theorem 3.1.3** ([106]). *The operators introduced above are linear and bounded between the spaces*

$$\begin{aligned} \tilde{V} &: H^{-1/2}(\Gamma) \rightarrow H_{\text{loc}}^1(\mathbb{R}^d), & \tilde{W} &: H^{1/2}(\Gamma) \rightarrow H^1(\mathbb{R}^d \setminus \Gamma), \\ V &: H^{-1/2}(\Gamma) \rightarrow H^{1/2}(\Gamma), & \tilde{K} &: H^{1/2}(\Gamma) \rightarrow H^{1/2}(\Gamma), \\ \tilde{K}' &: H^{-1/2}(\Gamma) \rightarrow H^{-1/2}(\Gamma), & D &: H^{1/2}(\Gamma) \rightarrow H^{-1/2}(\Gamma). \end{aligned}$$

Here,  $H_{\text{loc}}^1(\mathbb{R}^d)$  is the Fréchet space of functions which are  $H^1$  on every compact subset of  $\mathbb{R}^d$ .

The abstract representation of the boundary integral operators as traces of potentials is not convenient for computational purposes. Under sufficient regularity conditions, they admit representations as improper integrals over the surface  $\Gamma$ . Indeed, this is the reason they are referred to as boundary integral operators. For the hypersingular operator, the associated bilinear form  $\langle D \cdot, \cdot \rangle$  is rewritten using integration by parts.

**Theorem 3.1.4** ([106, 51]). *The boundary integral operators have the following representations.*

- For  $w \in L_\infty(\Gamma)$ , we have

$$(Vw)(x) = \int_{\Gamma} G(x-y)w(y) ds_y \quad \forall x \in \Gamma$$

as an improper integral.

- Assume that  $\Gamma$  is piecewise  $C^2$ -smooth and that  $v \in L_\infty(\Gamma)$  is piecewise  $C^1$ . Then we have

$$\begin{aligned} (\tilde{K}v)(x) &= (\gamma_{\text{int}}^0 \tilde{W}v)(x) = -\frac{1}{2}v(x) + (Kv)(x) & \forall x \in \Gamma \text{ a.e.}, \\ (\tilde{K}'v)(x) &= (\gamma_{\text{int}}^1 \tilde{V}v)(x) = \frac{1}{2}v(x) + (K'v)(x) & \forall x \in \Gamma \text{ a.e.} \end{aligned}$$

with the improper integrals

$$\begin{aligned} (Kv)(x) &:= \int_{\Gamma} \tilde{\gamma}_y^1 G(x-y)v(y) ds_y, \\ (K'v)(x) &:= \int_{\Gamma} \gamma_x^1 G(x-y)v(y) ds_y. \end{aligned}$$

- Let  $\phi, \psi \in H^{1/2}(\Gamma)$ . Then

$$\begin{aligned} \langle D\phi, \psi \rangle &= \int_{\Gamma \times \Gamma} G(x-y) \langle \text{curl}_{\Gamma, A, 0} \phi(x), \text{curl}_{\Gamma, A, 2b} \psi(y) \rangle ds_x ds_y \\ &\quad + c \int_{\Gamma \times \Gamma} G(x-y) \phi(x) \psi(y) \langle A^{1/2}n(x), A^{1/2}n(y) \rangle ds_x ds_y \end{aligned}$$

where, for  $\lambda \in H^{1/2}(\Gamma)$ ,

$$\text{curl}_{\Gamma, A, v} \lambda := (A^{1/2} \nabla \tilde{\lambda} + \lambda A^{-1/2} v) \times A^{1/2} n$$

is the surface curl operator, with  $\tilde{\lambda} \in H^1(T)$  being an extension of  $\lambda$ .

In the event that  $A$  is symmetric and  $b = 0$ ,  $V$  and  $D$  are self-adjoint operators, whereas  $K$  and  $K'$  are adjoint to each other. In this setting, we introduce the subspace

$$H_*^{-1/2}(\Gamma) := \{w \in H^{-1/2}(\Gamma) : \langle w, 1 \rangle_{\Gamma} = 0\},$$

in which  $V$  can be shown to be coercive. Using this coercivity, one proves that there exists a unique element  $w_{\text{eq}} \in H^{-1/2}(\Gamma)$  with

$$Vw_{\text{eq}} = \text{const.}, \quad \langle w_{\text{eq}}, 1 \rangle_{\Gamma} = 1.$$

This  $w_{\text{eq}}$  is called the natural density, and  $\lambda := Vw_{\text{eq}} \in \mathbb{R}$  is called the capacity of  $\Gamma$ . Thus, for any  $w \in H^{-1/2}(\Gamma)$ , there exists a unique splitting  $w = w_* + w_0 w_{\text{eq}}$  with  $w_* \in H_*^{-1/2}(\Gamma)$  and  $w_0 = \langle w, 1 \rangle_{\Gamma} \in \mathbb{R}$ .

Furthermore, we introduce the subspace

$$H_*^{1/2}(\Gamma) := \{v \in H^{1/2}(\Gamma) : \langle w_{\text{eq}}, v \rangle_\Gamma = 0\},$$

and it can be shown that the single layer potential operator  $V : H_*^{-1/2}(\Gamma) \rightarrow H_*^{1/2}(\Gamma)$  restricted to  $H_*^{-1/2}(\Gamma)$  is an isomorphism between these two spaces. Since  $\langle w_{\text{eq}}, 1 \rangle_\Gamma = 1$ , we have for any  $v \in H^{1/2}(\Gamma)$  a unique splitting  $v = v_* + v_0$  with  $v_* \in H_*^{1/2}(\Gamma)$  and  $v_0 = \langle w_{\text{eq}}, v \rangle_\Gamma \in \mathbb{R}$ .

Under certain conditions, the single layer potential operator  $V$  is coercive on all of  $H^{-1/2}(\Gamma)$ . In two dimensions, this full coercivity requires the additional technical assumption that the diameter of the domain  $T$  be less than one. In fact, a sufficient condition is that the capacity  $\lambda$  be positive, and this is guaranteed by the condition  $\text{diam } T < 1$  [64]. We will therefore always assume  $\text{diam } T < 1$  in the following when  $d = 2$ .

**Theorem 3.1.5** ([113, 106]). *Let  $L = -\Delta$ . Then*

$$\langle w, Vw \rangle_\Gamma \geq c_V \|w\|_{H^{-1/2}(\Gamma)}^2 \quad \forall w \in H_*^{-1/2}(\Gamma).$$

*Assuming  $d = 3$  or, if  $d = 2$ , then  $\text{diam } T < 1$ , we even have*

$$\langle w, Vw \rangle_\Gamma \geq c_V \|w\|_{H^{-1/2}(\Gamma)}^2 \quad \forall w \in H^{-1/2}(\Gamma).$$

*Above,  $c_V$  is a positive constant.*

The bilinear form induced by  $D$  is coercive on  $H_*^{1/2}(\Gamma)$ .

**Theorem 3.1.6** ([113, 106]). *Let  $L = -\Delta$ . Then*

$$\langle Dv, v \rangle_\Gamma \geq c_D \|v\|_{H^{1/2}(\Gamma)}^2 \quad \forall v \in H_*^{1/2}(\Gamma).$$

*Above,  $c_D$  is a positive constant.*

Finally, on  $H_*^{1/2}(\Gamma)$ , we have the contraction property [114, 113]

$$(1 - c_K) \|v\|_{V^{-1}} \leq \|(\frac{1}{2}I \pm K)v\|_{V^{-1}} \leq c_K \|v\|_{V^{-1}} \quad \forall v \in H_*^{1/2}(\Gamma), \quad (3.1)$$

with the contraction constants

$$c_0 := \inf_{v \in H_*^{1/2}(\Gamma)} \frac{\langle Dv, v \rangle_\Gamma}{\langle V^{-1}v, v \rangle_\Gamma} \in (0, \frac{1}{4}), \quad (3.2)$$

$$c_K := \frac{1}{2} + \sqrt{\frac{1}{4} - c_0} \in (\frac{1}{2}, 1), \quad (3.3)$$

where  $\|v\|_{V^{-1}} = \sqrt{\langle V^{-1}v, v \rangle}$ . Here and in the following we implicitly exclude  $v = 0$  in infima and suprema of the above form.

## 3.2 Boundary integral equations and the Steklov-Poincaré operator

There are two approaches to deriving boundary integral equations for partial differential equations: the *direct* and the *indirect* method. In the indirect approach, Lemma 3.1.1 is exploited by observing that for a suitable boundary function  $\phi$ , both  $\tilde{V}\phi$  and  $\tilde{W}\phi$  are  $\mathcal{L}$ -harmonic functions in  $T$ . Thus, by finding a proper  $\phi$  such that the Dirichlet traces of one of these potentials match the desired boundary data  $g$ , a solution of the boundary value problem is found. This approach is algebraically simple, but has the disadvantage that the function  $\phi$  has no obvious relationship to the solution of the boundary value problem. In the following, we will only consider the direct approach, where the involved unknowns are just the Dirichlet and Neumann traces of the sought solution. Indeed, the following theorem provides the base for this approach and states that a function which satisfies a homogeneous partial differential equation of the form (2.3) can be represented explicitly in terms of potentials of its Dirichlet and Neumann traces.

**Theorem 3.2.1** (Representation formula [106, 87]). *Let  $\Phi \in \mathcal{H}(T)$  be a  $\mathcal{L}$ -harmonic function. Then it has the representation*

$$\Phi = \tilde{V}(\gamma_{\text{int}}^1 \Phi) - \tilde{W}(\gamma_{\text{int}}^0 \Phi) \quad \text{in } T. \quad (3.4)$$

Denote by  $u = \gamma_{\text{int}}^0 \Phi$  and  $t = \gamma_{\text{int}}^1 \Phi$  the Dirichlet and Neumann traces of the  $\mathcal{L}$ -harmonic function  $\Phi$ . Then, by applying  $\gamma_{\text{int}}^0$  and  $\gamma_{\text{int}}^1$  to (3.4) and recalling the definitions of the boundary integral operators, Definition 3.1.1, we obtain the *Calderón system*

$$\begin{aligned} u &= Vt - \tilde{K}u, \\ t &= \tilde{K}'t + Du. \end{aligned}$$

Using the representations of the boundary integral operators described in Section 3.1, we may also write

$$\begin{pmatrix} u \\ t \end{pmatrix} = \begin{pmatrix} \frac{1}{2}I - K & V \\ D & \frac{1}{2}I + K' \end{pmatrix} \begin{pmatrix} u \\ t \end{pmatrix}.$$

The  $2 \times 2$ -block operator in this equation is also called the *Calderón projector*.

From the first line of the above equation, we obtain  $Vt = (\frac{1}{2}I + K)u$ , and hence, assuming that  $V$  is coercive and thus invertible,

$$t = V^{-1}(\frac{1}{2}I + K)u.$$

Inserting this into the second line of the Calderón system, we obtain

$$t = Du + (\frac{1}{2}I + K')V^{-1}(\frac{1}{2}I + K)u.$$

By definition,  $u$  and  $t$  are the Dirichlet and Neumann traces, respectively, of an  $\mathcal{L}$ -harmonic function  $\Phi$ . They are thus related through the Dirichlet-to-Neumann map  $S = S_T$  introduced in (2.8) via  $t = Su$ . From this, the following theorem follows.

**Theorem 3.2.2** ([87, 113, 106]). *The Dirichlet-to-Neumann map*

$$S : H^{1/2}(\Gamma) \rightarrow H^{-1/2}(\Gamma)$$

on  $T$  has the representations

$$S = V^{-1}(\frac{1}{2}I + K) = D + (\frac{1}{2}I + K')V^{-1}(\frac{1}{2}I + K) \quad (3.5)$$

in terms of the boundary integral operators. This operator is also called the Steklov-Poincaré operator.

The bilinear form  $\langle S \cdot, \cdot \rangle$  induced by this operator satisfies, for  $u, v \in H^{1/2}(\Gamma)$ ,

$$\begin{aligned} \langle Su, v \rangle &= \langle Du, v \rangle + \langle V^{-1}(\frac{1}{2}I + K)u, (\frac{1}{2}I + K)v \rangle \\ &= \langle Du, v \rangle + \langle t(u), Vt(v) \rangle \end{aligned} \quad (3.6)$$

with

$$t(u) = V^{-1}(\frac{1}{2}I + K)u = Su.$$

**Theorem 3.2.3.** *Under the assumptions of Theorem 3.1.5, we have the relation*

$$(1 - c_K)\langle V^{-1}v, v \rangle \leq \langle Sv, v \rangle \leq c_K\langle V^{-1}v, v \rangle \quad \forall v \in H_*^{1/2}(\Gamma), \quad (3.7)$$

where  $c_K \in (\frac{1}{2}, 1)$  is the contraction constant from (3.1).

*Proof.* With the first representation of  $S$  and the Cauchy-Schwarz inequality, we have

$$\langle Sv, v \rangle = \langle V^{-1}(\frac{1}{2}I + K)v, v \rangle \leq \|(\frac{1}{2}I + K)v\|_{V^{-1}}\|v\|_{V^{-1}} \leq c_K\|v\|_{V^{-1}}^2,$$

where the last inequality is due to (3.1). Similarly, for the lower bound, we have

$$\begin{aligned} \langle Sv, v \rangle &= \langle V^{-1}(\frac{1}{2}I + K)v, v \rangle = \langle V^{-1}v, v \rangle - \langle V^{-1}(\frac{1}{2}I - K)v, v \rangle \\ &\geq \langle V^{-1}v, v \rangle - \|(\frac{1}{2}I - K)v\|_{V^{-1}}\|v\|_{V^{-1}} \\ &\geq \|v\|_{V^{-1}}^2 - c_K\|v\|_{V^{-1}}^2, \end{aligned}$$

where again (3.1) was used. □

The constant functions form the kernel of both  $(\frac{1}{2}I + K)$  and  $S$ . We recall that, for every  $v \in H^{1/2}(\Gamma)$ , there exists a unique splitting  $v = v_* + v_0$  with  $v_0$  constant and  $v_* \in H_*^{1/2}(\Gamma)$ . Making use of these facts, we can derive the following inequality that we will make use of later:

$$\begin{aligned} \|(\frac{1}{2}I + K)v\|_{V^{-1}} &= \|(\frac{1}{2}I + K)v_*\|_{V^{-1}} \\ &\leq c_K\|v_*\|_{V^{-1}} \leq \frac{c_K}{\sqrt{1 - c_K}}|v_*|_S = \frac{c_K}{\sqrt{1 - c_K}}|v|_S. \end{aligned} \quad (3.8)$$

Above we have used the seminorm  $|v|_S = \sqrt{\langle Sv, v \rangle}$ .

### 3.3 Approximation of the Steklov-Poincaré operator

Recall from Theorem 3.2.2 that the Dirichlet-to-Neumann map has the two representations

$$Su = V^{-1}(\tfrac{1}{2}I + K)u = Du + (\tfrac{1}{2}I + K')V^{-1}(\tfrac{1}{2}I + K)u.$$

which are often called the “non-symmetric” and “symmetric” representations. This nomenclature stems from the case of the potential equation, where indeed  $S$  is a self-adjoint operator but the discretization of the first representation above yields a non-symmetric matrix, while discretizing the latter preserves symmetry. We will however abuse these terms by employing them also in the case where the partial differential operator  $L$  is not formally self-adjoint, e.g., for convection-diffusion equations. Then  $S$  is not self-adjoint either and discretization always results in a non-symmetric matrix.

Neither of these two representations immediately permits a computable Galerkin discretization due to the occurrence of the inverse of the single layer potential  $V$ . Therefore, we rewrite  $S$  as

$$Su = Du + (\tfrac{1}{2}I + K')t(u)$$

with  $t(u) = V^{-1}(\tfrac{1}{2}I + K)u = Su \in H^{-1/2}(\Gamma)$ . Let now  $t_h(u) \in \mathcal{Z}_h$  be the Galerkin projection of  $t(u)$  onto some finite-dimensional space  $\mathcal{Z}_h \subset H^{-1/2}(\Gamma)$ . That is,  $t_h(u)$  is determined by the variational problem

$$\langle z_h, Vt_h(u) \rangle_\Gamma = \langle z_h, (\tfrac{1}{2}I + K)u \rangle_\Gamma \quad \forall z_h \in \mathcal{Z}_h. \quad (3.9)$$

One possible approximation to  $Su$  is then already the choice  $t_h(u)$ , since by design  $t_h(u) \approx t(u) = Su$ . The drawback of this choice is however that, if  $S$  is self-adjoint, then this symmetry is lost in the approximation  $t_h(u)$ . We will therefore in the following use the outer symmetric BEM approximation of  $S$  given by

$$\begin{aligned} \tilde{S} : H^{1/2}(\Gamma) &\rightarrow H^{-1/2}(\Gamma), \\ u &\mapsto Du + (\tfrac{1}{2}I + K')t_h(u), \end{aligned}$$

see, e.g., [26, 112, 113]. In the symmetric case, we observe using (3.9) that for all  $u, v \in H^{1/2}(\Gamma)$ ,

$$\begin{aligned} \langle \tilde{S}u, v \rangle &= \langle Du, v \rangle + \langle (\tfrac{1}{2}I + K')t_h(u), v \rangle \\ &= \langle Du, v \rangle + \langle t_h(u), (\tfrac{1}{2}I + K)v \rangle \\ &= \langle Du, v \rangle + \langle t_h(v), Vt_h(u) \rangle, \end{aligned}$$

where the last expression is clearly symmetric with respect to  $u$  and  $v$  such that  $\tilde{S}$  is then, too, a self-adjoint operator. Even more, we see in the following theorem that a spectral equivalence holds.

**Theorem 3.3.1.** *Under the assumptions of Theorem 3.1.5, the symmetric approximation  $\tilde{S}$  of the Steklov-Poincaré operator  $S$  fulfills the spectral equivalence relation (cf. [94, 112])*

$$\frac{c_0}{c_K} \langle Sv, v \rangle_\Gamma \leq \langle \tilde{S}v, v \rangle_\Gamma \leq \langle Sv, v \rangle_\Gamma \quad \forall v \in H^{1/2}(\Gamma). \quad (3.10)$$

*Proof.* By definition, the exact and approximated Neumann traces  $t(v)$  and  $t_h(v)$  satisfy the variational equations

$$\begin{aligned} \langle z, Vt(v) \rangle &= \langle z, (\tfrac{1}{2}I + K)v \rangle & \forall z \in H^{-1/2}(\Gamma), \\ \langle z_h, Vt_h(v) \rangle &= \langle z_h, (\tfrac{1}{2}I + K)v \rangle & \forall z_h \in \mathcal{Z}_h. \end{aligned}$$

By assumption,  $V$  is self-adjoint, and thus by applying Theorem 2.2.2 (Ritz), we obtain that  $t(v)$  and  $t_h(v)$  minimize the functional

$$J(t) := \tfrac{1}{2} \langle t, Vt \rangle - \langle t, (\tfrac{1}{2}I + K)v \rangle$$

over  $H^{-1/2}(\Gamma)$  and  $\mathcal{Z}_h$ , respectively. Since  $H^{-1/2}(\Gamma) \supset \mathcal{Z}_h$ , we obtain  $J(t(v)) \leq J(t_h(v))$ , that is,

$$\tfrac{1}{2} \langle t(v), Vt(v) \rangle - \langle t(v), (\tfrac{1}{2}I + K)v \rangle \leq \tfrac{1}{2} \langle t_h(v), Vt_h(v) \rangle - \langle t_h(v), (\tfrac{1}{2}I + K)v \rangle.$$

Since, again by the defining relations above,  $\langle t(v), (\tfrac{1}{2}I + K)v \rangle = \langle t(v), Vt(v) \rangle$  and  $\langle t_h(v), (\tfrac{1}{2}I + K)v \rangle = \langle t_h(v), Vt_h(v) \rangle$ , we obtain

$$-\tfrac{1}{2} \langle t(v), Vt(v) \rangle \leq -\tfrac{1}{2} \langle t_h(v), Vt_h(v) \rangle. \quad (3.11)$$

Recalling the representations

$$\begin{aligned} \langle Sv, v \rangle &= \langle Dv, v \rangle + \langle t(v), Vt(v) \rangle, \\ \langle \tilde{S}v, v \rangle &= \langle Dv, v \rangle + \langle t_h(v), Vt_h(v) \rangle \end{aligned}$$

derived previously, the upper bound for  $\tilde{S}$  thus follows directly with (3.11).

For the lower bound, we split  $v = v_0 + v_*$  with  $v_0 \in \mathbb{R}$  and  $v_* \in H_*^{1/2}(\Gamma)$  and recall that the constants comprise the kernel of both  $S$  and  $D$ . Thus, using the ellipticity of  $V$ , we can estimate

$$\begin{aligned} \langle \tilde{S}v, v \rangle &= \langle Dv, v \rangle + \langle t_h(v), Vt_h(v) \rangle \\ &\geq \langle Dv, v \rangle = \langle Dv_*, v_* \rangle \geq c_0 \langle V^{-1}v_*, v_* \rangle \geq \frac{c_0}{c_K} \langle Sv_*, v_* \rangle = \frac{c_0}{c_K} \langle Sv, v \rangle, \end{aligned}$$

where we used the relations (3.2) and (3.7).  $\square$

We have above not explicitly specified the discrete space  $\mathcal{Z}_h$  in which  $t_h(u)$  is approximated. In this work, we will always stick to the simple and natural choice given by the space of piecewise constant functions on  $\Gamma$ . The boundary triangulation with respect to which these piecewise constant functions are defined will arise naturally in the construction of the BEM-based FEM in Chapter 4.

### 3.4 Explicit constants for boundary integral operators

The constant  $\tilde{c} := \frac{c_0}{c_K}$  from (3.10) is known to lie in the interval  $(0, \frac{1}{4}]$  and does not depend on the scaling of the domain  $T$ . However, it does depend on the shape of  $T$ . We will later see that robust error estimates for the BEM-based FEM require bounding  $\tilde{c}$  uniformly from below for a large family of possible shapes of  $T$ . It is only recently that bounds for these constants which are explicit with respect to the shape of  $T$  have been investigated, starting with a paper by Pechstein [95] which relied on the so-called Jones parameter and a constant in an isoperimetric inequality. These results were employed in the first rigorous *a priori* error analysis of the BEM-based FEM [61] as well as in the later analysis based on the mixed formulation which led to  $L_2$  error estimates [59]. Later, the assumptions were simplified such that, at least in the three-dimensional case, only relatively standard assumptions on mesh regularity have to be imposed [60].

In this section, we collect mostly results from [95], but present them in a slightly different way. Namely, we aim to write all arising constants only in terms of Poincaré constants as well as certain Sobolev extension constants, eliminating the dependencies on the Jones parameter as well as isoperimetric constants. Suitable explicitly bounded extension operators for polytopal domains will be constructed in Section 5.6. For simplicity, we focus on the Laplacian case,  $L = -\Delta$ , throughout this section.

We define the seminorm

$$|v|_{\star, H^{1/2}(\Gamma)} := \inf_{\substack{\tilde{v} \in H^1(T) \\ \tilde{v}|_{\Gamma} = v}} |\tilde{v}|_{H^1(T)} \quad \forall v \in H^{1/2}(\Gamma)$$

which coincides with the energy of the harmonic extension  $\mathcal{H} : H^{1/2}(\Gamma) \rightarrow H^1(T)$  defined in Chapter 2,

$$|v|_{\star, H^{1/2}(\Gamma)} = |\mathcal{H}v|_{H^1(T)},$$

as well as the full norm

$$\|v\|_{\star, H^{1/2}(\Gamma)}^2 := \left( \|\mathcal{H}v\|_{L_2(T)} + (\text{diam } T)^{-2} |v|_{\star, H^{1/2}(\Gamma)}^2 \right)^{1/2}.$$

Referring to Lemma 2.1.1, we see that the scaling factor  $(\text{diam } T)^{-2}$  in the definition of the norm makes both contributions behave identically with respect to uniform scaling of the domain  $T$ . Finally, we obtain a dual norm defined in the usual way,

$$\|w\|_{\star, H^{-1/2}(\Gamma)} := \sup_{v \in H^{1/2}(\Gamma)} \frac{\langle v, w \rangle}{\|v\|_{\star, H^{1/2}(\Gamma)}} \quad \forall w \in H^{-1/2}(\Gamma).$$

In the following, we will sometimes mention the space

$$H_{\text{loc}\star}^1(D) := \{v \in H_{\text{loc}}^1(D) : |v|_{H^1(D)} < \infty\},$$

where  $D \subseteq \mathbb{R}^d$  is a not necessarily bounded Lipschitz domain, or all of  $\mathbb{R}^d$ . While the functions in this space do not necessarily decay to 0 at infinity, they do have finite energy.

**Lemma 3.4.1.** *Assume that there exists a linear extension operator*

$$E_{\text{ext}} : H_{\text{loc}*}^1(T_{\text{ext}}) \rightarrow H_{\text{loc}*}^1(\mathbb{R}^d)$$

which satisfies, with some constant  $C_{E_{\text{ext}}} > 0$ ,

$$|E_{\text{ext}}v|_{H^1(T)} \leq C_{E_{\text{ext}}} |v|_{H^1(T_{\text{ext}})} \quad \forall v \in H_{\text{loc}*}^1(T_{\text{ext}}).$$

Then we have the Dirichlet trace inequalities

$$|\gamma_{\text{int}}^0 v|_{\star, H^{1/2}(\Gamma)} \leq |v|_{H^1(T)} \quad \forall v \in H^1(T), \quad (3.12)$$

$$|\gamma_{\text{ext}}^0 v|_{\star, H^{1/2}(\Gamma)} \leq C_{E_{\text{ext}}} |v|_{H^1(T_{\text{ext}})} \quad \forall v \in H_{\text{loc}*}^1(T_{\text{ext}}). \quad (3.13)$$

*Proof.* For  $v \in H^1(T)$ , we see immediately that

$$|\gamma_{\text{int}}^0 v|_{\star, H^{1/2}(\Gamma)} = |\mathcal{H}\gamma_{\text{int}}^0 v|_{H^1(T)} \leq |v|_{H^1(T)}.$$

For  $v \in H_{\text{loc}*}^1(T_{\text{ext}})$ , we have similarly

$$|\gamma_{\text{ext}}^0 v|_{\star, H^{1/2}(\Gamma)} = |\mathcal{H}\gamma_{\text{ext}}^0 v|_{H^1(T)} \leq |E_{\text{ext}}v|_{H^1(T)} \leq C_{E_{\text{ext}}} |v|_{H^1(T_{\text{ext}})}. \quad \square$$

The Neumann trace operator satisfies similar estimates.

**Lemma 3.4.2.** *Assume that there exists a linear, bounded extension operator*

$$E_{\text{int}} : H^1(T) \rightarrow H^1(\mathbb{R}^d)$$

with norm  $C_{E_{\text{int}}} \geq 1$ . Then we have the Neumann trace inequalities

$$\|\gamma_{\text{int}}^1 v\|_{\star, H^{-1/2}(\Gamma)} \leq |v|_{H^1(T)} \quad \forall v \in H^1(T), \Delta v = 0 \text{ weakly}, \quad (3.14)$$

$$\|\gamma_{\text{ext}}^1 v\|_{\star, H^{-1/2}(\Gamma)} \leq C_{E_{\text{int}}} |v|_{H^1(T_{\text{ext}})} \quad \forall v \in H_{\text{loc}*}^1(T_{\text{ext}}), \Delta v = 0 \text{ weakly}. \quad (3.15)$$

*Proof.* For  $w \in H^{1/2}(\Gamma)$ , we have that

$$\langle \gamma_{\text{int}}^1 v, w \rangle = \int_T \nabla v \cdot \nabla \mathcal{H}w \, dx \leq |v|_{H^1(T)} |\mathcal{H}w|_{H^1(T)} = |v|_{H^1(T)} |w|_{\star, H^{1/2}(\Gamma)},$$

and thus by the definition of the dual norm,

$$\|\gamma_{\text{int}}^1 v\|_{\star, H^{-1/2}(\Gamma)} \leq \sup_{w \in H^{1/2}(\Gamma)} \frac{|v|_{H^1(T)} |w|_{\star, H^{1/2}(\Gamma)}}{\|w\|_{\star, H^{1/2}(\Gamma)}} \leq |v|_{H^1(T)}.$$

For the second statement, assume first that  $\text{diam}(T) = 1$ . For  $w \in H^{1/2}(\Gamma)$ , we set  $\tilde{w} = E_{\text{int}} \mathcal{H}w \in H^1(\mathbb{R}^d)$ , and we can estimate

$$|w|_{H^1(T_{\text{ext}})} \leq C_{E_{\text{int}}} \|\mathcal{H}w\|_{H^1(T)} = C_{E_{\text{int}}} \|w\|_{\star, H^{1/2}(T)}.$$

Thus we have

$$\begin{aligned} -\langle \gamma_{\text{ext}}^1 v, w \rangle &= \int_{T_{\text{ext}}} \nabla v \cdot \nabla \tilde{w} \, dx \leq |v|_{H^1(T_{\text{ext}})} |\tilde{w}|_{H^1(T_{\text{ext}})} \\ &\leq C_{E_{\text{int}}} |v|_{H^1(T_{\text{ext}})} \|w\|_{\star, H^{1/2}(T)}, \end{aligned}$$

and the statement follows as above using the definition of the dual norm. For general  $\text{diam}(T)$ , the statement follows from a simple dilation argument.  $\square$

With these results, we obtain a version of Theorem 3.1.5 with an explicitly bounded coercivity constant.

**Lemma 3.4.3.** *Assume  $d = 3$  and the existence of an extension operator  $E_{\text{int}}$  as in Lemma 3.4.2. Then the single layer potential operator satisfies the coercivity estimate*

$$\langle w, Vw \rangle \geq c_V^* \|w\|_{\star, H^{-1/2}(\Gamma)}^2 \quad \forall w \in H^{-1/2}(\Gamma)$$

with

$$c_V^* = \frac{1}{2} C_{E_{\text{int}}}^{-2}.$$

*Proof.* Set  $u = \tilde{V}w$ , and thus, due to Lemma 3.1.1,  $Lu = 0$  in  $T$  and  $T_{\text{ext}}$  separately. In addition, it can be shown that  $u|_{T_{\text{ext}}} \in H_{\text{loc}\star}^1(T_{\text{ext}})$  (cf. [87]). Its traces then satisfy

$$\langle \gamma_{\text{int}}^1 u, \gamma_{\text{int}}^0 u \rangle = |u|_{H^1(T)}^2, \quad \langle \gamma_{\text{ext}}^1 u, \gamma_{\text{ext}}^0 u \rangle = -|u|_{H^1(T_{\text{ext}})}^2.$$

Using the jump relations from Lemma 3.1.2 and the above formulas, we obtain

$$\langle w, Vw \rangle = \langle w, \gamma^0 u \rangle = -\langle \llbracket \gamma^1 u \rrbracket, \gamma^0 u \rangle = |u|_{H^1(T)}^2 + |u|_{H^1(T_{\text{ext}})}^2. \quad (3.16)$$

Applying Lemma 3.4.2 and again Lemma 3.1.2 yields

$$\begin{aligned} \langle w, Vw \rangle &\geq \|\gamma_{\text{int}}^1 u\|_{\star, H^{-1/2}(\Gamma)}^2 + C_{E_{\text{int}}}^{-2} \|\gamma_{\text{ext}}^1 u\|_{\star, H^{-1/2}(\Gamma)}^2 \\ &\geq \frac{1}{2} C_{E_{\text{int}}}^{-2} \|\gamma_{\text{int}}^1 u - \gamma_{\text{ext}}^1 u\|_{\star, H^{-1/2}(\Gamma)}^2 = \frac{1}{2} C_{E_{\text{int}}}^{-2} \|w\|_{\star, H^{-1/2}(\Gamma)}^2. \quad \square \end{aligned}$$

**Corollary 3.4.4.** *Under the assumptions of Lemma 3.4.3, we have*

$$\langle V^{-1}v, v \rangle \leq (c_V^*)^{-1} \|v\|_{\star, H^{1/2}(\Gamma)}^2 \quad \forall v \in H^{1/2}(\Gamma).$$

*Proof.* By standard duality arguments.

**Lemma 3.4.5.** *Assume the existence of an extension operator  $E_{\text{ext}}$  as in Lemma 3.4.1. Then the hypersingular operator satisfies*

$$\langle Dv, v \rangle \geq c_D^* |v|_{\star, H^{1/2}(\Gamma)}^2 \quad \forall v \in H^{1/2}(\Gamma)$$

with

$$c_D^* = \frac{1}{2} (C_{E_{\text{ext}}})^{-2}.$$

*Proof.* We set  $u = \widetilde{W}v$ , and thus, due to Lemma 3.1.1,  $Lu = 0$  in  $T$  and  $T_{\text{ext}}$  separately. In addition, it can be shown that  $u|_{T_{\text{ext}}} \in H_{\text{loc}*}^1(T_{\text{ext}})$  (cf. [87]). Its traces then satisfy

$$\langle \gamma_{\text{int}}^1 u, \gamma_{\text{int}}^0 u \rangle = |u|_{H^1(T)}^2, \quad \langle \gamma_{\text{ext}}^1 u, \gamma_{\text{ext}}^0 u \rangle = -|u|_{H^1(T_{\text{ext}})}^2.$$

Using the jump relations from Lemma 3.1.2 and the above formulas, we obtain

$$\begin{aligned} \langle Dv, v \rangle &= \langle -\gamma^1 u, \llbracket \gamma^0 u \rrbracket \rangle = -\langle \gamma_{\text{ext}}^1 u, \gamma_{\text{ext}}^0 u \rangle + \langle \gamma_{\text{int}}^1 u, \gamma_{\text{int}}^0 u \rangle \\ &= |u|_{H^1(T)}^2 + |u|_{H^1(T_{\text{ext}})}^2. \end{aligned}$$

Applying Lemma 3.4.1 and again Lemma 3.1.2 yields

$$\begin{aligned} \langle Dv, v \rangle &\geq |\gamma_{\text{int}}^0 u|_{\star, H^{1/2}(\Gamma)}^2 + C_{E_{\text{ext}}}^{-2} |\gamma_{\text{ext}}^0 u|_{\star, H^{1/2}(\Gamma)}^2 \\ &\geq \frac{1}{2} C_{E_{\text{ext}}}^{-2} |\gamma_{\text{ext}}^0 u - \gamma_{\text{int}}^0 u|_{\star, H^{1/2}(\Gamma)}^2 = \frac{1}{2} C_{E_{\text{ext}}}^{-2} |v|_{\star, H^{1/2}(\Gamma)}^2. \quad \square \end{aligned}$$

In order to extend this coercivity result to the full norm, we need a Poincaré-type inequality for harmonic functions with a constant we can explicitly bound. The next theorem, for the proof of which we point to the given reference, provides such a result.

**Lemma 3.4.6** ([95]). *Let  $d = 3$ ,  $C_P$  be the Poincaré constant of  $T$ , and assume the existence of an extension operator  $E_{\text{int}}$  as in Lemma 3.4.2. Then it holds*

$$\text{diam}(T)^{-2} \|\mathcal{H}v\|_{L_2(T)}^2 \leq C_P^* |\mathcal{H}v|_{H^1(T)}^2 \quad \forall v \in H_*^{1/2}(\Gamma)$$

with

$$C_P^* = 2 \left( C_P^2 + (c_V^*)^{-1} (1 + C_P^2) \right)$$

and  $c_V^*$  as in Lemma 3.4.3.

This lemma allows us to obtain a version of Theorem 3.1.6 with explicit coercivity constant.

**Corollary 3.4.7.** *Assume the existence of an extension operator  $E_{\text{ext}}$  as in Lemma 3.4.1 and of an extension operator  $E_{\text{int}}$  as in Lemma 3.4.2. Then the hypersingular operator satisfies*

$$\langle Dv, v \rangle \geq \frac{c_D^*}{1 + C_P^*} \|v\|_{\star, H^{1/2}(\Gamma)}^2 \quad \forall v \in H_*^{1/2}(\Gamma)$$

with  $C_P^*$  as in Lemma 3.4.6 and  $c_D^*$  as in Lemma 3.4.5.

*Proof.* Lemma 3.4.6 implies that

$$\|v\|_{\star, H^{1/2}(\Gamma)}^2 \leq (1 + C_P^*) |v|_{\star, H^{1/2}(\Gamma)}^2,$$

and thus the statement follows with Lemma 3.4.5. □

In addition to the coercivity estimates of the previous theorems, we also require explicit bounds on the norms of the involved boundary integral operators.

**Lemma 3.4.8** ([95]). *Assume  $d = 3$  and the existence of an extension operator  $E_{int}$  as in Lemma 3.4.2. Then, for all  $w \in H^{-1/2}(\Gamma)$ , we have*

$$\begin{aligned} \|Vw\|_{\star, H^{1/2}(\Gamma)} &\leq C_V^* \|w\|_{\star, H^{-1/2}(\Gamma)}, \\ \langle w, Vw \rangle &\leq C_V^* \|w\|_{\star, H^{-1/2}(\Gamma)}^2, \end{aligned}$$

where

$$C_V^* = 1 + 2C_P^*$$

with  $C_P^*$  as in Lemma 3.4.6.

*Proof.* Fix  $w \in H^{-1/2}(\Gamma)$  with the unique splitting  $w = w_* + w_0 w_{eq}$ , where  $w_* \in H_*^{-1/2}(\Gamma)$ ,  $w_{eq} \in H^{-1/2}(\Gamma)$  is the natural density, and  $w_0 = \langle w, 1 \rangle \in \mathbb{R}$ , as described in Section 3.1. Since  $\langle w_*, Vw_{eq} \rangle = 0$  by definition, this splitting is orthogonal in the  $V$ -inner product, and we have

$$\langle w, Vw \rangle = \langle w_*, Vw_* \rangle + w_0^2 \langle w_{eq}, Vw_{eq} \rangle = \langle w_*, Vw_* \rangle + \lambda w_0^2, \quad (3.17)$$

where we recalling the definition of the capacity,  $\lambda = Vw_{eq} \in \mathbb{R}$ . We observe that, since  $Vw_{eq} = \text{const.}$ , we have

$$\begin{aligned} \|Vw\|_{\star, H^{1/2}(\Gamma)}^2 &= |Vw_*|_{\star, H^{1/2}(\Gamma)}^2 + (\text{diam } T)^{-2} \|\mathcal{H}Vw\|_{L_2(T)}^2 \\ &\leq |Vw_*|_{\star, H^{1/2}(\Gamma)}^2 + 2(\text{diam } T)^{-2} \left( \|\mathcal{H}Vw_*\|_{L_2(T)}^2 + \|\mathcal{H}Vw_0w_{eq}\|_{L_2(T)}^2 \right) \\ &\leq (1 + 2C_P^*) |Vw_*|_{\star, H^{1/2}(\Gamma)}^2 + 2(\text{diam } T)^{-2} \|\mathcal{H}Vw_0w_{eq}\|_{L_2(T)}^2, \end{aligned}$$

where we used using Lemma 3.4.6 for the last estimate. Since  $\mathcal{H}$  maps constants to constants, we have

$$\|\mathcal{H}Vw_0w_{eq}\|_{L_2(T)}^2 = |T| w_0^2 \lambda^2$$

Furthermore,

$$\begin{aligned} |Vw_*|_{\star, H^{1/2}(\Gamma)}^2 &= |\mathcal{H}Vw_*|_{H^1(T)}^2 \leq |\tilde{V}w_*|_{H^1(T)}^2 \\ &\leq |\tilde{V}w_*|_{H^1(T)}^2 + |\tilde{V}w_*|_{H^1(T_{ext})}^2 = \langle w_*, Vw_* \rangle, \end{aligned}$$

where we used relation (3.16). Thus we obtain

$$\|Vw\|_{\star, H^{1/2}(\Gamma)}^2 \leq (1 + 2C_P^*) \langle w_*, Vw_* \rangle + 2|T| (\text{diam } T)^{-2} w_0^2 \lambda^2.$$

In three dimensions, it can be shown that the capacity is bounded by  $\lambda \leq (\text{diam } T)^2 / |T|$  (see [95, Proof of Lemma 6.7]). Thus, since  $1 \leq C_P^*$ , we can estimate with  $C_V^* = 1 + 2C_P^*$  and equality (3.17)

$$\begin{aligned} \|Vw\|_{\star, H^{1/2}(\Gamma)}^2 &\leq C_V^* \left( \langle w_*, Vw_* \rangle + \lambda w_0^2 \right) = C_V^* \langle w, Vw \rangle \\ &\leq C_V^* \|w\|_{\star, H^{-1/2}(\Gamma)} \|Vw\|_{\star, H^{1/2}(\Gamma)}. \end{aligned}$$

This proves the first statement, and the second follows immediately.  $\square$

**Lemma 3.4.9** ([95]). *For all  $v \in H^{1/2}(\Gamma)$ , we have*

$$\begin{aligned} \|Dv\|_{\star, H^{-1/2}(\Gamma)} &\leq |v|_{\star, H^{1/2}(\Gamma)}, \\ \langle Dv, v \rangle &\leq |v|_{\star, H^{1/2}(\Gamma)}^2. \end{aligned}$$

*Proof.* We recall that

$$|v|_{\star, H^{1/2}(\Gamma)}^2 = \langle Sv, v \rangle = \langle Dv, v \rangle + \langle V^{-1}(\frac{1}{2}I + K)v, (\frac{1}{2}I + K)v \rangle \geq \langle Dv, v \rangle,$$

proving the second statement. For the first statement, let first  $v \in H_*^{1/2}(\Gamma)$ . Using duality, the fact that  $D$  vanishes on constants, the Cauchy-Schwarz inequality, and the above estimate, we obtain

$$\begin{aligned} \|Dv\|_{\star, H^{-1/2}(\Gamma)} &= \sup_{y \in H^{1/2}(\Gamma)} \frac{\langle Dv, y \rangle}{\|y\|_{\star, H^{1/2}(\Gamma)}} \\ &= \sup_{\substack{y_* \in H_*^{1/2}(\Gamma) \\ y_0 \in \mathbb{R}}} \frac{\langle Dv, y_* \rangle}{(|y_*|_{\star, H^{1/2}(\Gamma)}^2 + (\text{diam } T)^{-2} \|\mathcal{H}(y_* + y_0)\|_{L_2(T)}^2)^{1/2}} \\ &\leq \sup_{y_* \in H_*^{1/2}(\Gamma)} \frac{\langle Dv, v \rangle^{1/2} \langle Dy_*, y_* \rangle^{1/2}}{|y_*|_{\star, H^{1/2}(\Gamma)}} \\ &\leq |v|_{\star, H^{1/2}(\Gamma)}. \end{aligned}$$

Since  $Dv$  and  $|v|_{\star, H^{1/2}(\Gamma)}$  remain invariant if a constant is added to  $v$ , the above statement holds true for all  $v \in H^{1/2}(\Gamma)$ .  $\square$

Combining some of the above results, we get an explicit bound for the constant  $c_0$  from (3.2), and thus also for  $c_K$  from (3.3).

**Theorem 3.4.10.** *Assume the existence of an exterior extension operator  $E_{\text{ext}}$  as in Lemma 3.4.1 and of an interior extension operator  $E_{\text{int}}$  as in Lemma 3.4.2. Then we have the bound*

$$c_0 = \inf_{v \in H_*^{1/2}(\Gamma)} \frac{\langle Dv, v \rangle}{\langle V^{-1}v, v \rangle} \geq \frac{c_V^* c_D^*}{1 + C_P^*}$$

with  $c_V^*$  as in Lemma 3.4.3,  $C_P^*$  as in Lemma 3.4.6, and  $c_D^*$  as in Lemma 3.4.5. These constants can be tracked back to the Poincaré constant  $C_P$  of  $T$  and the extension constants  $C_{E_{\text{int}}}$  and  $C_{E_{\text{ext}}}$ .

*Proof.* Follows directly from Corollary 3.4.4 and Corollary 3.4.7. □

# Chapter 4

## Derivation of the BEM-based FEM

The aim of this chapter is the derivation of the BEM-based finite element method that is the main subject of this work. To this end, in Section 4.1, we first introduce the notion of a generalized polytopal mesh and then derive a variational formulation which involves spaces of functions which live only on the so-called skeleton, that is, the union of all element boundaries. This skeletal variational formulation is shown to be equivalent to the standard variational formulation (2.4). It turns out that, in some situations, a mixed variational equation which involves both the Dirichlet and Neumann data as unknowns has certain advantages over a primal formulation, and this mixed formulation is derived in Section 4.2. Finally, in Section 4.3, we discretize both variational formulations. This step involves both a discretization of the skeletal spaces and an approximation of the occurring Steklov-Poincaré operators by the boundary element techniques introduced in Chapter 3, which gives the method its name.

### 4.1 A skeletal variational formulation

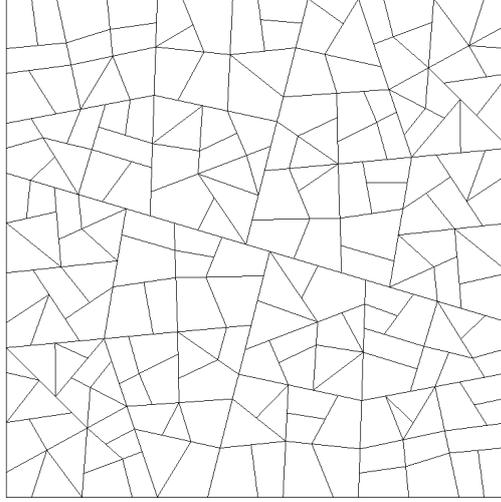
Consider a boundary value problem of the form (2.3) with piecewise constant coefficient functions  $A$ ,  $b$ , and  $c$ . Furthermore, for simplicity, we will assume that the right-hand side  $f$  is 0. For the general case  $f \neq 0$ , element-local Newton potentials can be used to homogenize the problem.

Finite element methods typically use the variational formulation (2.4) as their starting point. In our approach, however, we first introduce a mesh and derive a skeletal reformulation of (2.4). Later on, we will restrict it to discrete trial spaces.

Consider a non-overlapping decomposition  $\mathcal{T}$  of  $\Omega$ ,

$$\bar{\Omega} = \bigcup_{T \in \mathcal{T}} \bar{T}, \quad T_1 \cap T_2 = \emptyset \quad \forall T_1 \neq T_2 \text{ from } \mathcal{T}.$$

In contrast to the simplicial finite element mesh  $\Xi$  introduced in Section 2.5, we now allow each *element*  $T \in \mathcal{T}$  to be a  $d$ -dimensional Lipschitz polytope, i.e., a polygon or polyhedron. We will again call such a decomposition  $\mathcal{T}$  a *mesh* of  $\Omega$ . We require that the piecewise constant coefficients  $A$ ,  $b$ ,  $c$  of the partial differential operator  $L$  have their jumps aligned with the element boundaries. In other words, for each element  $T \in \mathcal{T}$ ,



**Figure 4.1:** A heterogeneous polygonal mesh.

there exists a partial differential operator  $L_T$  with constant coefficients  $A_T$ ,  $b_T$ ,  $c_T$  such that  $L|_T = L_T$ , and an associated local bilinear form

$$\forall u, v \in H^1(T) :$$

$$\mathcal{L}_T(u, v) = \int_T (A_T \nabla u(x) \cdot \nabla v(x) + 2b_T \cdot \nabla u(x) v(x) + c_T u(x) v(x)) dx$$

appearing in the weak formulation of the partial differential equation on  $T$ .

In the following we will frequently refer to the *local mesh sizes*  $h_T := \text{diam} T$  and the *global mesh size*  $h := \max\{h_T : T \in \mathcal{T}\}$ . In this work, we are interested in families of such meshes where the element diameters  $h_T$  uniformly tend to zero, while the number of boundary facets of every element remains uniformly bounded by a small constant. Within this framework we can treat typical element shapes like triangles or quadrilaterals in two dimensions, tetrahedra, hexahedra, prisms or pyramids in three dimensions, as well as other, less standard shapes. In particular, we do not necessarily assume convexity of the elements. We also retain the freedom to mix all these types of elements within one mesh; see Figure 4.1 for an example in 2D. Note especially that “hanging” nodes which are vertices of one element, but not a neighboring one, considered problematic in a standard FEM setting, can be easily treated in our framework by artificially introducing the hanging node as a new vertex of the latter element.

We define a restricted trial space by requiring that the trial functions fulfill the homogeneous PDE locally in every element, while being globally continuous. Thus, we

introduce the spaces

$$\begin{aligned}\mathcal{V}_{\mathcal{H}} &:= \{v \in H^1(\Omega) : v|_T \in \mathcal{H}(T) \quad \forall T \in \mathcal{T}\}, \\ \mathcal{V}_{\mathcal{H},0} &:= \mathcal{V}_{\mathcal{H}} \cap H_0^1(\Omega),\end{aligned}$$

with the space  $\mathcal{H}(T)$  of  $\mathcal{L}$ -harmonic functions on the element  $T$  defined, in analogy to (2.5), as

$$\mathcal{H}(T) := \left\{ u \in H^1(T) : \mathcal{L}_T(u, v) = 0 \quad \forall v \in H_0^1(T) \right\}.$$

Noting that  $\mathcal{V}_{\mathcal{H}} \subset H^1(\Omega)$  and  $\mathcal{V}_{\mathcal{H},0} \subset H_0^1(\Omega)$ , we state a restricted version of the variational problem (2.4) as follows: find  $u \in \mathcal{V}_{\mathcal{H}}$  which satisfies

$$\gamma_{\Omega} u = g, \quad \mathcal{L}(u, v) = 0 \quad \forall v \in \mathcal{V}_{\mathcal{H},0}. \quad (4.1)$$

**Lemma 4.1.1.** *Under the assumptions of Theorem 2.4.1, the problems (2.4) (with  $F = 0$ ) and (4.1) are equivalent.*

*Proof.* Owing to  $\mathcal{V}_{\mathcal{H}} \subset H^1(\Omega)$ , the boundedness and coercivity properties of the bilinear form  $\mathcal{L}(\cdot, \cdot)$  established in the proof of Theorem 2.4.1 carry over to (4.1). It follows by Lemma 2.2.1 (Lax-Milgram) that (4.1) has a unique solution. In order to establish equivalence of the two formulations, it suffices to show that the solution  $u \in H^1(\Omega)$  of (2.4) lies in  $\mathcal{V}_{\mathcal{H}}$ , that is, that for any  $T \in \mathcal{T}$ , we have  $u|_T \in \mathcal{H}(T)$ . To show this, fix  $T$  and choose an arbitrary function  $v_T \in H_0^1(T)$ , extending it by zero to  $v \in H_0^1(\Omega)$ . Testing with this choice of  $v$  in (2.4) yields

$$\mathcal{L}_T(u|_T, v_T) = 0,$$

and since  $v_T \in H_0^1(T)$  was arbitrary,  $u \in \mathcal{V}_{\mathcal{H}}$  follows.  $\square$

Let  $\gamma_T^0 : H^1(T) \rightarrow H^{1/2}(\partial T)$  and  $\gamma_T^1 : \mathcal{H}(T) \rightarrow H^{-1/2}(\partial T)$ , respectively, denote the Dirichlet and Neumann trace operators on  $T$  as defined in Chapter 2. We recall the Green's identity (2.7) for harmonic functions  $u_T \in \mathcal{H}(T)$  which now reads

$$\langle \gamma_T^1 u_T, \gamma_T^0 v_T \rangle_{\partial T} = \mathcal{L}_T(u_T, v_T) \quad \forall v_T \in H^1(T). \quad (4.2)$$

It allows us to rewrite the variational problem (4.1) as follows: we seek  $u \in \mathcal{V}_{\mathcal{H}}$  satisfying

$$\gamma_{\Omega} u = g, \quad \sum_{T \in \mathcal{T}} \langle \gamma_T^1 u, \gamma_T^0 v \rangle_{\partial T} = 0 \quad \forall v \in \mathcal{V}_{\mathcal{H},0}. \quad (4.3)$$

The only values of  $u$  occurring in this formulation are the Neumann traces on the element boundaries. This gives rise to the idea of representing  $u$  solely via its values on the *skeleton*  $\Gamma_S = \bigcup_{T \in \mathcal{T}} \partial T$ . We remark that  $\Gamma_S$  is not a manifold due to its locally branching

structure at interior vertices and edges of the mesh, and therefore the usual definition of Sobolev spaces on manifolds is not applicable. We can, however, define

$$H^{1/2}(\Gamma_S) := \left\{ (v_T)_{T \in \mathcal{T}} \mid \exists \varphi \in H^1(\Omega) \forall T \in \mathcal{T} : v_T = \gamma_T^0 \varphi \right\} \subset \bigotimes_{T \in \mathcal{T}} H^{1/2}(\partial T)$$

as collections of element-local traces of  $H^1(\Omega)$ -functions. Indeed, this space is the image of the *skeletal (Dirichlet) trace operator*

$$\begin{aligned} \gamma_S = \gamma_S^0 : H^1(\Omega) &\rightarrow H^{1/2}(\Gamma_S) \\ \phi &\mapsto (\gamma_T^0 \phi)_{T \in \mathcal{T}}. \end{aligned}$$

It is clear that the values of the components  $(v_T)$  match across element interfaces. Also, the restriction  $\gamma_S|_{\mathcal{V}_{\mathcal{H}}}$  to the locally  $\mathcal{L}$ -harmonic functions yields an isomorphism between the spaces  $\mathcal{V}_{\mathcal{H}}$  and  $H^{1/2}(\Gamma_S)$ .

As in Section 2.1, we denote by  $\mathcal{H}_T : H^{1/2}(\partial T) \rightarrow \mathcal{H}(T)$  the local harmonic extension operator which maps  $g_T \in H^{1/2}(\partial T)$  to the unique solution  $u_T \in H^1(T)$  of the local variational problem

$$\gamma_T^0 u_T = g_T, \quad \mathcal{L}_T(u_T, v_T) = 0 \quad \forall v_T \in H_0^1(T).$$

It is easy to see that  $\mathcal{H}_T$  is bijective, with its inverse given by  $\gamma_T^0$ . We introduce the *skeletal harmonic extension operator*

$$\begin{aligned} \mathcal{H}_S : H^{1/2}(\Gamma_S) &\rightarrow \mathcal{V}_{\mathcal{H}}, \\ (\mathcal{H}_S v)|_T &:= \mathcal{H}_T(v_T) \quad \forall T \in \mathcal{T}. \end{aligned} \tag{4.4}$$

Clearly,  $\mathcal{H}_S$  is just the inverse of the bijection  $\gamma_S|_{\mathcal{V}_{\mathcal{H}}}$ . Similarly, with the subspace  $\mathcal{W}_0 \subset \mathcal{W} := H^{1/2}(\Gamma_S)$  given by

$$\begin{aligned} \mathcal{W}_0 &:= \left\{ (v_T)_{T \in \mathcal{T}} \mid \exists \varphi \in H_0^1(\Omega) \forall T \in \mathcal{T} : v_T = \gamma_T^0 \varphi \right\} \\ &= \{v \in H^{1/2}(\Gamma_S) : v|_{\partial\Omega} = 0\}, \end{aligned}$$

the operator  $\mathcal{H}_S$  is a bijection between  $\mathcal{W}_0$  and  $\mathcal{V}_{\mathcal{H},0}$ .

These bijections allow us to represent any piecewise  $\mathcal{L}$ -harmonic function by its skeletal traces, giving us the equivalent reformulation of (4.3): find  $u \in H^{1/2}(\Gamma_S)$  such that

$$\gamma_\Omega \mathcal{H}_S u = g, \quad \sum_{T \in \mathcal{T}} \langle \gamma_T^1 \mathcal{H}_T u, \gamma_T^0 \mathcal{H}_T v \rangle_{\partial T} = 0 \quad \forall v \in \mathcal{W}_0.$$

Noting that  $\gamma_T^0 \circ \mathcal{H}_T = \text{id}$  and recalling the definition of the *Dirichlet-to-Neumann map* (2.8), now applied locally to an element  $T$ , we arrive at the formulation: find  $u \in H^{1/2}(\Gamma_S)$  satisfying

$$u|_{\partial\Omega} = g, \quad \sum_{T \in \mathcal{T}} \langle S_T u_T, v_T \rangle_{\partial T} = 0 \quad \forall v \in \mathcal{W}_0. \tag{4.5}$$

We point out the similarity of this formulation to the relation (1.2) derived by Trefftz in his method. This is not surprising as the underlying idea of using (PDE-)harmonic trial functions presented in his work was also fundamental in our derivation.

We have arrived at (4.5) by nothing but equivalent reformulations of (4.1), which, in turn, Lemma 4.1.1 has shown to be equivalent to the standard variational formulation (2.4). We have therefore proved the following proposition.

**Proposition 4.1.2.** *Let  $g \in H^{1/2}(\partial\Omega)$ . Under the assumptions of Theorem 2.4.1, the variational problems to find  $u_\Omega \in H^1(\Omega)$  such that*

$$\gamma_\Omega u_\Omega = g, \quad \mathcal{L}(u_\Omega, v_\Omega) = 0 \quad \forall v_\Omega \in H_0^1(\Omega)$$

and  $u \in H^{1/2}(\Gamma_S)$  such that

$$u|_{\partial\Omega} = g, \quad \sum_{T \in \mathcal{T}} \langle S_T u_T, v_T \rangle_{\partial T} = 0 \quad \forall v \in \mathcal{W}_0$$

are both well-posed. They are equivalent in the sense that their unique solutions  $u_\Omega$  and  $u$  are related by

$$u = \gamma_S u_\Omega \quad \text{and} \quad u_\Omega = \mathcal{H}_S u.$$

We have treated only the pure Dirichlet boundary value problem above. Before we proceed, we point out that mixed boundary value problems, where Neumann data is prescribed on a part of the boundary, can be formulated as skeletal variational problems in an analogous way. Indeed, if the boundary of  $\Omega$  is split into two parts  $\Gamma_D$  and  $\Gamma_N$  with nonzero surface measure on which we prescribe Dirichlet data  $g_D \in H^{1/2}(\Gamma_D)$  and Neumann data  $g_N \in H^{-1/2}(\Gamma_N)$ , respectively, leading to the mixed boundary value problem

$$\begin{aligned} Lu(x) &= 0, & x \in \Omega, \\ \gamma_\Omega u(x) &= g_D(x), & x \in \Gamma_D, \\ \gamma_\Omega (A \nabla u(x)) \cdot n(x) &= g_N(x), & x \in \Gamma_N, \end{aligned}$$

then the corresponding skeletal variational formulation is easily seen to be: find  $u \in H^{1/2}(\Gamma_S)$  such that  $u|_{\Gamma_D} = g$  and

$$\sum_{T \in \mathcal{T}} \langle S_T u_T, v_T \rangle_{\partial T} = \int_{\Gamma_N} g_N v \, ds$$

for all  $v \in H^{1/2}(\Gamma_S)$  which vanish on  $\Gamma_D$ . This translates naturally into the discretization that follows later in this chapter, and thus the BEM-based FEM can be applied to problems of this type without difficulties. In order not to unnecessarily complicate the notation and analysis, we refrain from treating the mixed boundary value problem separately.

## 4.2 A mixed variational formulation

We introduce the space of elementwise Neumann traces,

$$\mathcal{Z} := \bigotimes_{T \in \mathcal{T}} H^{-1/2}(\partial T).$$

In contrast to the space  $\mathcal{W} = H^{1/2}(\Gamma_S)$ , whose members are globally continuous on the skeleton,  $\mathcal{Z}$  contains functions which are discontinuous and double-valued on element interfaces. In this space, we choose the auxiliary variable

$$t := (t_T)_{T \in \mathcal{T}} \in \mathcal{Z}, \quad t_T = V_T^{-1}(\frac{1}{2}I + K_T)u_T \quad \text{for } T \in \mathcal{T}.$$

Equivalently,  $t_T \in H^{-1/2}(\partial T)$  is determined by the local variational equation

$$\langle z_T, V_T t_T \rangle = \langle z_T, (\frac{1}{2}I + K_T)u_T \rangle \quad \forall z_T \in H^{-1/2}(\partial T).$$

Note that  $t_T = S_T u_T$  is just the Neumann trace belonging to  $u_T$ . With (3.5), we have  $S_T u_T = D_T u_T + (\frac{1}{2}I + K_T')t_T$ , and hence we can write the following equivalent mixed formulation for (4.5): find  $(u, t) \in \mathcal{X} := \mathcal{W} \times \mathcal{Z}$  such that

$$\begin{aligned} u|_{\partial\Omega} &= g, \\ d(u, v) + b(v, t) &= 0 \quad \forall v \in \mathcal{W}_0, \\ -b(u, z) + c(z, t) &= 0 \quad \forall z \in \mathcal{Z} \end{aligned}$$

with the bilinear forms

$$\begin{aligned} d(u, v) &= \sum_{T \in \mathcal{T}} \langle D_T u_T, v_T \rangle, \\ b(v, t) &= \sum_{T \in \mathcal{T}} \langle t_T, (\frac{1}{2}I + K_T)v_T \rangle, \\ c(z, t) &= \sum_{T \in \mathcal{T}} \langle z_T, V_T t_T \rangle. \end{aligned}$$

Here we assumed for simplicity that the problem is self-adjoint such that the operators  $K_T$  and  $K_T'$  are adjoint to each other. If this is not the case, the term  $b(v, t)$  in the first equation has to be modified accordingly.

With the combined bilinear form

$$\mathcal{A}((u, t), (v, z)) := d(u, v) + b(v, t) - b(u, z) + c(z, t)$$

and the space  $\mathcal{X}_0 := \mathcal{W}_0 \times \mathcal{Z}$ , we may write the above variational problem more compactly: find  $(u, t) \in \mathcal{X}$  such that

$$u|_{\partial\Omega} = g, \quad \mathcal{A}((u, t), (v, z)) = 0 \quad \forall (v, z) \in \mathcal{X}_0. \quad (4.6)$$

### 4.3 A BEM-based finite element method

In this section we derive the BEM-based FEM discretization of the skeletal variational formulation (4.5). Since we work with skeletal functions spaces which only incorporate boundary values of the involved functions on every element, it is natural to use a representation of the Dirichlet-to-Neumann map  $S_T$  in terms of boundary integral operators. For symmetric problems, we use symmetric approximations of the local Steklov-Poincaré operators in order to obtain a symmetric stiffness matrix.

In the following, we give first the discretization of the primal skeletal formulation and then of the mixed formulation.

#### 4.3.1 Discrete function spaces

Recall that the elements  $T \in \mathcal{T}$  are bounded Lipschitz polytopes. Let  $\mathcal{F}_T$  denote a triangulation of the element boundary  $\partial T$  into  $(d-1)$ -dimensional simplices, i.e., line segments for two-dimensional elements and triangles for three-dimensional elements. We will refer to these boundary simplices as *facets* in the following. We assume that the triangulations are *matching*, i.e., for two elements  $T$  and  $T'$  and facets  $f \in \mathcal{F}_T$  and  $f' \in \mathcal{F}_{T'}$ , we have  $f \cap f' \neq \emptyset \Leftrightarrow f = f' \in \mathcal{F}_T \cap \mathcal{F}_{T'}$ . In other words, faces from neighboring elements are either identical or do not intersect at all. This allows us to introduce a triangulation of the skeleton via  $\mathcal{F} := \bigcup_{T \in \mathcal{T}} \mathcal{F}_T$ .

On these boundary triangulations, we set up the boundary element spaces

$$\begin{aligned} \mathcal{W}_h &:= \left\{ v \in H^{1/2}(\Gamma_S) : v|_{\bar{f}} \in P^1(\bar{f}) \ \forall f \in \mathcal{F} \right\} \subset H^{1/2}(\Gamma_S), \\ \mathcal{Z}_{h,T} &:= \left\{ z \in L_2(\partial T) : z|_f \in P^0(f) \ \forall f \in \mathcal{F}_T \right\} \subset H^{-1/2}(\partial T), \\ \mathcal{Z}_h &:= \bigotimes_{T \in \mathcal{T}} \mathcal{Z}_{h,T} \subset \mathcal{Z}, \end{aligned}$$

where  $P^k(f)$  means the space of polynomials of total degree  $k$  on the facet  $f$ . Thus,  $\mathcal{W}_h$  contains piecewise linear, continuous skeletal functions,  $\mathcal{Z}_{h,T}$  contains piecewise constant, discontinuous functions on  $\partial T$ , and  $\mathcal{Z}_h$  is a broken space of piecewise constant functions. We also define  $\mathcal{W}_{h,0} := \mathcal{W}_h \cap \mathcal{W}_0$  as the subspace with homogeneous boundary conditions.

#### 4.3.2 Discretization of the primal formulation

Let us restate the skeletal variational formulation (4.5) derived in Section 4.1: we seek  $u \in H^{1/2}(\Gamma_S)$  such that

$$u|_{\partial\Omega} = g, \quad a(u, v) = 0 \quad \forall v \in \mathcal{W}_0 \quad (4.7)$$

with the bilinear form

$$a(u, v) := \sum_{T \in \mathcal{T}} \langle S_T u_T, v_T \rangle. \quad (4.8)$$

The solution of the boundary value problem (2.3) is then given by  $u_\Omega = \mathcal{H}_S u$ , the piecewise harmonic extension of the skeletal solution  $u$ .

In some instances, it is convenient to consider the homogenized version of (4.7). Observe that the given Dirichlet data  $g \in H^{1/2}(\partial\Omega)$  can always be extended to the skeleton. A simple, if not necessarily practical, such extension is given by  $\gamma_S \mathcal{H}_\Omega g$ . In 2D, a simple constructive extension can be done by setting all interior vertex values of the extension to 0 and extending linearly on all edges which do not lie on the boundary. We denote the extension again by  $g \in H^{1/2}(\Gamma_S)$  since there is no possibility of confusion. Homogenizing the above equation, we have  $u = u_0 + g$  with an  $u_0 \in \mathcal{W}_0$  which fulfils

$$a(u_0, v) = \langle F, v \rangle \quad \forall v \in \mathcal{W}_0 \quad (4.9)$$

with the linear functional

$$\langle F, v \rangle := \sum_{T \in \mathcal{T}} \langle -S_T g_T, v_T \rangle = -a(g, v).$$

For each element  $T \in \mathcal{T}$ , let

$$\tilde{S}_T : H^{1/2}(\partial T) \rightarrow H^{-1/2}(\partial T)$$

denote the BEM-approximated Steklov-Poincaré operator on the boundary of the element  $T$ , defined as in Section 3.3. The choice of the discrete space for approximating the Neumann traces which was left open there is now simply  $\mathcal{Z}_{h,T}$ .

Replacing  $S_T$  by  $\tilde{S}_T$  above, we get an approximate bilinear form,

$$\tilde{a}(u, v) := \sum_{T \in \mathcal{T}} \langle \tilde{S}_T u_T, v_T \rangle. \quad (4.10)$$

Restricting the trial and test functions to the space of piecewise linear skeletal functions  $\mathcal{W}_h$ , we obtain the discretized variational formulation: find  $u_h \in \mathcal{W}_h$  such that

$$u_h|_{\partial\Omega} = g, \quad \tilde{a}(u_h, v_h) = 0 \quad \forall v_h \in \mathcal{W}_{h,0}. \quad (4.11)$$

Again, we state also the homogenized formulation: with  $u_h = u_{h,0} + g$ , we seek  $u_{h,0} \in \mathcal{W}_{h,0}$  such that

$$\tilde{a}(u_{h,0}, v_h) = \langle \tilde{F}, v \rangle_h \quad \forall v_h \in \mathcal{W}_{h,0} \quad (4.12)$$

with the linear functional

$$\langle \tilde{F}, v \rangle := \sum_{T \in \mathcal{T}} \langle -\tilde{S}_T g_T, v_T \rangle = -\tilde{a}(g, v).$$

As basis functions for  $\mathcal{W}_h$ , we choose the skeletal nodal basis functions which are one in a designated vertex of the skeleton and zero in all others while being piecewise linear

on the skeletal facets. To assemble the stiffness matrix corresponding to (4.11), we only need an efficient method for computing the local stiffness matrices arising from  $\tilde{S}_T$ . These are standard BEM matrices, and any existing BEM software package can be leveraged for this task.

As we have pointed out in Section 3.3, if the partial differential operator is formally self-adjoint, then also the approximated Steklov-Poincaré operators  $\tilde{S}_T$  are self-adjoint. Thus, in this case, the resulting stiffness matrix is symmetric. Due to

$$\sum_T \langle \tilde{S}_T v_T, v_T \rangle \geq C \sum_T \langle S_T v_T, v_T \rangle = \mathcal{L}(\mathcal{H}_S v, \mathcal{H}_S v) \quad \forall v \in H^{1/2}(\Gamma_S),$$

which follows immediately from the spectral equivalence in Theorem 3.3.1, it is easy to see that this bilinear form  $\tilde{a}(\cdot, \cdot)$  is coercive on  $\mathcal{W}_{h,0}$  if the partial differential operator  $L$  is, and thus the stiffness matrix is positive definite in this case. Furthermore, note that the nodal skeletal basis functions have local support and thus, the stiffness matrix has non-zero entries only for such pairs of vertices which share an element. Just as in a standard finite element method, the stiffness matrix is thus sparse.

It is interesting to note that, in the case of the Laplace equation on a purely simplicial mesh,

- the piecewise harmonic trial functions with linear boundary data are just the piecewise linear functions,
- the space  $\mathcal{Z}_h$  of piecewise constant boundary functions can represent the Neumann derivatives of the piecewise linear functions exactly,
- the local Galerkin projections  $t_{h,T}(u_h)$  of the Neumann derivative  $t_T(u_h)$  (cf. Section 3.3) are thus just the identity, i.e.,  $t_{h,T} \equiv t_T$ , and therefore also  $\tilde{S}_T = S_T$  and  $\tilde{a}(\cdot, \cdot) = a(\cdot, \cdot)$ .

This means that in this special case, the scheme can be realized exactly and is equivalent to a standard nodal FEM with piecewise linear trial functions. Indeed, the resulting stiffness matrices from the BEM-based FEM and this standard FEM are then identical, up to quadrature errors. However, as soon as the partial differential operator  $L$  is not the Laplace operator, this is not true anymore, even on simplicial meshes, since then the piecewise  $\mathcal{L}$ -harmonic trial functions are no longer the affine linear functions.

### 4.3.3 Discretization of the mixed formulation

We discretize the variational formulation (4.6) by looking for a pair of unknowns

$$(u_h, t_h) \in \mathcal{X}_h := \mathcal{W}_h \times \mathcal{Z}_h \subset \mathcal{X}$$

such that

$$u_h|_{\partial\Omega} = g, \quad \mathcal{A}((u_h, t_h), (v_h, z_h)) = 0 \quad \forall (v_h, z_h) \in \mathcal{X}_{h,0}, \quad (4.13)$$

where  $\mathcal{X}_{h,0} := \mathcal{W}_{h,0} \times \mathcal{Z}_h \subset \mathcal{X}_0$ . In practice, the auxiliary variable  $t_h$  can be eliminated locally on each element by static condensation. To see this, fix an element  $T$  and chose a test function  $z_h$  which has an arbitrary component  $z_{h,T} \in \mathcal{Z}_{h,T}$  on  $T$  and is 0 for all the components associated with other elements. Testing with  $(0, z_h)$  in (4.13), we obtain the variational equation

$$\langle z_{h,T}, V_T t_{h,T} \rangle = \langle z_{h,T}, (\frac{1}{2}I + K_T)u_{h,T} \rangle,$$

which means that  $t_{h,T} = t_{h,T}(u_{h,T})$  with the notation  $t_{h,T}(\cdot)$  for the approximated Neumann traces introduced in (3.9). We can thus eliminate the second line from the mixed variational problem by inserting this choice for  $t_h$  in the first line and recover the primal discretized formulation (4.11), in which only the primal unknowns  $u_h$  enter. In this way, we obtain the same numerical scheme as for the primal scheme from Section 4.3.2, even though the variational formulation is now a mixed one.

The advantage of the mixed formulation is that, by subtracting (4.6) and (4.13), we obtain the Galerkin orthogonality

$$\mathcal{A}((u - u_h, t - t_h), (v_h, z_h)) = 0 \quad \forall (v_h, z_h) \in \mathcal{X}_{h,0}. \quad (4.14)$$

This will be of vital importance in Section 6.6.3, where we derive  $L_2$ -error estimates. By contrast, the discretization of the primal formulation does not yield a Galerkin orthogonality due to the approximation of the bilinear form.

We state also the homogenized form of (4.13): find  $(u_{h,0}, t_h) \in \mathcal{X}_0$  such that

$$\mathcal{A}((u_{h,0}, t_h), (v_h, z_h)) = -d(g, v) + b(g, z) \quad \forall (v_h, z_h) \in \mathcal{X}_0. \quad (4.15)$$

# Chapter 5

## Properties of polytopal elements

For the elements used in classical finite element methods, a wide array of analytical tools is available in the literature, among these trace theorems, approximation properties, inverse inequalities, and so on. All of these are typically proved under quite standard assumptions on the shape regularity of the elements. Most commonly, proofs proceed via the mapping principle by transforming some quantity of interest to a reference element, proving the desired estimate there and transforming the result back to the actual element. See, for instance, the classical book by Ciarlet [22] for many instances of this line of reasoning.

For the non-standard polytopal elements considered in this thesis, we cannot assume the existence of a reference element to which we can map our element since many different element shapes might occur in a single mesh. It is therefore not clear how to generalize the mapping principle to our setting, which means that the standard theory is not applicable.

In large parts, the goal of this chapter is to fill these gaps by providing important analytical tools for the polytopal elements used in the BEM-based FEM. In particular, we prove approximation properties and trace theorems for Lipschitz polytopes and make all estimates explicit in shape regularity parameters. Here, shape regularity is not defined via mapping to a reference element or angle conditions. Instead, we assume that every element can be triangulated by simplices and impose standard regularity conditions on these local element triangulations. Some of the estimates are proved in certain non-standard norms which are most suitable for our purposes.

Additionally, in Section 5.6, we provide two ways of constructing extension operators for polytopes whose operator norms may be explicitly bounded in terms of mesh parameters. We are mainly interested in such operators because they allow us to bound the contraction constants and other quantities related to the boundary integral equations on the elements, as shown in Section 3.4. In particular, in Section 5.6.2, we will see that by a particular construction, we can reuse a very similar notion of mesh regularity as that used in the remaining sections of this chapter.

Throughout this chapter, we assume that  $T$  satisfies the following conditions.

**Assumption 5.0.1.** *We assume that  $T \subset \mathbb{R}^d$ ,  $d = 2$  or  $3$ , is an open Lipschitz polytope and that there exists a conforming simplicial triangulation  $\Xi$  of  $T$  which is shape-regular according to Definition 2.5.1 with constants  $c_1$ ,  $\bar{c}_1$ ,  $c_2$ , and  $\bar{c}_2 > 0$ . Furthermore, we*

assume that the number of simplices in the triangulation  $\Xi$  is bounded by a small integer,  $\#\Xi \leq N_\Xi$ .

By  $\mathcal{F}$ , we denote the collection of boundary facets of simplices  $\tau \in \Xi$  which lie on  $\partial T$ , such that  $\mathcal{F}$  is a triangulation of  $\partial T$ . From Assumption 5.0.1, it follows that this boundary triangulation consists of at most  $N_{\mathcal{F}}$  facets with some small integer  $N_{\mathcal{F}}$ . The facets are  $(d-1)$ -dimensional simplices, i.e., line segments or triangles. Note that for every boundary facet  $f \in \mathcal{F}$  of  $T$ , there exists exactly one simplex in  $\Xi$  which has  $f$  as one of its facets, and we will denote it by  $\tau_f \in \Xi$ .

## 5.1 Norms

### 5.1.1 The energy norm

Because we use harmonic extensions heavily, it is natural to work with norms which measure the energy of the harmonic extension. For simplicity, we consider here only the case  $L = -\Delta$ , such that  $\mathcal{H}_T$  is the classical harmonic extension operator. Thus, we equip the trace space  $H^{1/2}(\partial T)$  with the seminorm and norm

$$\begin{aligned} |v_T|_{H^{1/2}(\partial T)} &:= |\mathcal{H}_T v_T|_{H^1(T)} = \inf_{\substack{\phi \in H^1(T) \\ \gamma_T^0 \phi = v_T}} |\phi|_{H^1(T)}, \\ \|v_T\|_{H^{1/2}(\partial T)}^2 &:= (\text{diam } T)^{-2} \|\mathcal{H}_T v_T\|_{L_2(T)}^2 + |\mathcal{H}_T v_T|_{H^1(T)}^2. \end{aligned}$$

The term  $(\text{diam } T)^{-2}$  here serves to make both components scale identically with respect to  $\text{diam } T$  as per Lemma 2.1.1. The norm  $\|\cdot\|_{H^{1/2}(\partial T)}$  induces the associated dual norm

$$\|w\|_{H^{-1/2}(\partial T)} := \sup_{v \in H^{1/2}(\partial T)} \frac{\langle w, v \rangle}{\|v\|_{H^{1/2}(\partial T)}} \quad \forall w \in H^{-1/2}(\partial T)$$

on the dual space  $H^{-1/2}(\partial T)$  of  $H^{1/2}(\partial T)$ .

The close relationship of the seminorm to the skeletal variational problem (4.5) is evident from the fact that it is induced by the bilinear form  $\langle S_T \cdot, \cdot \rangle$ . Indeed, for all  $v \in H^{1/2}(\partial T)$ ,

$$\begin{aligned} |v|_{S_T}^2 = \langle S_T v, v \rangle &= \langle \gamma_T^1(\mathcal{H}_T v), \gamma_T^0(\mathcal{H}_T v) \rangle \\ &\stackrel{(2.7)}{=} \int_T \nabla(\mathcal{H}_T v) \cdot \nabla(\mathcal{H}_T v) \, dx \\ &= |\mathcal{H}_T v|_{H^1(T)}^2 = |v|_{H^{1/2}(\partial T)}^2. \end{aligned} \tag{5.1}$$

### 5.1.2 The Sobolev-Slobodeckii norm

For technical reasons, we will also need the Sobolev-Slobodeckii seminorm in addition to the harmonic extension norm introduced above. For a smooth  $(d-1)$ -dimensional surface  $f$  embedded in  $\mathbb{R}^d$ , in particular for a boundary facet  $f \in \mathcal{F}$ , we define the seminorm

$$|u|_{H^{\sim 1/2}(f)}^2 := \int_f \int_f \frac{(u(x) - u(y))^2}{|x - y|^d} ds_x ds_y. \quad (5.2)$$

On the space

$$H_{\text{pw}}^{1/2}(\partial T) := \left\{ v \in L_2(\partial T) : v|_f \in H^{1/2}(f) \quad \forall f \in \mathcal{F} \right\},$$

of piecewise  $H^{1/2}$  functions on  $\partial T$ , this gives rise to the piecewise Sobolev-Slobodeckii seminorm

$$|u|_{H_{\text{pw}}^{1/2}(\partial T)}^2 := \sum_{f \in \mathcal{F}} |u|_{H^{\sim 1/2}(f)}^2.$$

## 5.2 Transformation properties

At this point, we recall the definition of a shape-regular simplicial mesh, Definition 2.5.1, the notation  $F_\tau : \mathbb{R}^d \rightarrow \mathbb{R}^d$  for the affine mapping from the unit simplex  $\Delta_d$  to an element  $\tau$ , and  $J_\tau = \nabla F_\tau \in \mathbb{R}^{d \times d}$  for its Jacobian. From the regularity conditions (2.14) and (2.15), we easily derive the property

$$c_2(\text{diam } \tau) |\xi| \leq |J_\tau \xi| \leq \bar{c}_2(\text{diam } \tau) |\xi| \quad \forall \xi \in \mathbb{R}^d, \quad (5.3)$$

which describes how lengths transform under  $F_\tau$ .

In the following we show that regularity of  $\Xi$  implies regularity of the boundary triangulation  $\mathcal{F}$ .

**Lemma 5.2.1.** *Let  $\Xi$  be a shape-regular simplicial mesh. Then for every facet  $f \in \mathcal{F}$  and every simplex  $\tau = \tau_f \in \Xi$  with  $f \subset \partial\tau$ , we have*

$$c_2 \text{diam } \tau \leq \text{diam } f \leq \text{diam } \tau, \quad (5.4)$$

$$\frac{c_1}{2c_2} (\text{diam } \tau)^2 \leq |f| \leq \frac{1}{2} (\text{diam } \tau)^2, \quad (5.5)$$

where  $|f|$  denotes the surface area of the facet  $f$ . The second statement (5.5) is only applicable for  $d = 3$  as in 2D,  $\text{diam } f = |f|$ .

*Proof.* The estimate  $\text{diam } f \leq \text{diam } \tau$  is trivial as  $\bar{f} \subset \bar{\tau}$ . In 3D, we easily get from this that the area of the triangle  $f$  satisfies

$$|f| \leq \frac{1}{2} (\text{diam } f)^2 \leq \frac{1}{2} (\text{diam } \tau)^2,$$

and thus the upper bounds are proved.

For the lower bounds, let  $\{\xi_1, \dots, \xi_{d+1}\}$  denote the vertices of the unit simplex  $\Delta_d$ , and  $x_i = F_\tau(\xi_i)$ ,  $i = 1, \dots, d+1$ , the vertices of  $\tau$ . Clearly, the diameter of  $f$  is the length of an edge, say  $(x_i, x_j)$ , of  $\tau$ . We have

$$\begin{aligned} \text{diam } f = |x_i - x_j| &= |F_\tau(\xi_i) - F_\tau(\xi_j)| = \\ &= |J_\tau(\xi_i - \xi_j)| \stackrel{(5.3)}{\geq} \underline{c}_2 \text{diam } \tau |\xi_i - \xi_j|. \end{aligned}$$

Since  $|\xi_i - \xi_j|$  is the length of an edge of the unit simplex, it is clear that  $|\xi_i - \xi_j| \geq 1$ , which finishes the proof of (5.4).

For the lower area bound in 3D, let  $(x_i, x_j, x_k)$  be the vertices of the triangle  $f$ . With  $y_1 := x_j - x_i$  and  $y_2 := x_k - x_i$ , the area of the triangle is given by  $|f| = \frac{1}{2} |y_1 \times y_2|$ . Furthermore,  $\hat{f} := F_\tau^{-1}(f)$  is a face of  $\Delta_3$ , and we have  $|\hat{f}| = \frac{1}{2} |\eta_1 \times \eta_2|$  with

$$\eta_1 = \xi_j - \xi_i = F_\tau^{-1}(x_j) - F_\tau^{-1}(x_i) = J_\tau^{-1}(x_j - x_i) = J_\tau^{-1}y_1,$$

and analogously  $\eta_2 = J_\tau^{-1}y_2$ . Thus we may estimate

$$\begin{aligned} \frac{1}{2} = |\hat{f}| &= \frac{1}{2} |\eta_1 \times \eta_2| = \frac{1}{2} |J_\tau^{-1}y_1 \times J_\tau^{-1}y_2| \\ &\stackrel{(*)}{=} \frac{1}{2} \left| \det J_\tau^{-1} \right| \left| J_\tau^\top(y_1 \times y_2) \right| \leq \frac{1}{2} \underline{c}_1^{-1} (\text{diam } \tau)^{-3} \bar{c}_2 (\text{diam } \tau)^2 |f|, \end{aligned}$$

where we have used that  $\det(J_\tau^{-1}) = (\det J_\tau)^{-1}$  and  $\|J_\tau^\top\|_{\ell_2} = \|J_\tau\|_{\ell_2}$ . The identity marked with (\*) stems from the following elementary property of the cross product that can easily be checked by direct calculation: for any non-singular matrix  $A \in \mathbb{R}^{3 \times 3}$ ,

$$Ay_1 \times Ay_2 = (\det A)A^{-\top}(y_1 \times y_2). \quad \square$$

We also need some norm scaling relations for transforming functions to and from the unit simplex.

**Lemma 5.2.2.** *Let  $f$  be a facet of a simplex  $\tau$  from a shape-regular triangulation and  $\hat{f} := F_\tau^{-1}(f)$  the corresponding facet on the unit simplex  $\Delta_d$ .*

(a) *Let  $\phi \in H^{1/2}(f)$  and denote by  $\hat{\phi} = \phi \circ F_\tau$  the pullback of  $\phi$  to  $\hat{f}$ . Then*

$$|\phi|_{H^{1/2}(f)} \leq \bar{c} (\text{diam } \tau)^{d/2-1} |\hat{\phi}|_{H^{1/2}(\hat{f})} \quad (5.6)$$

*with the Sobolev-Slobodeckii seminorm as defined in (5.2).*

(b) *Let  $u \in H^1(\tau)$  and denote by  $\hat{u} = u \circ F_\tau$  the pullback of  $u$  to  $\Delta_d$ . Then*

$$\underline{c} (\text{diam } \tau)^{d/2-1} |\hat{u}|_{H^1(\Delta_d)} \leq |u|_{H^1(\tau)} \leq \bar{c} (\text{diam } \tau)^{d/2-1} |\hat{u}|_{H^1(\Delta_d)}. \quad (5.7)$$

*Above,  $\underline{c}$  and  $\bar{c}$  are generic positive constants which depend only on the regularity parameters from Definition 2.5.1.*

*Proof.* Statement (b) is shown by standard transformation arguments from finite element analysis. For instance, [22, Theorem 3.1.2] states that there is a constant  $C = C(k, d)$  such that for any Sobolev function  $u \in W_p^k(\tau)$  with  $k \in \mathbb{N}_0$  and  $1 \leq p \leq \infty$ ,

$$C^{-1} \|J\|_{\ell_2}^{-k} |\det J|^{1/p} |\hat{u}|_{W_p^k(\Delta_d)} \leq |u|_{W_p^k(\tau)} \leq C \|J^{-1}\|_{\ell_2}^k |\det J|^{1/p} |\hat{u}|_{W_p^k(\Delta_d)}$$

from which together with the regularity conditions (2.13)-(2.15) the statement follows.

We now prove (a) in the case  $d = 3$ . Let  $F_f, F_{\hat{f}} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  denote affine mappings such that  $F_f(\Delta_2) = f$ ,  $F_{\hat{f}}(\Delta_2) = \hat{f}$ , and  $F_f = F_\tau \circ F_{\hat{f}}$ . Note that  $\left| \frac{\partial F_f}{\partial x_1} \times \frac{\partial F_f}{\partial x_2} \right| = 2|f|$ . For any function  $\psi \in L_1(f)$ , we see that

$$\begin{aligned} \int_f \psi(x) ds_x &= 2|f| \int_{\Delta_2} \psi(F_f(\xi)) d\xi = 2|f| \int_{\Delta_2} \psi(F_\tau(F_{\hat{f}}(\xi))) d\xi \\ &= \frac{|f|}{|\hat{f}|} \int_{\hat{f}} \psi(F_\tau(x)) ds_x = \frac{|f|}{|\hat{f}|} \int_{\hat{f}} \hat{\psi}(x) ds_x. \end{aligned}$$

For the Sobolev-Slobodeckii seminorm (5.2), the above identity gives us

$$\begin{aligned} |\phi|_{H^1(f)}^2 &= \int_f \int_f \frac{|\phi(x) - \phi(y)|^2}{|x - y|^3} ds_x ds_y \\ &= \left( \frac{|f|}{|\hat{f}|} \right)^2 \int_{\hat{f}} \int_{\hat{f}} \frac{|\hat{\phi}(\xi) - \hat{\phi}(\eta)|^2}{|J_\tau(\xi - \eta)|^3} ds_\xi ds_\eta. \end{aligned}$$

Using the regularity relations (5.5) and (5.3) we obtain

$$|\phi|_{H^1(f)}^2 \leq c_2^{-3} \left( \frac{(\text{diam } \tau)^2}{2|\hat{f}|} \right)^2 (\text{diam } \tau)^{-3} \int_{\hat{f}} \int_{\hat{f}} \frac{|\hat{\phi}(\xi) - \hat{\phi}(\eta)|^2}{|\xi - \eta|^3} ds_\xi ds_\eta.$$

Noting finally that  $|\hat{f}| \geq \frac{1}{2}$ , we get (5.6).

It remains to show (a) for  $d = 2$ . We observe that in this case  $f$  and  $\hat{f}$  are just line segments which can be mapped to each other by an affine mapping which consists just of a rotation and a uniform scaling. Therefore, for any  $\psi \in L_1(f)$ , it is easy to see that again

$$\int_f \psi ds = \frac{|f|}{|\hat{f}|} \int_{\hat{f}} \hat{\psi} ds.$$

Applying this identity to the definition (5.2) of the Sobolev-Slobodeckii seminorm and observing that the distance of the coordinates too scales with the same factor,  $|x - y| = \frac{|f|}{|\hat{f}|} |\xi - \eta|$ , we find that

$$|\phi|_{H^1(f)} = |\hat{\phi}|_{H^1(\hat{f})},$$

and thus the statement is trivially true.  $\square$

### 5.3 Trace inequalities

In this section we derive trace inequalities for  $T$  with constants which depend solely on the regularity parameters of its triangulation. First we consider a single simplex  $\tau$  with associated Dirichlet trace operator  $\gamma_\tau^0 = \gamma_\tau : H^1(\tau) \rightarrow H^{1/2}(\partial\tau)$ .

**Lemma 5.3.1.** *For a simplex  $\tau$  from a shape-regular triangulation and one of its facets,  $f$ , we have the Dirichlet trace inequality*

$$|\gamma_\tau u|_{H^{1/2}(f)} \leq c_\gamma |u|_{H^1(\tau)} \quad \forall u \in H^1(\tau) \quad (5.8)$$

with a trace constant  $c_\gamma > 0$  which depends solely on the regularity parameters.

*Proof.* By a standard embedding argument, there exists a fixed constant  $\hat{c}_\gamma > 0$  such that for every facet  $\hat{f}$  of the unit simplex  $\Delta_d$ , we have

$$|\gamma_{\Delta_d} u|_{H^{1/2}(\hat{f})} \leq \hat{c}_\gamma |u|_{H^1(\Delta_d)} \quad \forall u \in H^1(\Delta_d), \quad (5.9)$$

with the trace operator  $\gamma_{\Delta_d} : H^1(\Delta_d) \rightarrow H^{1/2}(\partial\Delta_d)$ . Using the transformation relations from Lemma 5.2.2, we obtain

$$\begin{aligned} |\gamma_\tau u|_{H^{1/2}(f)} &\stackrel{(5.6)}{\leq} \bar{c} (\text{diam } \tau)^{d/2-1} |\gamma_{\Delta_3} \hat{u}|_{H^{1/2}(\hat{f})} \\ &\stackrel{(5.9)}{\leq} \hat{c}_\gamma \bar{c} (\text{diam } \tau)^{d/2-1} |\hat{u}|_{H^1(\Delta_3)} \\ &\stackrel{(5.7)}{\leq} \hat{c}_\gamma \bar{c} \bar{c}^{-1} |u|_{H^1(\tau)}. \quad \square \end{aligned}$$

This result extends straightforwardly to the piecewise Sobolev-Slobodeckii seminorm on the boundary of a polytopal element.

**Lemma 5.3.2.** *If the element  $T$  has a shape-regular triangulation, then*

$$|\gamma_T u|_{H^{1/2}_{\text{pw}}(\partial T)} \leq 2 c_\gamma |u|_{H^1(T)} \quad \forall u \in H^1(T) \quad (5.10)$$

with the trace constant  $c_\gamma$  from Lemma 5.3.1.

*Proof.* We fix  $u \in H^1(T)$  and calculate

$$|\gamma_T u|_{H^{1/2}_{\text{pw}}(\partial T)}^2 = \sum_{f \in \mathcal{F}} |\gamma_{\tau_f} u|_{H^{1/2}(f)}^2 \stackrel{(5.8)}{\leq} c_\gamma^2 \sum_{f \in \mathcal{F}} |u|_{H^1(\tau_f)}^2.$$

In 2D, every triangle  $\tau_f$  has three edges, and in 3D, every tetrahedron  $\tau_f$  has four boundary triangles. Thus, every  $\tau \in \Xi$  occurs at most four times in the rightmost sum. Thus we may further estimate

$$|\gamma_T u|_{H^{1/2}_{\text{pw}}(\partial T)}^2 \leq 4 c_\gamma^2 \sum_{\tau \in \Xi} |u|_{H^1(\tau)}^2 = 4 c_\gamma^2 |u|_{H^1(T)}^2. \quad \square$$

With this result, we are able to prove a Neumann trace inequality for functions which are sufficiently smooth such that the classical definition of the (co-)normal derivative applies.

**Theorem 5.3.3** (Neumann trace inequality). *Let  $T$  satisfy Assumption 5.0.1 and assume  $A \in \mathbb{R}^{d \times d}$  with  $\|A\|_{\ell_2} \leq \bar{\alpha} \in \mathbb{R}$ . Then, for all  $u \in H^2(T)$ , the estimate*

$$|\gamma_T^1 u|_{H_{\sim \text{pw}}^{1/2}(\partial T)} = |\gamma_T^0(A \nabla u) \cdot n|_{H_{\sim \text{pw}}^{1/2}(\partial T)} \leq C \bar{\alpha} |u|_{H^2(T)}$$

holds, where the constant  $C$  depends solely on the constants from Assumption 5.0.1,

*Proof.* On every boundary facet  $f \in \mathcal{F}$ , there is a uniquely defined and constant outwards normal vector  $n_f \in \mathbb{R}^d$  of unit length. On a single facet  $f \in \mathcal{F}$  lying on the simplex  $\tau$ , by using the triangle inequality and then the Cauchy-Schwarz inequality, we get

$$\begin{aligned} |\gamma_\tau^1 u|_{H_{\sim}^{1/2}(f)} &= |(\gamma_\tau A \nabla u) \cdot n_f|_{H_{\sim}^{1/2}(f)} = \left| \sum_{k=1}^d (\gamma_\tau \nabla u)_k (A^\top n_f)_k \right|_{H_{\sim}^{1/2}(f)} \\ &\leq \sum_{k=1}^d |(A^\top n_f)_k| |(\gamma_\tau \nabla u)_k|_{H_{\sim}^{1/2}(f)} \\ &\leq |A^\top n_f| \left( \sum_{k=1}^d |(\gamma_\tau \nabla u)_k|_{H_{\sim}^{1/2}(f)}^2 \right)^{1/2} \\ &\leq \bar{\alpha} \left( \sum_{k=1}^d \left| \gamma_\tau \frac{\partial u}{\partial x_k} \right|_{H_{\sim}^{1/2}(f)}^2 \right)^{1/2}. \end{aligned}$$

With this we obtain that on the entire boundary,

$$\begin{aligned} |\gamma_T^1 u|_{H_{\sim \text{pw}}^{1/2}(\partial T)}^2 &= \sum_{f \in \mathcal{F}} |\gamma_{\tau_f}^1 u|_{H_{\sim}^{1/2}(f)}^2 \leq \bar{\alpha}^2 \sum_{f \in \mathcal{F}} \sum_{k=1}^d \left| \gamma_{\tau_f} \frac{\partial u}{\partial x_k} \right|_{H_{\sim}^{1/2}(f)}^2 \\ &= \bar{\alpha}^2 \sum_{k=1}^d \left| \gamma_T \frac{\partial u}{\partial x_k} \right|_{H_{\sim \text{pw}}^{1/2}(\partial T)}^2 \stackrel{(5.10)}{\leq} 4 \bar{\alpha}^2 c_\gamma^2 \sum_{k=1}^d \left| \frac{\partial u}{\partial x_k} \right|_{H^1(T)}^2 = 4 \bar{\alpha}^2 c_\gamma^2 |u|_{H^2(T)}^2. \quad \square \end{aligned}$$

## 5.4 An auxiliary harmonic extension norm

For our final approximation result, we will make use of a more general version of the norm defined via the harmonic extension, namely one which is defined on arbitrary parts of the surface. This first requires a generalization of the harmonic extension operator. We again restrict ourselves to  $L = -\Delta$ . For any Lipschitz domain  $D$  and some surface

component  $t \subseteq \partial D$  with positive measure, we define

$$\begin{aligned} \mathcal{H}_{t \rightarrow D} : H^{1/2}(t) &\rightarrow H^1(D), \\ u &\mapsto \arg \min_{\substack{\phi \in H^1(D) \\ \phi|_t = u}} |\phi|_{H^1(D)}. \end{aligned}$$

By Theorem 2.1.2 (Friedrichs inequality) and Lemma 2.2.1 (Lax-Milgram), it is easy to show that the operator is well-defined. The previously introduced harmonic extension operator may be seen as a special case of this definition:  $\mathcal{H}_T = \mathcal{H}_{\partial T \rightarrow T}$ . With this notation, we define a seminorm on  $H^{1/2}(t)$  given by

$$|u|_{H^{1/2}(t,D)} := |\mathcal{H}_{t \rightarrow D} u|_{H^1(D)} = \inf_{\substack{\phi \in H^1(D) \\ \phi|_t = u}} |\phi|_{H^1(D)} \quad \forall u \in H^{1/2}(t). \quad (5.11)$$

Again, this may be viewed as a generalization of the previously introduced energy seminorm  $|\cdot|_{H^{1/2}(\partial T)} = |\cdot|_{H^{1/2}(\partial T, T)}$ .

The following lemma gives some indication of the monotonic behavior of the new seminorm when either the domain into which it extends or the surface component on which it is defined is restricted.

**Lemma 5.4.1.** *Let  $D' \subseteq D$  be Lipschitz domains and  $t' \subseteq t \subseteq \partial D' \cap \partial D$  surface components with positive measure. Then, for every  $v \in H^{1/2}(t)$ , we have*

$$|v|_{H^{1/2}(t,D')} \leq |v|_{H^{1/2}(t,D)}, \quad (5.12)$$

$$|v|_{H^{1/2}(t',D)} \leq |v|_{H^{1/2}(t,D)}. \quad (5.13)$$

*Proof.* We observe that

$$|\mathcal{H}_{t \rightarrow D'} v|_{H^1(D')} \leq |\mathcal{H}_{t \rightarrow D} v|_{H^1(D')} \leq |\mathcal{H}_{t \rightarrow D} v|_{H^1(D)},$$

where the first inequality holds because of the energy-minimizing property of the harmonic extension. This proves the first statement.

For the proof of (5.13), we observe that because of  $t' \subseteq t$ , it is clear that

$$\{u \in H^1(D) : u|_{t'} = v\} \supseteq \{u \in H^1(D) : u|_t = v\}.$$

These are just the sets over which we minimize in the definition (5.11), and thus the minimum that is attained over the left set is less or equal to that over the right one. This proves the second statement.  $\square$

It is of interest to know how this seminorm relates to the previously introduced Sobolev-Slobodeckii seminorm. For our purposes, the following simple result will suffice.

**Lemma 5.4.2.** *Let  $\tau \in \Xi$  be a simplex from a shape-regular triangulation, and let  $f \subset \partial\tau$  be one of its facets. For every  $v \in H^{1/2}(f)$ , we have*

$$|v|_{H^{1/2}(f)} \leq C |v|_{H^{1/2}(f,\tau)} \quad (5.14)$$

with a constant  $C$  that depends solely on the regularity parameters.

*Proof.* Using the trace inequality for a regular simplex, Lemma 5.3.1, we get

$$|v|_{H^{1/2}(f)} = |\gamma_\tau \mathcal{H}_{f \rightarrow \tau} v|_{H^{1/2}(f)} \stackrel{(5.8)}{\leq} c_\gamma |\mathcal{H}_{f \rightarrow \tau} v|_{H^1(\tau)} = c_\gamma |v|_{H^{1/2}(f,\tau)}. \quad \square$$

We now return to the polytopal element  $T$ . For  $u \in H_{\text{pw}}^{1/2}(\partial T)$ , we define the seminorm

$$|u|_{H_{\text{pw}}^{1/2}(\partial T)}^2 := \sum_{f \in \mathcal{F}} |u|_{H^{1/2}(f,\tau_f)}^2.$$

This norm is defined for the larger space of broken  $H^{1/2}$ -functions. If, however,  $u \in H^{1/2}(\partial T)$ , then by applying (5.12) and (5.13) we immediately obtain

$$|u|_{H_{\text{pw}}^{1/2}(\partial T)} \leq \sqrt{N_{\mathcal{F}}} |u|_{H^{1/2}(\partial T, T)} = \sqrt{N_{\mathcal{F}}} |u|_{H^{1/2}(\partial T)}. \quad (5.15)$$

## 5.5 Approximation properties

We now have all the tools at hand to study approximation properties for piecewise constant boundary functions on  $\partial T$ . We follow quite closely the approach by Steinbach [113], but use our particular norms and take care to track the dependencies of all constants.

Let

$$\mathcal{Z}_h := \{v \in L_2(\partial T) : v|_f \equiv \text{const.} \quad \forall f \in \mathcal{F}\}$$

denote the space of piecewise constant functions on  $\partial T$ . We introduce the  $L_2$ -projector  $Q_h : L_2(\partial T) \rightarrow \mathcal{Z}_h$  given by the variational problem

$$\langle Q_h u, v_h \rangle_{L_2(\partial T)} = \langle u, v_h \rangle_{L_2(\partial T)} \quad \forall v_h \in \mathcal{Z}_h$$

which is uniquely solvable for any given  $u \in L_2(\partial T)$ , as is easy to show by Lemma 2.2.1 (Lax-Milgram). It is easy to see that the values of the projection are just the means over the facets,

$$(Q_h u)|_f \equiv \frac{1}{|f|} \int_f u(y) ds_y \quad \forall f \in \mathcal{F}. \quad (5.16)$$

**Lemma 5.5.1.** *Let  $\Xi$  be a shape-regular triangulation of  $T$  and  $f \in \mathcal{F}$  a boundary facet. For  $u \in H_{\text{pw}}^{1/2}(\partial T)$ , we have the error estimates*

$$\begin{aligned} \|u - Q_h u\|_{L_2(f)} &\leq C (\text{diam } f)^{1/2} |u|_{H^{1/2}(f)}, \\ \|u - Q_h u\|_{L_2(\partial T)} &\leq C (\text{diam } T)^{1/2} |u|_{H_{\text{pw}}^{1/2}(\partial T)} \end{aligned} \quad (5.17)$$

with a constant  $C$  which depends solely on the regularity parameters.

*Proof.* Because of (5.16), we have

$$u(x) - Q_h u(x) = \frac{1}{|f|} \int_f [u(x) - u(y)] ds_y \quad \text{for } x \in f.$$

Squaring this relation and using the Cauchy-Schwarz inequality yields

$$\begin{aligned} |u(x) - Q_h u(x)|^2 &= \frac{1}{|f|^2} \left( \int_f [u(x) - u(y)] ds_y \right)^2 \\ &= \frac{1}{|f|^2} \left( \int_f \frac{[u(x) - u(y)]}{|x - y|^{d/2}} |x - y|^{d/2} ds_y \right)^2 \\ &\leq \frac{1}{|f|^2} \int_f \frac{[u(x) - u(y)]^2}{|x - y|^d} ds_y \int_f |x - y|^d ds_y \\ &\leq (\text{diam } f)^d \frac{1}{|f|} \int_f \frac{[u(x) - u(y)]^2}{|x - y|^d} ds_y. \end{aligned}$$

By integrating over  $f$ , we obtain the estimate

$$\|u - Q_h u\|_{L_2(f)}^2 \leq \frac{(\text{diam } f)^d}{|f|} |u|_{H^{1/2}(f)}.$$

In 2D, the first statement follows with  $\text{diam } f = |f|$ . In 3D, we estimate  $|f|$  from below in terms of  $\text{diam } f$  using the regularity conditions (5.5) and (5.4) to see that  $C(\text{diam } f)^2 \leq |f|$ . The second statement follows by summing up over all  $f \in \mathcal{F}$  and using that  $\text{diam } f \leq \text{diam } T$ .  $\square$

Using the above result, we can prove an approximation property for piecewise  $H^{1/2}$ -functions by piecewise constant functions in the  $H^{-1/2}$ -norm using an Aubin-Nitsche duality argument.

**Theorem 5.5.2.** *Let  $T$  satisfy Assumption 5.0.1. For all  $w \in H_{\text{pw}}^{1/2}(\partial T)$ , we have the error estimate*

$$\|w - Q_h w\|_{H^{-1/2}(\partial T)} \leq C \text{diam } T |w|_{H_{\text{pw}}^{1/2}(\partial T)}, \quad (5.18)$$

where the constant  $C$  depends solely on the constants from Assumption 5.0.1.

*Proof.* By the definition of the dual norm and of the  $L_2$ -projection  $Q_h$ , and per the Cauchy-Schwarz inequality, we have

$$\begin{aligned} \|w - Q_h w\|_{H^{-1/2}(\partial T)} &= \sup_{v \in H^{1/2}(\partial T)} \frac{\langle w - Q_h w, v \rangle_{L_2(\partial T)}}{\|v\|_{H^{1/2}(\partial T)}} \\ &= \sup_{v \in H^{1/2}(\partial T)} \frac{\langle w - Q_h w, v - Q_h v \rangle_{L_2(\partial T)}}{\|v\|_{H^{1/2}(\partial T)}} \\ &\leq \|w - Q_h w\|_{L_2(\partial T)} \sup_{v \in H^{1/2}(\partial T)} \frac{\|v - Q_h v\|_{L_2(\partial T)}}{\|v\|_{H^{1/2}(\partial T)}}. \end{aligned}$$

We estimate  $\|w - Q_h w\|_{L_2(\partial T)}$  using (5.17). For  $\|v - Q_h v\|_{L_2(\partial T)}$ , we again use (5.17) and then further estimate using Lemma 5.4.2,

$$\begin{aligned}
\|v - Q_h v\|_{L_2(\partial T)} &\leq C (\text{diam } T)^{1/2} |v|_{H_{\sim \text{pw}}^{1/2}(\partial T)} \\
&= C (\text{diam } T)^{1/2} \left( \sum_{f \in \mathcal{F}} |v|_{H_{\sim}^{1/2}(f)}^2 \right)^{1/2} \\
&\stackrel{(5.14)}{\leq} C (\text{diam } T)^{1/2} \left( \sum_{f \in \mathcal{F}} |v|_{H^{1/2}(f, \tau_f)}^2 \right)^{1/2} \\
&= C (\text{diam } T)^{1/2} |v|_{H_{\text{pw}}^{1/2}(\partial T)} \\
&\stackrel{(5.15)}{\leq} C \sqrt{N_{\mathcal{F}}} (\text{diam } T)^{1/2} |v|_{H^{1/2}(\partial T)}.
\end{aligned}$$

Since we assumed that  $N_{\mathcal{F}}$  is a uniform, small bound on the number of boundary facets per element, we may subsume it into the generic constant  $C$ . Combined, these estimates yield the statement.  $\square$

## 5.6 Explicit extension results

As we have seen in Section 3.4, explicitly bounding the coercivity constants and norms of boundary integral operators can be reduced to explicitly bounding certain extension operators as well as the Poincaré constant  $C_P(T)$ . The characterization of extension domains for Sobolev spaces, i.e., domains  $D \subset \mathbb{R}^d$  for which there exists a linear and bounded extension operator  $E : W_p^k(D) \rightarrow W_p^k(\mathbb{R}^d)$  for all  $1 \leq p \leq \infty$  and  $k \in \mathbb{N}$ , is a classical problem. For instance, it is known that every Lipschitz domain is an extension domain [21, 110]. Many results in this field, however, only give existence results without providing a way of bounding the norm of an extension operator in terms of geometric properties of the extension domain. In the following, we describe two constructions of extension operators which allow us to obtain such explicit bounds.

### 5.6.1 The Jones extension operator

Jones [73] gives sufficient conditions for a domain to be an extension domain in terms of a certain twisted cone condition. For finitely connected domains in two dimensions, he even shows that these conditions are necessary and sufficient and thus characterize such extension domains completely. The following definition gives the crucial geometric parameter in his construction; in fact, it is the special case of a  $(\varepsilon, \infty)$ -domain in his notation.

**Definition 5.6.1** (Jones [73]). A bounded and connected set  $D \subset \mathbb{R}^d$  is called a *uniform domain* if there exists a constant  $C_U(D) > 0$  such that any pair of points  $x_1 \in D$  and  $x_2 \in D$  can be joined by a rectifiable curve

$$\gamma(t) : [0, 1] \rightarrow D \quad \text{with} \quad \gamma(0) = x_1 \quad \text{and} \quad \gamma(1) = x_2,$$

such that the Euclidean arc length of  $\gamma$  is bounded by  $C_U(D) |x_1 - x_2|$  and

$$\min_{i=1,2} |x_i - \gamma(t)| \leq C_U(D) \operatorname{dist}(\gamma(t), \partial D) \quad \forall t \in [0, 1].$$

Any Lipschitz domain is also a uniform domain. In the following, for any Lipschitz domain  $D$ , we call the smallest constant  $C_U(D)$  that complies with Definition 5.6.1 the Jones parameter of  $D$ .

We paraphrase Jones' main extension result in the following theorem.

**Theorem 5.6.1** (Jones [73]). *Let  $T \subset \mathbb{R}^d$  be a bounded, uniform domain with  $\operatorname{diam} T = 1$ . Then there exists a bounded linear operator*

$$E : H^1(T) \rightarrow H^1(\mathbb{R}^d), \quad \forall v \in H^1(T) : (Ev)|_T = v$$

and a positive constant  $C_E(T)$  depending only on  $C_U(T)$  and  $d$  such that

$$\|Ev\|_{H^1(\mathbb{R}^d)} \leq C_E(T) \|v\|_{H^1(T)}.$$

*Remark.* The condition  $\operatorname{diam} T = 1$  in the above theorem stems from the fact that the  $L_2$ - and  $H^1$ -components of the full  $H^1$ -norm behave differently with respect to uniform scalings of  $T$ . If we drop this condition and set  $H = \operatorname{diam} T$ , it is easy to show by a dilation argument using Lemma 2.1.1 that the theorem remains true with the modified estimate

$$H^{-2} \|Ev\|_{L_2(\mathbb{R}^d)}^2 + |Ev|_{H^1(\mathbb{R}^d)}^2 \leq C_E(T)^2 (H^{-2} \|Ev\|_{L_2(T)}^2 + |Ev|_{H^1(T)}^2).$$

The second parameter that we use is the constant  $C_P(T)$  in the Poincaré inequality, Theorem 2.1.3. It has been shown in [95, Lemma 3.4] by combining a famous result by Maz'ya [86] and Federer and Fleming [39] with an auxiliary result by Kim that the constant  $C_P(T)$  can be tracked back to the constant in an isoperimetric inequality. For convex domains, Payne and Weinberger [93] have shown that  $C_P(T) \leq 1/\pi$ , with the proof later corrected in the 3D case by Bebendorf [6]. Estimates for shar-shaped domains can be found in [116, 97, 121].

Since each individual element  $T$  is Lipschitz, the Jones parameter  $C_U(T)$  and the Poincaré constant  $C_P(T)$  are both bounded. If we prescribe *a priori* uniform bounds for these geometric constants for all elements as well as their exterior domains, the results from Section 3.4 allow us to bound the BEM contraction constants uniformly.

**Lemma 5.6.2** ([95]). *Given a domain  $T$ , we fix a ball  $B$  enclosing  $T$  such that*

$$B \supset \bar{T}, \quad \text{dist}(\partial B, \partial T) \geq \frac{1}{2} \text{diam}(T), \quad (5.19)$$

*and let the Jones parameter  $C_U(B \setminus \bar{T})$  and the Poincaré constant  $C_P(B \setminus \bar{T})$  be bounded. Then, there exists a positive constant  $\tilde{c}_{0,T}$  depending solely on  $C_U(T)$ ,  $C_P(T)$ ,  $C_U(B \setminus \bar{T})$  and  $C_P(B \setminus \bar{T})$  such that*

$$c_{0,T} \geq \tilde{c}_{0,T} > 0.$$

### 5.6.2 An extension operator in the spirit of Stein

The Jones extension operator described in the previous section allows us to bound the BEM constants explicitly, however the geometric regularity assumptions we have to make, in particular the twisted cone condition in Definition 5.6.1, are non-standard and difficult to quantify. It turns out that, for the polytopal domains we consider in this work, an extension operator can be explicitly constructed in a fashion similar to that presented in the classical book by Stein [110] in such a way that its norm depends only on relatively standard geometric parameters, namely, the shape-regularity of an auxiliary triangulation. The results of this section were originally published in [60].

Assume that we have a polytopal, open, connected neighborhood  $T' \subset \mathbb{R}^d$  such that  $\bar{T} \subset T'$ , i.e.,  $T$  does not touch the boundary of  $T'$ . Further assume there is a shape-regular simplicial triangulation  $\Xi'$  of  $T'$  such that there is a subset  $\Xi \subset \Xi'$  which triangulates  $T$ . Let  $P$  denote the set of vertices of the triangulation  $\Xi$  which lie on the boundary of  $T$ . For every vertex  $p \in P$ , we define

$$\Xi_p := \{\tau \in \Xi' : p \in \bar{\tau}\}, \quad \bar{\omega}_p := \bigcup \{\bar{\tau} : \tau \in \Xi_p\}$$

such that  $\omega_p$  is the vertex patch of simplices which are neighbors of  $p$ , and the interior and exterior components

$$\omega_p^{\text{int}} := \omega_p \cap T, \quad \omega_p^{\text{ext}} := \omega_p \subset \bar{T}.$$

Without loss of generality, we assume that both  $\omega_p^{\text{int}}$  and  $\omega_p^{\text{ext}}$  contain at least one vertex of  $\Xi'$  which does not lie on  $\partial T$ , i.e., which is not in  $P$ . This can always be ensured by subdividing elements of  $\Xi'$  where necessary.

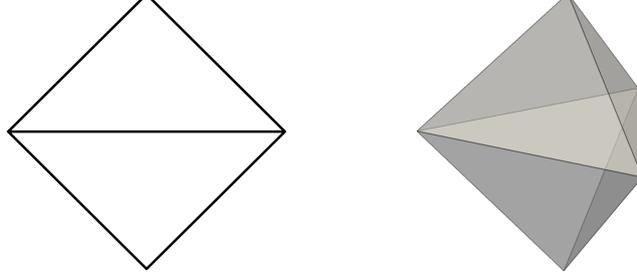
We define the reference patch

$$\hat{\omega} := \begin{cases} \text{conv}^\circ \{(-1, 0), (1, 0), (0, 1), (0, -1)\}, & d = 2, \\ \text{conv}^\circ \{(-1, 0, 0), (1, 1, 0), (1, -1, 0), (0, 0, 1), (0, 0, -1)\}, & d = 3, \end{cases}$$

where  $\text{conv}^\circ$  denotes the interior of the convex hull of the given set. The reference patch is thus a square in 2D and consists of two triangular pyramids joined at their bases in 3D. We define the subsets

$$\hat{\omega}^{\text{int}} := \{x \in \hat{\omega} : x_d < 0\}, \quad \hat{\omega}^{\text{ext}} := \{x \in \hat{\omega} : x_d > 0\},$$

where  $x_d$  refers to the  $d$ -th Cartesian coordinate of  $x$ . See Figure 5.1 for a sketch of the reference patches in 2D and 3D.



**Figure 5.1:** Reference patches  $\hat{\omega}$  in 2D and 3D, split into  $\hat{\omega}^{\text{ext}}$  and  $\hat{\omega}^{\text{int}}$ . In both cases, the two halves intersect at a  $(d - 1)$ -dimensional simplex.

For every vertex  $p \in P$ , let  $F_p : \hat{\omega} \rightarrow \omega_p$  denote a bijective, continuous mapping from the reference patch to the vertex patch  $\omega_p$  and set  $\hat{\Xi}_p := \{F_p^{-1}(\tau) : \tau \in \Xi_p\}$  of  $\hat{\omega}$ . Also let the following additional conditions hold:

- $F_p(0) = p$ ,
- if  $x_d = 0$ , then  $F_p(x) \in \partial T$ ,
- $F_p(\hat{\omega}^{\text{int}}) = \omega_p^{\text{int}}$  and  $F_p(\hat{\omega}^{\text{ext}}) = \omega_p^{\text{ext}}$ ,
- for all  $\tau \in \hat{\Xi}_p$ , the restriction  $F_p|_{\tau}$  is affine linear and satisfies

$$c_1 H^d \leq \det(F_p'|_{\tau}) \leq c_2 H^d,$$

$$\|F_p'|_{\tau}\|_{\ell_2} \leq c_3 H, \quad \|(F_p'|_{\tau})^{-1}\|_{\ell_2} \leq c_4 H^{-1},$$

where  $H = \text{diam} T$  and the constants  $c_1, c_2, c_3, c_4$  depend only on the shape-regularity parameters of  $\Xi'$ .

Clearly, under these conditions,  $\hat{\Xi}_p$  is a shape-regular triangulation of  $\hat{\omega}$ . By careful choice of  $F_p$ , these conditions can always be met for shape-regular  $\Xi'$ .

On the reference patch we define

$$\hat{E} : H^1(\hat{\omega}^{\text{int}}) \rightarrow H^1(\hat{\omega}^{\text{ext}}),$$

$$(\hat{E}v)(x_1, \dots, x_d) := v(x_1, \dots, x_{d-1}, -x_d)$$

by the reflection of  $v$  across the hyperplane  $x_d = 0$ . For each vertex  $p \in P$  we define

$$E_p : H^1(\omega_p^{\text{int}}) \rightarrow H^1(\omega_p^{\text{ext}}),$$

$$E_p v := (\hat{E}(v \circ F_p)) \circ F_p^{-1}.$$

Since  $F_p$  is continuous and piecewise affine,  $E_p v$  is indeed in  $H^1(\omega_p^{\text{ext}})$ . Furthermore, by assumption on  $F_p$  we have

$$(E_p v)|_{\omega_p \cap \partial T} = v|_{\omega_p \cap \partial T} \quad (5.20)$$

in the usual sense of  $H^1$ -traces, and thus the piecewise defined function

$$\begin{cases} v(x), & x \in \omega_p^{\text{int}}, \\ (E_p v)(x), & x \in \omega_p^{\text{ext}} \end{cases}$$

lies in  $H^1(\omega_p)$ .

Finally, let  $\varphi_p$  denote the standard nodal finite element basis function with respect to  $\Xi'$  which is 1 in  $p$  and 0 in all other vertices. Observe that its support is just  $\omega_p$ . Given a function  $v \in H^1(T)$ , we define, for  $x \in T'$ ,

$$(Ev)(x) := \begin{cases} v(x), & x \in T, \\ \sum_{p \in P} \varphi_p(x) \cdot (E_p v)(x) & x \in T' \setminus T. \end{cases}$$

Clearly,  $E$  is a linear operator. Since  $\{\varphi_p\}$  is a partition of unity on  $\partial T$  and due to (5.20), the interior and exterior parts of the definition have identical traces on  $\partial T$  and thus the operator  $E$  maps into  $H^1(T')$ . Furthermore, since by construction every  $\varphi_p$  vanishes on  $\partial T'$ , we see that  $Ev \in H_0^1(T')$ . By a simple extension by 0 into the exterior of  $T'$ , we thus have

$$E : H^1(T) \rightarrow H^1(\mathbb{R}^d),$$

and since  $(Ev)|_T \equiv v$ ,  $E$  is an extension operator. It remains to bound the norm of  $E$  in terms of the mesh parameters.

**Theorem 5.6.3.** *Assume that  $\Xi$  consists of at most  $N_\Xi \in \mathbb{N}$  simplicial elements. Then there exists a constant  $C_E$  depending only on  $N_\Xi$  and on the shape-regularity parameters of  $\Xi'$  such that, for all  $v \in H^1(T)$ ,*

$$|Ev|_{H^1(\mathbb{R}^d)}^2 + H^{-2} \|Ev\|_{L_2(\mathbb{R}^d)}^2 \leq C_E \left( |v|_{H^1(T)}^2 + H^{-2} \|v\|_{L_2(T)}^2 \right).$$

*Proof.* Let  $v \in H^1(T)$ . Using standard finite element techniques (cf. [10, 22]), one easily shows

$$|E_p v|_{H^1(\omega_p^{\text{ext}})} \leq C |v|_{H^1(\omega_p^{\text{int}})}, \quad \|E_p v\|_{L_2(\omega_p^{\text{ext}})} \leq C \|v\|_{L_2(\omega_p^{\text{int}})}.$$

The constant  $C$  depends only on the shape-regularity parameters of  $\Xi'$  because there are only a small number of different triangulations  $\hat{\Xi}_p$ .

Since  $\|\varphi_p\|_\infty = 1$ , it follows that

$$\|\varphi_p \cdot E_p v\|_{L_2(\omega_p^{\text{ext}})} \leq C \|v\|_{L_2(\omega_p^{\text{int}})}.$$

Since  $\|\nabla\varphi_p\|_\infty \leq CH^{-1}$ , we can conclude by the product rule that

$$\begin{aligned} |\varphi_p \cdot E_p v|_{H^1(\omega_p^{\text{ext}})}^2 &\leq C \left( |E_p v|_{H^1(\omega_p^{\text{ext}})}^2 + H^{-2} \|E_p v\|_{L_2(\omega_p^{\text{ext}})}^2 \right) \\ &\leq C \left( |v|_{H^1(\omega_p^{\text{int}})}^2 + H^{-2} \|v\|_{L_2(\omega_p^{\text{int}})}^2 \right) \\ &\leq C \left( |v|_{H^1(T)}^2 + H^{-2} \|v\|_{L_2(T)}^2 \right). \end{aligned}$$

Since  $\#P$  can be bounded in terms of  $N_\Xi$ , the statement follows by summing the above estimates over all  $p \in P$ .  $\square$

Reversing the roles of the interior and exterior domains in the above construction, we can completely analogously define an exterior extension operator

$$E^{\text{ext}} : H^1(T' \setminus T) \rightarrow H^1(T'),$$

and we obtain a similar estimate for its norm using the same proof technique.

**Theorem 5.6.4.** *Under the assumptions of Theorem 5.6.3, we have for all  $v \in H^1(T' \setminus T)$*

$$|E^{\text{ext}} v|_{H^1(T')}^2 \leq C_E |v|_{H^1(T' \setminus T)}^2$$

with a constant  $C_E$  depending only on  $m$ , the shape-regularity parameters of  $\Xi'$ , and the Poincaré constant  $C_P(T' \setminus T)$ .

*Proof.* As in the proof of Theorem 5.6.3, we obtain first

$$|E^{\text{ext}} v|_{H^1(T')}^2 \leq C \left( |v|_{H^1(T' \setminus T)}^2 + H^{-2} \|v\|_{L_2(T' \setminus T)}^2 \right).$$

Let  $\bar{v} := |T' \setminus T|^{-1} \int_{T' \setminus T} v$  denote the mean of  $v$ , and observe that  $E^{\text{ext}}$ , by construction, preserves constants. We thus have

$$\begin{aligned} |E^{\text{ext}} v|_{H^1(T')}^2 &= |E^{\text{ext}}(v - \bar{v}) + \bar{v}|_{H^1(T')}^2 = |E^{\text{ext}}(v - \bar{v})|_{H^1(T')}^2 \\ &\leq C \left( |v|_{H^1(T' \setminus T)}^2 + H^{-2} \|v - \bar{v}\|_{L_2(T' \setminus T)}^2 \right). \end{aligned}$$

The statement then follows with the Poincaré inequality, Theorem 2.1.3.  $\square$

It is clear that, since  $T' \setminus T$  is a bounded subset of  $\mathbb{R}^d \setminus T$ , the exterior extension operator extends trivially to the spaces

$$E^{\text{ext}} : H_{\text{loc}*}^1(\mathbb{R}^d \setminus T) \rightarrow H_{\text{loc}*}^1(\mathbb{R}^d),$$

and the estimate in Theorem 5.6.4 then implies

$$|E^{\text{ext}} v|_{H^1(T)}^2 \leq C_E |v|_{H^1(\mathbb{R}^d \setminus T)}^2.$$

Thus, the exterior extension operator constructed here satisfies the assumptions made in Lemma 3.4.1 and in the further theorems of that section.

# Chapter 6

## Error analysis

The aim of this chapter is to derive rigorous error estimates for the numerical scheme described by (4.11). Recall that the discretization of the variational formulation (4.5) proceeded in two steps: we chose a finite-dimensional trial space  $\mathcal{W}_h \subset \mathcal{W}$ , and, to make the scheme computable, we chose an approximation  $\tilde{S}_T$  of the Dirichlet-to-Neumann map  $S_T$ . While the first step leads to a standard Galerkin method which is easily analyzed using the Céa lemma, the second step introduces a consistency error which demands analysis by a Strang lemma. Using this tool and applying the estimates derived in Chapter 5, we obtain an error estimate for the BEM-based FEM in the  $H^1(\Omega)$ -norm which is quasi-optimal with respect to the approximation properties of the skeletal space, and also exhibits the same convergence rate as a standard piecewise linear FEM on a simplicial mesh.

We are also interested in  $L_2$ -error estimates. However, they are difficult to obtain from the primal formulation since it involves an approximated bilinear form. A “variational crime” of this type means that no Galerkin orthogonality holds for our variational problem, and thus the standard Aubin-Nitsche duality trick cannot be applied. We have seen, however, that the equivalent mixed discretization derived in Section 4.3.3 does permit a Galerkin orthogonality, and thus we will use it in order to derive an  $L_2$ -error estimate which, again, is quasi-optimal.

In this chapter, for simplicity, we consider only the Laplace equation, i.e., the choice  $L = -\Delta$  for the partial differential operator.

### 6.1 Error of the inexact Galerkin scheme

On the skeletal space  $\mathcal{W} = H^{1/2}(\Gamma_S)$ , we define the skeletal energy norm by

$$\|v\|_S := \sqrt{a(v, v)} = \left( \sum_{T \in \mathcal{T}} \langle S_T v_T, v_T \rangle \right)^{1/2}. \quad (6.1)$$

Using previous results and definitions, in particular (5.1), we have the equivalent representations

$$\|v\|_S = \left( \sum_{T \in \mathcal{T}} |v_T|_{H^{1/2}(\partial T)}^2 \right)^{1/2} = \left( \sum_{T \in \mathcal{T}} |\mathcal{H}_T v_T|_{H^1(T)}^2 \right)^{1/2} = |\mathcal{H}_S v|_{H^1(\Omega)}. \quad (6.2)$$

On the space  $\mathcal{W}_0$ , whose members satisfy homogeneous boundary conditions, this is indeed a full norm, as one easily proves using the Friedrichs inequality (Theorem 2.1.2).

We use the Strang lemma, Lemma 2.2.4, to prove a first Céa-type error estimate in this skeletal energy norm.

**Lemma 6.1.1.** *Let  $u \in \mathcal{W}$  be the solution of (4.7), and  $u_h \in \mathcal{W}_h$  the solution of (4.11). Denote, for all  $T \in \mathcal{T}$ , by  $t_T(u) = S_T u \in H^{-1/2}(\partial T)$  the elementwise Neumann data corresponding to the exact solution. Then we have the error estimate*

$$\begin{aligned} |\mathcal{H}_S(u - u_h)|_{H^1(\Omega)} &= \|u - u_h\|_S \\ &\leq C \left\{ \inf_{\substack{v_h \in \mathcal{W}_h \\ v_h|_{\partial\Omega} = g}} \|u - v_h\|_S + \left( \sum_{T \in \mathcal{T}} \inf_{z_{h,T} \in \mathcal{Z}_{h,T}} \|t_T(u) - z_{h,T}\|_{V_T}^2 \right)^{1/2} \right\}, \end{aligned} \quad (6.3)$$

where

$$\begin{aligned} C &= \left( 1 + \frac{1}{\underline{c}_S} \right) \max \left\{ 1, \frac{\bar{c}_K}{\sqrt{1 - \bar{c}_K}} \right\}, \\ \bar{c}_K &= \max_{T \in \mathcal{T}} c_{K,T} < 1, \\ \underline{c}_S &= \min_{T \in \mathcal{T}} \frac{c_{0,T}}{c_{K,T}} > 0, \end{aligned}$$

and  $c_{0,T} > 0$  and  $c_{K,T} < 1$  are the BEM constants as introduced in Chapter 3 on the boundary of the element  $T$ .

*Proof.* In the notation of Lemma 2.2.4 (Strang), we use the Hilbert spaces  $\mathcal{W}_{h,0} \subset \mathcal{W}_0$  with the norm  $\|\cdot\|_S$  and consider the homogenized equations (4.9) and (4.12) since clearly

$$\|u - u_h\|_S = \|u_0 - u_{h,0}\|_S.$$

The bilinear form  $a(\cdot, \cdot)$  as defined in (4.8) is nothing but the inner product which induces the norm  $\|\cdot\|_S$ , and thus it has the bounds  $c_1 = c_2 = 1$ . Due to relation (3.10), for every  $T \in \mathcal{T}$  the local approximations  $\tilde{S}_T$  satisfy the bounds

$$\frac{c_{0,T}}{c_{K,T}} S_T \leq \tilde{S}_T \leq S_T,$$

and thus the approximate bilinear form  $\tilde{a}(\cdot, \cdot)$  as defined in (4.10) has bounds  $\tilde{c}_1 = \underline{c}_S$  and  $\tilde{c}_2 = 1$ . (The upper bounds follow from the spectral estimates via the Cauchy-Schwarz

inequality,  $\langle S_T v_T, t_T \rangle^2 \leq \langle S_T v_T, v_T \rangle \langle S_T t_T, t_T \rangle$ , due to symmetry.) Lemma 2.2.4 then implies the error estimate

$$\|u - u_h\|_S \leq C_1 \left( \inf_{v_h \in \mathcal{W}_{h,0}} \|u_0 - v_h\|_S + \sup_{v_h \in \mathcal{W}_{h,0}} \frac{|\tilde{a}(u_0, v_h) + \tilde{a}(g, v_h)|}{\|v_h\|_S} \right), \quad (6.4)$$

where  $C_1 = 1 + \frac{1}{c_S}$ .

We now estimate the consistency error, i.e., the numerator of the rightmost term in (6.4). Since  $u = u_0 + g$ , this error is given by  $|\tilde{a}(u, v_h)|$ . Recall that, by definition,  $a(u, v) = 0$  for all  $v \in \mathcal{W}_0$ . Hence,  $|\tilde{a}(u, v_h)| = |a(u, v_h) - \tilde{a}(u, v_h)|$ , and by the definitions of  $S_T$  and  $\tilde{S}_T$  we see that

$$\begin{aligned} a(u, v_h) - \tilde{a}(u, v_h) &= \sum_{T \in \mathcal{T}} \left( \langle S_T u_T, v_{h,T} \rangle - \langle \tilde{S}_T u_T, v_{h,T} \rangle \right) \\ &= \sum_{T \in \mathcal{T}} \left\langle \left( \frac{1}{2}I + K'_T \right) (t_T(u) - t_{h,T}(u)), v_{h,T} \right\rangle \\ &= \sum_{T \in \mathcal{T}} \left\langle \left( \frac{1}{2}I + K_T \right) v_{h,T}, t_T(u) - t_{h,T}(u) \right\rangle, \end{aligned}$$

where  $t_{h,T}(u)$  is determined by relation (3.9). In order to bound the local consistency error on each element boundary  $\partial T$ , we use the fact that  $\|\cdot\|_{V_T}$  is the associated dual norm to  $\|\cdot\|_{V_T^{-1}}$ , i.e.,

$$\sup_{v \in H^{1/2}(\partial T)} \frac{\langle w, v \rangle_{\partial T}}{\|v\|_{V_T^{-1}}} = \|w\|_{V_T},$$

which is easily obtained by standard duality arguments. Hence,

$$\begin{aligned} & \left| \left\langle \left( \frac{1}{2}I + K_T \right) v_{h,T}, t_T(u) - t_{h,T}(u) \right\rangle_{\partial T} \right| \\ & \leq \left\| \left( \frac{1}{2}I + K_T \right) v_{h,T} \right\|_{V_T^{-1}} \|t_T(u) - t_{h,T}(u)\|_{V_T} \\ & \leq \frac{c_{K,T}}{\sqrt{1 - c_{K,T}}} |v_{h,T}|_{S_T} \|t_T(u) - t_{h,T}(u)\|_{V_T}, \end{aligned} \quad (6.5)$$

where in the last line we have used inequality (3.8).

It remains to estimate the rightmost term in (6.5). By the defining relations  $V_T t_T(u) = (\frac{1}{2}I + K_T)u_T$  for  $t_T(u)$  and (3.9) for  $t_{h,T}(u)$ , we have the Galerkin orthogonality

$$\langle V_T(t_T(u) - t_{h,T}(u)), z_{h,T} \rangle = 0 \quad \forall z_{h,T} \in \mathcal{Z}_{h,T}.$$

By a simple application of Lemma 2.2.3 (Céa), we therefore get

$$\|t_T(u) - t_{h,T}(u)\|_{V_T} = \inf_{z_{h,T} \in \mathcal{Z}_{h,T}} \|t_T(u) - z_{h,T}\|_{V_T}.$$

Combining the above results, we obtain

$$\begin{aligned} |a(u, v_h) - \tilde{a}(u, v_h)| &\leq \sum_{T \in \mathcal{T}} \frac{c_{K,T}}{\sqrt{1 - c_{K,T}}} |v_{h,T}|_{S_T} \|t_T(u) - t_{h,T}(u)\|_{V_T}, \\ &\leq \max \left\{ 1, \frac{\bar{c}_K}{\sqrt{1 - \bar{c}_K}} \right\} \|v_h\|_S \left( \sum_{T \in \mathcal{T}} \inf_{z_{h,T} \in \mathcal{Z}_{h,T}} \|t_T(u) - z_{h,T}\|_{V_T}^2 \right)^{1/2} \end{aligned}$$

by the Cauchy-Schwarz inequality in  $\mathbb{R}^{\#\mathcal{T}}$ . Inserting this estimate in (6.4), we obtain the desired statement.  $\square$

The error estimate (6.3) contains the constants  $\bar{c}_K$  and  $\underline{c}_S$ . To clarify their dependence on the mesh, or to be more precise, on the shapes of the elements, we will make use of the results from Section 3.4. Furthermore, estimating the error in terms of the Dirichlet and Neumann errors on the skeleton is not desirable since these terms are inherently mesh-dependent. To bound these, the approximation properties derived in Chapter 5 will be used. The remainder of our error analysis is thus concerned with estimating the expressions on the right-hand side of (6.3) only in terms of the exact solution and certain regularity parameters of the mesh.

In the sequel we restrict ourselves to the three-dimensional case.

## 6.2 Geometric assumptions on the mesh

The main regularity assumption for the polyhedral mesh  $\mathcal{T}$  which allows us to obtain error estimates for the BEM-based FEM is the following. It is chosen such that Assumption 5.0.1 holds for every element  $T \in \mathcal{T}$ , and thus the statements from Chapter 5 hold for every element.

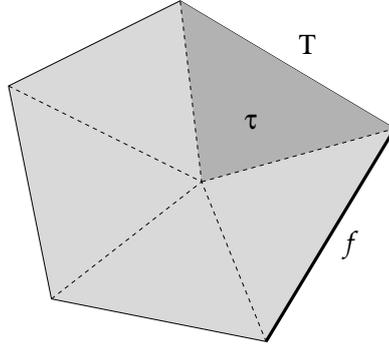
**Assumption 6.2.1.** *There exists a conforming simplicial triangulation  $\Xi$  of  $\Omega$  which is shape-regular according to Definition 2.5.1 with constants  $\underline{c}_1$ ,  $\bar{c}_1$ ,  $\underline{c}_2$ , and  $\bar{c}_2 > 0$  such that every element  $T \in \mathcal{T}$  is triangulated by a subset  $\Xi_T \subset \Xi$ , and every element consists of at most  $N_\Xi$  simplices,  $\#\Xi_T \leq N_\Xi \forall T \in \mathcal{T}$ , with a small integer  $N_\Xi$ .*

By this assumption, every element  $T \in \mathcal{T}$  has a conforming shape-regular triangulation  $\Xi_T$  consisting of mutually disjoint tetrahedra  $\tau$ ,

$$\bar{T} = \bigcup_{\tau \in \Xi_T} \bar{\tau}.$$

By  $\mathcal{F}_T$ , we denote the collection of all triangular faces  $f$  of tetrahedra  $\tau \in \Xi_T$  which lie on the element boundary  $\partial T$ , such that  $\partial T = \bigcup_{f \in \mathcal{F}_T} \bar{f}$ . This setting is illustrated in Figure 6.1 for the two-dimensional case.

By assumption, the triangulations are *matching*, that is, faces from neighboring elements are either identical or do not intersect at all.



**Figure 6.1:** Sketch of a pentagonal element  $T$  with auxiliary triangulation  $\Xi_T$ , one of its constituting simplices  $\tau \in \Xi_T$ , and a boundary facet  $f \in \mathcal{F}_T$ .

We emphasize that the local triangulations  $\Xi$  and  $\Xi_T$  are a purely analytical device and not required for the numerical realization.

In the standard finite element analysis, we usually obtain uniform constants by transforming domain and surface integrals to reference elements. In this way, the constants appearing in the estimates depend only on mesh regularity parameters as well as on some fixed constants stemming from inequalities on the reference elements. For general polyhedral meshes, no such technique is available. In particular, we cannot estimate the constants  $c_{0,T}$  by transformation to reference elements. In order to nonetheless get uniform bounds, we make use of shape-explicit bounds on the constants  $c_{0,T}$  described in Section 3.4, in conjunction with explicit extension results from Section 5.6. To be able to apply these results, we need to make the following slightly stronger assumption.

**Assumption 6.2.2.** *We assume that Assumption 6.2.1 holds. Furthermore, we assume that for an open, bounded neighborhood  $\Omega' \supset \bar{\Omega}$ , there exists a triangulation  $\Xi' \supset \Xi$  with the same regularity properties.*

We first cite a result that allows us to bound Poincaré constants of polytopes constructed from a shape-regular simplicial mesh.

**Lemma 6.2.3.** *Let  $\Xi$  be a shape-regular triangulation, and let  $T$  be the union of at most  $m$  simplices from  $\Xi$ . Then the Poincaré constant  $C_P(T)$  of  $T$  can be bounded from above by an expression which depends only on  $m$  and the regularity parameters of  $\Xi$ .*

*Proof.* Is a direct consequence of [97, Lemma 4.1]. Compare also [121].

Using the results outlined in previous sections, the following inequalities hold uniformly for all elements under the above regularity assumption.

**Theorem 6.2.4.** *Let Assumption 6.2.2 hold. For every element  $T \in \mathcal{T}$ , let  $V_T$ ,  $K_T$ ,  $K_T^l$ ,  $D_T$  denote the boundary integral operators on  $\partial T$  as introduced in Chapter 3. Then we have*

$$\begin{aligned} \|z\|_{V_T} &\leq C_V^* \|z\|_{H^{-1/2}(\partial T)} & \forall z \in H^{-1/2}(\partial T), \\ c_{0,T} \langle V_T^{-1}v, v \rangle &\leq \langle D_T v, v \rangle & \forall v \in H_*^{1/2}(\partial T), \\ (1 - c_{K,T}) \|v\|_{V_T^{-1}} &\leq \|(\frac{1}{2}I \pm K_T)v\|_{V_T^{-1}} \leq c_{K,T} \|v\|_{V_T^{-1}} & \forall v \in H_*^{1/2}(\partial T), \\ \frac{c_{0,T}}{c_{K,T}} \langle S_T v, v \rangle_\Gamma &\leq \langle \tilde{S}_T v, v \rangle_\Gamma \leq \langle S_T v, v \rangle_\Gamma & \forall v \in H^{1/2}(\partial T). \end{aligned}$$

with constants  $C_V^*$ ,  $c_{0,T}$ ,  $c_{K,T}$  that are uniformly bounded in terms of the regularity parameters of  $\Xi$ .

*Proof.* Assumption 6.2.2 implies that for every element  $T \in \mathcal{T}$ , there exists a neighborhood  $T'$  with a shape-regular triangulation  $\Xi'_T \subset \Xi'$  as assumed in the construction of the extension operator in Section 5.6.2. Therefore, Theorem 5.6.3 and Theorem 5.6.4 guarantee for every  $T$  the existence of extension operators

$$E : H^1(T) \rightarrow H^1(\mathbb{R}^d), \quad E^{\text{ext}} : H^1(T' \setminus T) \rightarrow H^1(T')$$

with norms which are bounded in terms of the regularity parameters of  $\Xi$  as well as the Poincaré constant  $C_P(T' \setminus T)$ . The latter can be bounded in terms of the mesh regularity parameters by Lemma 6.2.3 since every element, and thus also its neighborhood  $T'$ , was assumed to consist of only a few simplices of  $\Xi'$ .

The inequalities which form the statement of the theorem have all been introduced in Chapter 3, and it has been shown in Section 3.4 that all involved constants can be tracked back to the norms of the extension operators  $E$  and  $E^{\text{ext}}$  as well as the Poincaré constants.  $\square$

### 6.3 Approximation error in the Dirichlet data

Drawing upon the standard finite element interpolation theory that was outlined in Section 2.5, we can construct a skeletal interpolation operator with similar properties as the Scott-Zhang quasi-interpolator from Theorem 2.5.2.

**Theorem 6.3.1.** *Let the mesh  $\mathcal{T}$  satisfy Assumption 6.2.1. Then there exists a skeletal quasi-interpolation operator  $\Pi_S : H^1(\Omega) \rightarrow \mathcal{W}_h$  which preserves piecewise linear boundary data,*

$$(\Pi_S u_\Omega)|_{\partial\Omega} = u_\Omega|_{\partial\Omega} \quad \text{for all } u_\Omega \in H^1(\Omega) \text{ with } u_\Omega|_{\partial\Omega} \in \mathcal{V}_h(\partial\Omega),$$

and satisfies the skeletal estimate

$$\|\gamma_S u_\Omega - \Pi_S u_\Omega\|_S \leq c_\Pi h^{k-1} |u_\Omega|_{H^k(\Omega)} \quad \forall u_\Omega \in H^k(\Omega),$$

where  $k \in [1, 2]$ ,  $h$  is the mesh size of  $\mathcal{T}$ , and the constant  $c_\Pi > 0$  depends only on the shape regularity parameters of  $\mathcal{T}$ .

*Proof.* Let  $\mathcal{V}_h \subset H^1(\Omega)$  denote the standard piecewise linear finite element space over the triangulation  $\Xi$  of  $\Omega$  as introduced in Section 2.5. We set

$$\Pi_S := \gamma_S \circ \Pi,$$

where  $\Pi$  is the Scott-Zhang interpolation operator for  $\Xi$  from Theorem 2.5.2. The first statement, preservation of piecewise linear boundary data, follows directly from the corresponding statement in Theorem 2.5.2.

For the second statement, let  $u_\Omega \in H^k(\Omega)$  and denote by  $u := \gamma_S u_\Omega \in \mathcal{W}$  its skeletal trace. Furthermore, let  $\phi_h := \Pi u_\Omega \in \mathcal{V}_h$  and denote by  $\Phi_h := \Pi_S u_\Omega = \gamma_S \phi_h \in \mathcal{W}_h$  the skeletal trace of the interpolant. We have

$$\|\gamma_S u_\Omega - \Pi_S u_\Omega\|_S^2 = \sum_{T \in \mathcal{T}} |\mathcal{H}_T(u - \Phi_h)|_{H^1(T)}^2.$$

By construction, on the boundary of every element  $T$ , we have  $(u_\Omega - \phi_h)|_{\partial T} = u - \Phi_h$ , and hence, by the energy-minimizing property of the harmonic extension (Theorem 2.4.2),

$$|\mathcal{H}_T(u - \Phi_h)|_{H^1(T)}^2 \leq |u_\Omega - \phi_h|_{H^1(T)}^2 \quad \forall T \in \mathcal{T}.$$

By summing up over all  $T \in \mathcal{T}$ , we get

$$\|\gamma_S u_\Omega - \Pi_S u_\Omega\|_S \leq |u_\Omega - \Pi u_\Omega|_{H^1(\Omega)},$$

and the statement follows with Theorem 2.5.2.  $\square$

By this interpolation result, we get a skeletal approximation result in analogy to Corollary 2.5.3.

**Corollary 6.3.2.** *Let the mesh  $\mathcal{T}$  satisfy Assumption 6.2.1. Let  $u_\Omega \in H^k(\Omega)$  with  $k \in [1, 2]$  and  $u = \gamma_S u_\Omega \in H^{1/2}(\Gamma_S)$  its skeletal trace. Assume furthermore that  $u|_{\partial\Omega}$  is piecewise linear with respect to the mesh facets which lie on  $\partial\Omega$ . Then we have*

$$\inf_{\substack{v_h \in \mathcal{W}_h \\ v_h|_{\partial\Omega} = u|_{\partial\Omega}}} \|u - v_h\|_S \leq C h^{k-1} |u_\Omega|_{H^k(\Omega)}, \quad (6.6)$$

where the constant  $C$  depends only on the regularity parameters from Assumption 6.2.1.

*Proof.* With the choice  $w_h := \Pi_S u_\Omega$ , we have

$$\inf_{\substack{v_h \in \mathcal{W}_h \\ v_h|_{\partial\Omega} = u|_{\partial\Omega}}} \|u - v_h\|_S \leq \|u - w_h\|_S,$$

and the statement follows directly from Theorem 6.3.1.  $\square$

## 6.4 Approximation error in the Neumann data

The error estimate derived in Section 6.1 also involves a best-approximation error of the elementwise Neumann traces of the exact solution, measured in the single layer potential norm,  $\|\cdot\|_{V_T}$ . The following theorem provides a bound on this approximation error.

**Theorem 6.4.1.** *Let the mesh  $\mathcal{T}$  satisfy Assumption 6.2.2 and fix an element  $T \in \mathcal{T}$ . Let  $\phi \in H^2(T)$  and let  $w = \gamma_T^1 \phi \in H_{pw}^{1/2}(\partial T)$  be its normal derivative. Then,*

$$\inf_{z_{h,T} \in \mathcal{Z}_{h,T}} \|w - z_{h,T}\|_{V_T} \leq C h_T |\phi|_{H^2(T)}$$

where  $h_T = \text{diam } T$  and the constant  $C$  depends solely on the regularity parameters from Assumption 6.2.1.

*Proof.* Let  $Q_{h,T} : L_2(\partial T) \rightarrow \mathcal{Z}_{h,T}$  be the  $L_2$ -projector to the piecewise constant functions introduced in Section 5.5. Using Theorem 6.2.4, Theorem 5.5.2, and Theorem 5.3.3, we estimate

$$\begin{aligned} \inf_{z_{h,T} \in \mathcal{Z}_{h,T}} \|w - z_{h,T}\|_{V_T} &\leq \|w - Q_{h,T} w\|_{V_T} \\ &\stackrel{\text{Thm. 6.2.4}}{\leq} C_V^* \|w - Q_{h,T} w\|_{H^{-1/2}(\partial T)} \\ &\stackrel{\text{Thm. 5.5.2}}{\leq} C h_T |w|_{H_{pw}^{1/2}(\partial T)} \\ &= C h_T |\gamma_T^1 u_\Omega|_{H_{pw}^{1/2}(\partial T)} \\ &\stackrel{\text{Thm. 5.3.3}}{\leq} C h_T |u_\Omega|_{H^2(T)}. \quad \square \end{aligned}$$

## 6.5 $H^1$ -error estimate

With the above approximation error bounds, we have the tools at hand to derive an  $H^1$ -error estimate for the BEM-based FEM.

**Theorem 6.5.1.** *Let the mesh  $\mathcal{T}$  satisfy Assumption 6.2.2. Assume further that the given Dirichlet data  $g$  is piecewise linear. If  $u_\Omega \in H^2(\Omega)$  is the exact solution of the variational formulation (2.4), and  $u_h \in \mathcal{W}_h$  is the solution of the discrete skeletal formulation (4.11), we have the error estimate*

$$|u_\Omega - \mathcal{H}_S u_h|_{H^1(\Omega)} \leq C h |u_\Omega|_{H^2(\Omega)},$$

where the constant  $C$  depends solely on the regularity parameters from Assumption 6.2.1.

*Proof.* From Proposition 4.1.2, we know that  $u_\Omega = \mathcal{H}_S u$ , and thus  $u_\Omega - \mathcal{H}_S u_h = \mathcal{H}_S(u - u_h)$ . From Lemma 6.1.1, we have

$$|\mathcal{H}_S(u - u_h)|_{H^1(\Omega)} \leq C \left\{ \inf_{\substack{v_h \in \mathcal{W}_h \\ v_h|_{\partial\Omega} = g}} \|u - v_h\|_S + \left( \sum_{T \in \mathcal{T}} \inf_{z_{h,T} \in Z_{h,T}} \|t_T(u) - z_{h,T}\|_{V_T}^2 \right)^{1/2} \right\}$$

with

$$C = \left( 1 + \frac{1}{\underline{c}_S} \right) \max \left\{ 1, \frac{\bar{c}_K}{\sqrt{1 - \bar{c}_K}} \right\}.$$

Due to Theorem 6.2.4, the constants  $c_{0,T}$  and  $c_{K,T}$  can be bounded in terms of the regularity parameters of the mesh, and thus the same holds true for  $C$  since it depends only on these constants. Corollary 6.3.2 yields the Dirichlet approximation property

$$\inf_{\substack{v_h \in \mathcal{W}_h \\ v_h|_{\partial\Omega} = g}} \|u - v_h\|_S \leq C h |u_\Omega|_{H^2(\Omega)}.$$

By definition,  $t_T(u) = S_T u_T$ , and since  $u_T$  is the Dirichlet trace of the  $\mathcal{L}$ -harmonic function  $u_\Omega$ , we have  $t_T(u) = \gamma_T^1 u_\Omega \in H_{\text{pw}}^{1/2}(\partial T)$ . Thus, the remaining terms can be treated using the Neumann approximation property from Theorem 6.4.1,

$$\inf_{z_{h,T} \in Z_{h,T}} \|t_T(u) - z_{h,T}\|_{V_T} \leq C h_T |u_\Omega|_{H^2(T)}. \quad \square$$

For the case that we cannot assume full  $H^2(\Omega)$ -regularity of the exact solution, we get at the very least a stability result for the error. Estimates for solutions with reduced regularity could be obtained by means of interpolation theory (cf. [10, 8, 119]).

**Theorem 6.5.2.** *Let the mesh  $\mathcal{T}$  satisfy Assumption 6.2.2. Assume further that the given Dirichlet data  $g$  is piecewise linear. If  $u_\Omega \in H^1(\Omega)$  is the exact solution of the variational formulation (2.4), and  $u_h \in \mathcal{W}_h$  is the solution of the discrete skeletal formulation (4.11), we have the stability estimate*

$$|u_\Omega - \mathcal{H}_S u_h|_{H^1(\Omega)} \leq C |u_\Omega|_{H^1(\Omega)},$$

where the constant  $C$  depends solely on the regularity parameters from Assumption 6.2.1.

*Proof.* We start again from the estimate

$$|\mathcal{H}_S(u - u_h)|_{H^1(\Omega)} \leq C \left\{ \inf_{\substack{v_h \in \mathcal{W}_h \\ v_h|_{\partial\Omega} = g}} \|u - v_h\|_S + \left( \sum_{T \in \mathcal{T}} \inf_{z_{h,T} \in Z_{h,T}} \|t_T(u) - z_{h,T}\|_{V_T}^2 \right)^{1/2} \right\}$$

as in the proof of Theorem 6.5.1. We bound the infima in the second summand with the special choice  $z_{h,T} = 0$  and observe that then

$$\|t_T(u)\|_{V_T} = \|(\frac{1}{2}I + K_T)u_T\|_{V_T^{-1}} \leq C |u_T|_{S_T} = C |u_\Omega|_T|_{H^1(\Omega)}$$

with a uniform constant  $C$  due to the inequality (3.8). Thus we get

$$\|u - u_h\|_S \leq C \left( \inf_{\substack{v_h \in \mathcal{W}_h \\ v_h|_{\partial\Omega} = g}} \|u - v_h\|_S + C|u_\Omega|_{H^1(\Omega)} \right),$$

and the remaining infimum can be estimated by Corollary 6.3.2 with  $k = 1$ .  $\square$

## 6.6 Analysis of the mixed formulation

In this section, we give an analysis of the discretized mixed problem (4.13). As in the analysis of the primal formulation, we first derive error estimates in skeletal function spaces. While inherently mesh-dependent, they are an important intermediate result in the derivation of mesh-independent estimates. We proceed by rederiving an error estimate in the  $H^1$ -norm using the mixed variational framework. The true advantage of the mixed formulation is that it provides us with a Galerkin orthogonality. This allows us to apply an Aubin-Nitsche duality argument in order to prove an error estimate in the  $L_2$ -norm, which is difficult in the primal formulation.

For the convergence and approximation results that follow, we equip the mixed skeletal space  $\mathcal{X} = \mathcal{W} \times \mathcal{Z}$  with the norm

$$\|(v, z)\|_{\mathcal{X}}^2 := \|v\|_S^2 + \|z\|_V^2 := \sum_{T \in \mathcal{T}} \langle S_T v_T, v_T \rangle + \sum_{T \in \mathcal{T}} \langle z_T, V_T z_T \rangle.$$

The approximation properties proved previously transfer directly to the mixed space with this norm.

**Theorem 6.6.1.** *Let the mesh  $\mathcal{T}$  satisfy Assumption 6.2.2. If  $w_\Omega \in H^2(\Omega)$  with piecewise linear boundary conditions  $g$ , and if  $(\phi, \eta) \in \mathcal{X}$  denotes its skeletal Dirichlet and Neumann data, respectively, that is,*

$$\phi = \gamma_S w_\Omega, \quad \eta_T = \gamma_T^1 w_\Omega \text{ for } T \in \mathcal{T},$$

then we have

$$\inf_{\substack{(\phi_h, \eta_h) \in \mathcal{X}_h \\ \phi_h|_{\partial\Omega} = g}} \|(\phi - \phi_h, \eta - \eta_h)\|_{\mathcal{X}} \leq C h |w_\Omega|_{H^2(\Omega)} \quad (6.7)$$

with a uniform constant  $C$ .

*Proof.* Follows directly from Corollary 6.3.2 and Theorem 6.4.1.  $\square$

### 6.6.1 Convergence on the skeleton

**Theorem 6.6.2.** *Let the mesh  $\mathcal{T}$  satisfy Assumption 6.2.2 and  $g$  be piecewise linear. Then the discrete solution  $(u_h, t_h) \in \mathcal{X}_h$  of (4.13) is a quasi-optimal approximation to the solution  $(u, t) \in \mathcal{X}$  of (4.6). That is,*

$$\|(u - u_h, t - t_h)\|_{\mathcal{X}} \leq C \inf_{\substack{(v_h, z_h) \in \mathcal{X}_h \\ v_h|_{\partial\Omega} = g}} \|(u - v_h, t - z_h)\|_{\mathcal{X}} \quad (6.8)$$

with a uniform constant  $C$ .

*Proof.* The result is proved using Céa's Lemma (Lemma 2.2.3). Hence, only uniform coercivity and boundedness of the bilinear form  $\mathcal{A}$  need to be shown.

We take note of the spectral equivalence

$$c_{D,T}^* \langle S_T v_T, v_T \rangle \leq \langle D_T v_T, v_T \rangle \leq \langle S_T v_T, v_T \rangle \quad \forall v \in H^{1/2}(\partial T) \quad (6.9)$$

where the lower bound follows from Lemma 3.4.5 and the upper bound is a direct consequence of (3.6) and the coercivity of  $V$ . The assumptions guarantee that the constant  $c_{D,T}^* = \frac{1}{2}(C_{E_{\text{ext}}})^{-2} \in (0, 1)$  is uniformly bounded away from 0. Hence we obtain coercivity of the bilinear form  $\mathcal{A}$  via

$$\begin{aligned} \mathcal{A}((v, z), (v, z)) &= \sum_{T \in \mathcal{T}} \langle D_T v_T, v_T \rangle + \sum_{T \in \mathcal{T}} \langle z_T, V_T z_T \rangle \\ &\geq C \sum_{T \in \mathcal{T}} \langle S_T v_T, v_T \rangle + \sum_{T \in \mathcal{T}} \langle z_T, V_T z_T \rangle \geq C \|(v, z)\|_{\mathcal{X}}^2 \end{aligned}$$

with  $C = \min_{T \in \mathcal{T}} c_{D,T}^* \in (0, 1)$ .

In order to get upper bounds, we again use (6.9) as well as the Cauchy-Schwarz inequality for the symmetric and positive definite or semidefinite forms  $\langle \cdot, V_T \cdot \rangle$  and  $\langle D_T \cdot, \cdot \rangle$  to see that

$$|d(u, v)| \leq \|u\|_S \|v\|_S, \quad |c(t, z)| \leq \|t\|_V \|z\|_V.$$

By duality of the norms  $\|\cdot\|_{V_T}$  and  $\|\cdot\|_{V_T^{-1}}$ , we get

$$\begin{aligned} b(v, t) &= \sum_{T \in \mathcal{T}} \langle t_T, (\frac{1}{2}I + K_T)v_T \rangle \leq \sum_{T \in \mathcal{T}} \|t_T\|_{V_T} \|(\frac{1}{2}I + K_T)v_T\|_{V_T^{-1}} \\ &\stackrel{(3.8)}{\leq} C \sum_{T \in \mathcal{T}} \|t_T\|_{V_T} |v_T|_{S_T} \leq C \|t\|_V \|v\|_S. \end{aligned}$$

Here we used (3.8), that is,  $\|(\frac{1}{2}I + K_T)v_T\|_{V_T^{-1}} \leq c_{K,T}(1 - c_{K,T})^{-1/2}|v_T|_{S_T}$ . Combined, the above bounds yield

$$\begin{aligned} |\mathcal{A}((u, t), (v, z))| &\leq C(\|u\|_S \|v\|_S + \|t\|_V \|v\|_S + \|u\|_S \|z\|_V + \|t\|_V \|z\|_V) \\ &= C(\|u\|_S + \|t\|_V)(\|v\|_S + \|z\|_V) \\ &\leq 2C \|(u, t)\|_{\mathcal{X}} \|(v, z)\|_{\mathcal{X}}. \end{aligned}$$

All in all, the bilinear form  $\langle \mathcal{A}\cdot, \cdot \rangle$  is uniformly bounded and coercive in the  $\mathcal{X}$ -norm, and thus the statement follows, after passing to the homogenized formulation (4.15), with Lemma 2.2.3(Céa).  $\square$

While error estimates on the skeleton follow directly from this result and Theorem 6.6.1, they are inherently mesh-dependent and therefore of limited use. More interesting is the error within the domain with respect to the exact solution  $u_\Omega$  of (2.4), which will typically have additional regularity. As before, we will assume  $u_\Omega \in H^2(\Omega)$ . We have already seen in the proof of Theorem 6.5.1 that it suffices to bound the error  $\mathcal{H}_S(u - u_h)$ .

### 6.6.2 Convergence in the $H^1$ -norm

Let  $(u, t) \in \mathcal{X}$  be the exact solution of (4.6) and  $(u_h, t_h) \in \mathcal{X}_h$  be the solution of (4.13). From (6.2), Theorem 6.6.2, and Theorem 6.6.1, we immediately obtain

$$|\mathcal{H}_S(u - u_h)|_{H^1(\Omega)} = \|u - u_h\|_S \leq \|(u - u_h, t - t_h)\|_{\mathcal{X}} \leq Ch |u_\Omega|_{H^2(\Omega)}.$$

This is nothing but the statement of Theorem 6.5.1, which was there proved using the primal formulation. The equivalent mixed formulation thus leads to the same  $H^1$ -error estimate.

### 6.6.3 Convergence in the $L_2$ -norm

The proof of the error estimate in the  $L_2$ -norm proceeds by a standard Aubin-Nitsche duality argument. We thus consider the adjoint variational problem: find  $w \in H_0^1(\Omega)$  such that

$$\int_{\Omega} \nabla v \cdot \nabla w \, dx = \int_{\Omega} \mathcal{H}_S(u - u_h) v \, dx \quad \forall v \in H_0^1(\Omega). \quad (6.10)$$

Under assumptions on the regularity of this adjoint problem, we have the following result.

**Theorem 6.6.3.** *Let the assumptions of Theorem 6.6.1 be satisfied. Furthermore, assume that the adjoint problem (6.10) is  $H^2$ -coercive, and that the solution  $u_\Omega$  of the variational problem (2.4) belongs to  $H^2(\Omega)$ . Then the quasi-optimal  $L_2$  discretization error estimate*

$$\|\mathcal{H}_S(u - u_h)\|_{L_2(\Omega)} \leq Ch^2 |u_\Omega|_{H^2(\Omega)} \quad (6.11)$$

*holds.*

*Proof.* Due to the assumption of the regularity of the adjoint problem, the solution  $w \in H_0^1(\Omega)$  of the adjoint problem (6.10) lies in  $H^2(\Omega)$  and satisfies the estimate

$$|w|_{H^2(\Omega)} \leq C \|\mathcal{H}_S(u - u_h)\|_{L_2(\Omega)}. \quad (6.12)$$

Because of the equivalence of the standard and the skeletal variational formulation, Proposition 4.1.2, its skeletal traces  $(\phi, \eta)$ , where  $\phi_T := \gamma_T^0 w$ ,  $\eta_T := \gamma_T^1 w$  for  $T \in \mathcal{T}$ , satisfy the (adjoint) mixed skeletal variational formulation (4.6), i.e.,

$$\mathcal{A}((v, z), (\phi, \eta)) = \int_{\Omega} \mathcal{H}_S(u - u_h) \mathcal{H}_S v \, dx \quad \forall (v, z) \in \mathcal{X}_0.$$

In particular, with the choice  $(v, z) = (u - u_h, t - t_h)$  and exploiting the Galerkin orthogonality (4.14) as well as the uniform boundedness of  $\mathcal{A}$  shown in the proof of Theorem 6.6.2, we get

$$\begin{aligned} \|\mathcal{H}_S(u - u_h)\|_{L_2(\Omega)}^2 &= \mathcal{A}((u - u_h, t - t_h), (\phi, \eta)) \\ &= \mathcal{A}((u - u_h, t - t_h), (\phi - \phi_h, \eta - \eta_h)) \\ &\leq C \|(u - u_h, t - t_h)\|_{\mathcal{X}} \|(\phi - \phi_h, \eta - \eta_h)\|_{\mathcal{X}} \end{aligned}$$

for arbitrary  $(\phi_h, \eta_h) \in \mathcal{X}_{h,0}$ . Taking the infimum over  $(\phi_h, \eta_h)$  and applying Theorem 6.6.2 and Theorem 6.6.1, we obtain

$$\|\mathcal{H}_S(u - u_h)\|_{L_2(\Omega)}^2 \leq C h^2 |u_{\Omega}|_{H^2(\Omega)} |w|_{H^2(\Omega)}.$$

Using now estimate (6.12), we arrive at the desired  $L_2$ -error estimate.  $\square$



# Chapter 7

## Fast solution methods

As we have pointed out earlier, the system of linear equations that the BEM-based FEM discretization results in shares many properties with that obtained in a standard finite element method: it is large for any reasonably small mesh size, but sparsely populated, and furthermore symmetric (if the underlying partial differential operator is formally self-adjoint) and positive definite (if the partial differential operator is coercive). Techniques for solving such linear systems are generally divided into direct and iterative solvers.

While the classical direct solvers using Gaussian elimination or LU decomposition are not a good match for these kinds of systems due to excessive fill-in of the zeroes of the matrix, sophisticated direct solvers have been developed with sparse systems in mind which perform very well and are often competitive for two-dimensional problems of moderately large size. We mention, for example, UMFPACK [28], PARDISO [108], and SuperLU [81]. For a numerical comparison of several direct solvers for sparse systems, see [53].

Iterative solvers are a highly successful class of methods for solving large, sparse systems of linear equations. Among the most important developments in this field are the invention of the method of conjugate gradients (CG) [57] for the solution of symmetric, positive definite systems, of the method of minimal residuals (MinRes) [92] for symmetric, indefinite systems, and of the method of generalized minimal residuals (GMRes) [105] for nonsymmetric systems. Iterative solvers share the characteristic that, instead of storing the full matrix, only the action of the matrix on a given vector needs to be available, and that they benefit greatly from preconditioning in order to speed up the convergence of the iterative procedure.

Modern solvers for discretized partial differential equations are often iterative solvers based on either a multigrid or a domain decomposition paradigm. While early numerical tests using algebraic multigrid solvers for BEM-based FEM discretizations have shown promising results [25], in this chapter we derive an iterative solver which is based on domain decomposition. In particular, we follow the ideas of the one-level finite-element tearing/interconnecting (FETI) substructuring technique that was originally proposed by Farhat and Roux [35]. Since then, FETI methods have been established in theory and practice as highly efficient solvers for discretized partial differential equations. A major advantage of domain decomposition and in particular FETI solvers over multigrid

approaches is the ease with which they can be parallelized due to the independence of the local subproblems.

In this chapter, we restrict ourselves to diffusion problems of the form

$$\begin{aligned} -\operatorname{div}(\alpha \nabla u) &= 0 && \text{in } \Omega, \\ \gamma_{\Omega}^0 u &= g && \text{on } \partial\Omega \end{aligned}$$

with piecewise constant diffusion coefficient  $\alpha(x) = \alpha_T > 0 \forall x \in T$  and piecewise linear Dirichlet data  $g$ . Due to (2.7), we have the energy

$$\langle S_T v, v \rangle = \alpha_T |\mathcal{H}_T v|_{H^1(T)}^2. \quad (7.1)$$

with the Dirichlet-to-Neumann map  $S_T$  on every element  $T$ .

The discretized variational problem (4.11), or rather the homogenized version (4.12), can be written as the operator equation: find  $u_h \in \mathcal{W}_{h,0}$  with

$$A u_h = \tilde{F} \quad (7.2)$$

with  $A : \mathcal{W}_h \rightarrow \mathcal{W}_h^*$  given by the mapping

$$A(u) : v \mapsto \sum_{T \in \mathcal{T}} \langle \tilde{S}_T u_T, v_T \rangle$$

and  $\tilde{F} = -A g \in \mathcal{W}_h^*$ , where  $g$  was assumed to be piecewise linear on the skeleton. The associated stiffness matrix is given by

$$\underline{A} = (\langle A \phi_l, \phi_k \rangle)_{k,l},$$

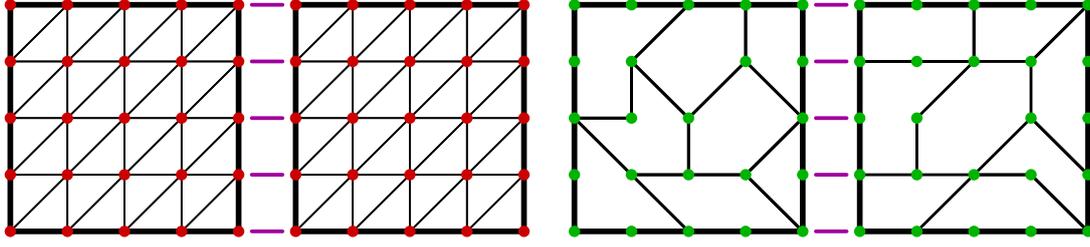
where  $\{\phi_k\}$  forms a nodal basis for  $\mathcal{W}_{h,0}$  such that  $\phi_k$  is 1 in the  $k$ -th skeletal node, 0 in all others, and interpolated linearly on every (simplicial) facet  $\tau \in \mathcal{F}$  of the skeleton.

## 7.1 A FETI-type solver

### 7.1.1 Derivation

Our derivation follows that of the classical FETI method closely, and we therefore refer to the literature, e.g., [35, 83, 76, 117], for further details and some omitted proofs.

We decompose the computational domain  $\Omega$  into non-overlapping subdomains  $(\Omega_i)_{i=1}^N$  in agreement with the polytopal mesh  $\mathcal{T}$ , that is,  $\bar{\Omega}_i = \bigcup_{T \in \mathcal{T}_i} \bar{T}$  with a corresponding decomposition  $(\mathcal{T}_i)_{i=1}^N$  of the mesh. We set  $H_i := \operatorname{diam} \Omega_i$  and  $H := \max_{i=1}^N H_i$ . Every subdomain  $\Omega_i$  has an associated skeleton  $\bigcup_{T \in \mathcal{T}_i} \partial T$  and discrete skeletal trial spaces  $\mathcal{W}_h(\Omega_i)$  and  $\mathcal{W}_{h,0}(\Omega_i)$  constructed as in Section 4.3.1. Both the operator  $A$  and the



**Figure 7.1:** Sketch of domain decomposition approach in 2D for a rectangular domain with  $N = 2$  subdomains. *Left:* FETI substructuring. *Right:* FETI-like substructuring for the BEM-based FEM.

functional  $\tilde{F}$  can be written as a sum of local contributions  $A_i : \mathcal{W}_h(\Omega_i) \rightarrow \mathcal{W}_h(\Omega_i)^*$  given by

$$A_i(u) : v \mapsto \sum_{T \in \mathcal{T}_i} \langle \tilde{S}_T u_T, v_T \rangle$$

and, analogously,  $\tilde{F}_i \in \mathcal{W}_h(\Omega_i)^*$ . This yields the formulation

$$\sum_{i=1}^N A_i(u_h|_{\Omega_i}) = \sum_{i=1}^N \tilde{F}_i.$$

Motivated by Theorem 2.2.2, we rewrite this discrete problem as the minimization problem

$$u = \arg \min_{v \in \mathcal{W}_{h,0}(\Omega)} \frac{1}{2} \sum_{i=1}^N \langle A_i v|_{\Omega_i}, v|_{\Omega_i} \rangle - \sum_{i=1}^N \langle \tilde{F}_i, v|_{\Omega_i} \rangle,$$

where here and in the sequel we drop the subscript  $h$  since all functions are discrete from now on. Indeed, all relevant functions live in spaces of piecewise linear functions which have natural nodal bases. Therefore, we will not distinguish in the following between functions and the coefficient vectors representing them with respect to the nodal basis, nor between operators and their matrix representations.

It turns out that the interior unknowns in every subdomain  $\Omega_i$ , i.e., those not lying on  $\partial\Omega_i$ , can be eliminated by passing to a Schur complement formulation. For this, we introduce the Schur complement

$$\tilde{S}_i = A_{i,\Gamma\Gamma} - A_{i,\Gamma I} (A_{i,II})^{-1} A_{i,I\Gamma},$$

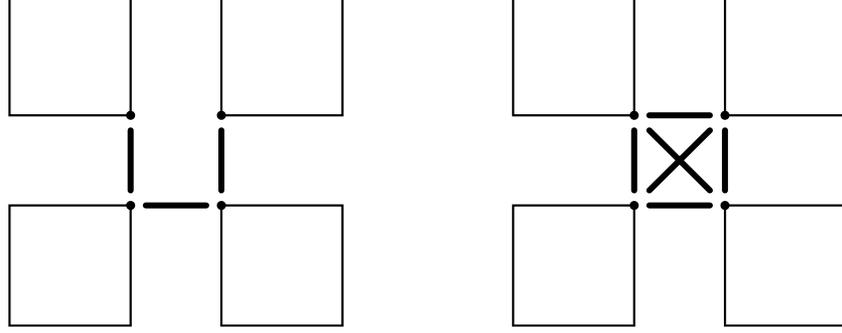
where the blocks  $A_{i,\Gamma\Gamma}, A_{i,\Gamma I}, A_{i,I\Gamma}, A_{i,II}$  are chosen such that for a discrete function  $w \in \mathcal{W}_h(\Omega_i)$  with boundary degrees of freedom  $w_\Gamma$  and inner degrees of freedom  $w_I$ , there holds

$$A_i w = \begin{bmatrix} A_{i,\Gamma\Gamma} & A_{i,\Gamma I} \\ A_{i,I\Gamma} & A_{i,II} \end{bmatrix} \begin{bmatrix} w_\Gamma \\ w_I \end{bmatrix}.$$

With this, we can write the equivalent formulation to the above minimization problem,

$$u = \arg \min_{v \in \mathcal{W}_{h,0}(\Gamma_S^H)} \frac{1}{2} \sum_{i=1}^N \langle \tilde{S}_i v|_{\partial\Omega_i}, v|_{\partial\Omega_i} \rangle - \sum_{i=1}^N \langle g_i, v|_{\partial\Omega_i} \rangle, \quad (7.3)$$

where  $\Gamma_S^H = \bigcup_{i=1}^N \partial\Omega_i$  is the coarse skeleton,  $\mathcal{W}_{h,0}(\Gamma_S^H)$  is the trace space of discrete functions  $\mathcal{W}_{h,0}(\Omega)$  onto  $\Gamma_S^H$ , and  $g_i$  is a suitably adjusted forcing term.



**Figure 7.2:** Constraints at the intersection between four subdomains. *Left:* a choice of non-redundant constraints. *Right:* fully redundant constraints.

Let  $\mathcal{W}_h(\partial\Omega_i) := \{v|_{\partial\Omega_i} : v \in \mathcal{W}_h(\Omega_i)\}$  denote a space of discrete boundary functions. We then introduce the broken space  $Y$  by

$$Y_i := \{v \in \mathcal{W}_h(\partial\Omega_i) : v|_{\Gamma_D} = 0\}, \quad Y := \prod_{i=1}^N Y_i.$$

Functions from  $Y$  may have two different values on either side of a subdomain interface. Only if their values match across interfaces can they be identified with functions in  $\mathcal{W}_{h,0}(\Gamma_S^H)$ . In order to enforce this, we introduce the jump operator

$$B : Y \rightarrow \mathbb{R}^{N_\Lambda},$$

where  $N_\Lambda \in \mathbb{N}$  is the total number of constraints. In the nodal basis,  $B$  has exactly two non-zero contributions of opposite sign per constraint and may thus be viewed as a signed Boolean matrix. In this work, we assume fully redundant constraints, i.e., for every node on a subdomain interface, constraints corresponding to all neighboring subdomains are introduced. This is in contrast to the non-redundant case, where only a minimal set of constraints to ensure continuity is introduced. See Figure 7.2 for an illustration. The choice of redundant constraints implies that the jump operator  $B$  is not surjective, and we define the space of Lagrange multipliers as the range

$$\Lambda := \text{Range } B \subseteq \mathbb{R}^{N_\Lambda}$$

and consider  $B$  as a mapping  $Y \rightarrow \Lambda$  in the following in order to avoid some subtle technical difficulties in the analysis.

The jump operator can be written as a sum of local contributions  $B_i : Y_i \rightarrow \Lambda$ , and the globally consistent functions in  $Y$  are those which satisfy

$$By = \sum_{i=1}^N B_i y_i = 0,$$

that is,  $y \in \ker B$ . In light of this, we rewrite (7.3) as

$$u = \arg \min_{\substack{y \in Y \\ By=0}} \frac{1}{2} \sum_{i=1}^N \langle \tilde{S}_i y_i, y_i \rangle - \sum_{i=1}^N \langle g_i, y_i \rangle.$$

Introducing Lagrange multipliers to enforce the constraint  $By = 0$ , we obtain the saddle point formulation

$$\begin{bmatrix} \tilde{S} & B^\top \\ B & 0 \end{bmatrix} \begin{bmatrix} u \\ \lambda \end{bmatrix} = \begin{bmatrix} g \\ 0 \end{bmatrix} \quad (7.4)$$

for  $u \in Y$  and  $\lambda \in \Lambda$ , where we used the block matrices and vectors  $\tilde{S} = \text{diag}(\tilde{S}_1, \dots, \tilde{S}_N)$ ,  $B = (B_1, \dots, B_N)$ ,  $u = (u_1, \dots, u_N)^\top$ ,  $g = (g_1, \dots, g_N)^\top$ . From (7.4), we see that the local skeletal functions  $u_i$  satisfy the relationship

$$\tilde{S}_i u_i = g_i - B_i^\top \lambda. \quad (7.5)$$

For a *non-floating* domain  $\Omega_i$ , that is, one that shares a part of the Dirichlet boundary such that  $\partial\Omega_i \cap \Gamma_D \neq \emptyset$ ,  $\tilde{S}_i$  is positive definite and thus invertible. For a *floating* domain  $\Omega_i$ , the kernel of  $\tilde{S}_i$  consists only of the constant functions, and we parameterize it by the operator

$$\begin{aligned} R_i : \mathbb{R} &\rightarrow \ker \tilde{S}_i \subset Y_i, \\ \xi_i &\mapsto \xi_i. \end{aligned}$$

Under the condition that the right-hand side is orthogonal to the kernel, i.e.,

$$\langle g_i - B_i^\top \lambda, R_i \zeta \rangle = 0 \quad \forall \zeta \in \mathbb{R}, \quad (7.6)$$

the local problem (7.5) is solvable and we have

$$u_i = \tilde{S}_i^\dagger (g_i - B_i^\top \lambda) + R_i \xi_i$$

with some  $\xi_i \in \mathbb{R}$ . Above,  $\tilde{S}_i^\dagger$  denotes a pseudo-inverse of  $\tilde{S}_i$ . For easier analysis later on, we make the particular choice

$$\tilde{S}_i^\dagger = (\tilde{S}_i + \beta_i R_i R_i^\top)^{-1}$$

for some  $\beta_i > 0$ . We note that for practical reasons, different choices might be preferable and refer to [36] for further discussion.

We set  $Z := \prod_{i=1}^N \mathbb{R}^{\dim(\ker \tilde{S}_i)}$ , define the operator

$$R : Z \rightarrow Y, \quad (R\xi)|_{\Omega_i} := \begin{cases} R_i \xi_i, & \text{if } \Omega_i \text{ is floating,} \\ 0, & \text{if } \Omega_i \text{ is non-floating,} \end{cases}$$

and set  $\tilde{S}_i^\dagger = \tilde{S}_i^{-1}$  for non-floating domains  $\Omega_i$ . The local solutions  $u$  can then be expressed by

$$u = \tilde{S}^\dagger (g - B^\top \lambda) + R\xi \quad (7.7)$$

under the compatibility condition (derived from (7.6))

$$R^\top B^\top \lambda = R^\top g.$$

Formula (7.7) allows us to eliminate the unknowns  $u \in Y$  from the saddle point problem (7.4). Indeed, inserting (7.7) into the second line of (7.4) yields

$$B\tilde{S}^\dagger g - B\tilde{S}^\dagger B^\top \lambda + BR\xi = 0,$$

and together with the compatibility condition and using the notations  $F = B\tilde{S}^\dagger B^\top$  and  $G = BR$ , we obtain the dual saddle point problem

$$\begin{bmatrix} F & -G \\ G^\top & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \xi \end{bmatrix} = \begin{bmatrix} B\tilde{S}^\dagger g \\ R^\top g \end{bmatrix}. \quad (7.8)$$

With a self-adjoint operator  $Q : \Lambda \rightarrow \Lambda$  which is positive definite on the range of  $G$  and to be chosen later, we define

$$P = I - QG(G^\top QG)^{-1}G^\top.$$

It can be shown that  $G^\top QG$  is positive definite and thus indeed invertible, and that  $P$  is a projection from  $\Lambda$  onto the subspace  $\Lambda_0$  of admissible increments,

$$\Lambda_0 := \ker G^\top \subset \Lambda.$$

The choice

$$\lambda_g := QG(G^\top QG)^{-1}R^\top g \in \Lambda$$

ensures that  $G^\top \lambda_g = R^\top g$ , and thus with  $\lambda = \lambda_0 + \lambda_g$  we can homogenize (7.8) such that we only search for a  $\lambda_0 \in \Lambda_0$  with

$$F\lambda_0 - G\xi = B\tilde{S}^\dagger g - F\lambda_g. \quad (7.9)$$

Applying the projector  $P^\top$  to this equation and noting that  $P^\top G = 0$ , we obtain the following formulation of the dual problem: find  $\lambda_0 \in \Lambda_0$  such that

$$P^\top F \lambda_0 = P^\top (B \tilde{S}^\dagger g - F \lambda_g) = P^\top B \tilde{S}^\dagger (g - B^\top \lambda_g). \quad (7.10)$$

It can be shown that  $P^\top F$  is self-adjoint and positive definite on  $\Lambda_0$ , and thus the problem (7.10) has a unique solution which may be computed by Conjugate Gradient (CG) iteration in the subspace  $\Lambda_0$ . Once  $\lambda = \lambda_0 + \lambda_g$  has been computed, we see that applying  $(G^\top QG)^{-1} G^\top Q$  to (7.9) yields the formula

$$\xi = (G^\top QG)^{-1} G^\top Q B \tilde{S}^\dagger (B^\top \lambda - g)$$

for  $\xi$ . The local unknowns  $u_i$  may then be obtained by solving the local problems (7.7), and the eliminated unknowns in the interior of each  $\Omega_i$  may be recovered by solving local Dirichlet problems.

### 7.1.2 Preconditioning

Preconditioners for FETI are typically constructed in the form  $PM^{-1}$  with a suitable operator  $M^{-1} : \Lambda \rightarrow \Lambda$ . The Dirichlet preconditioner proposed by Farhat, Mandel, and Roux [37], adapted to our setting, is given by the choice

$$M^{-1} = B \tilde{S} B^\top.$$

Applying  $M^{-1}$  to a vector requires solving  $N$  Dirichlet boundary value problems, from which its name stems. It is known to work well for constant or mildly varying diffusion coefficient  $\alpha$ , and in this case, the choice  $Q = I$  works satisfactorily.

To deal with coefficient jumps, we need to employ a *scaled* or *weighted jump operator* as introduced by Rixen and Farhat [101] in a mechanical setting and later analyzed by Klawonn and Widlund [76]. For this, let  $x^h \in \partial\Omega_i$  refer to an arbitrary boundary node and introduce scalar weights

$$\rho_i(x^h) > 0.$$

We will restrict ourselves to the case of subdomain-wise constant coefficient  $\alpha$  in the following, i.e.,

$$\alpha(x) = \alpha_i \quad \forall x \in \Omega_i.$$

For a comprehensive treatment of the case of an unresolved diffusion coefficient, we refer to the forthcoming monograph [96]. In the setting of piecewise constant  $\alpha$ , we simply choose the weight

$$\rho_i(x^h) = \alpha_i.$$

We use these weights to define weighted counting functions  $\delta_j$ ,  $j = 1, \dots, N$ , by the nodal values

$$\delta_j(x^h) := \begin{cases} \frac{\rho_j(x^h)}{\sum_{k \in \mathcal{N}(x^h)} \rho_k(x^h)}, & x^h \in \partial\Omega_j, \\ 0, & \text{otherwise,} \end{cases}$$

and piecewise linear interpolation on the facets of the coarse skeleton  $\Gamma_S^H$ . Above,  $\mathcal{N}(x^h) = \{i \in \{1, \dots, N\} : x^h \in \partial\Omega_i\}$  is the set of subdomains on whose boundaries  $x^h$  lies. The family of counting functions  $\{\delta_j\}_{j=1}^N$  forms a partition of unity on the skeleton.

We now introduce diagonal scaling matrices  $D_i : \Lambda \rightarrow \Lambda$ ,  $i = 1, \dots, N$ , operating on the space of Lagrange multipliers. Consider two neighboring domains  $\Omega_i$  and  $\Omega_j$  sharing a node  $x^h \in \partial\Omega_i \cap \partial\Omega_j$ . Let  $k \in \{1, \dots, N_\Lambda\}$  denote the index of the Lagrange multiplier associated with this node and pair of subdomains. Then, the  $k$ -th diagonal entry of  $D_i$  is set to  $\delta_j(x^h)$ , and the  $k$ -th diagonal entry of  $D_j$  is set to  $\delta_i(x^h)$ . Diagonal entries of  $D_i$  not associated with a node on  $\partial\Omega_i$  are set to zero.

The *weighted jump operator*  $B_D : Y \rightarrow \Lambda$  is now given by

$$B_D = [D_1 B_1, \dots, D_N B_N],$$

and the weighted Dirichlet preconditioner by

$$M_D^{-1} = B_D \tilde{S} B_D^\top.$$

In this case, a possible choice for  $Q$  is simply  $Q = M_D^{-1}$ . Alternatively,  $Q$  can be replaced by a suitable diagonal matrix as described in [76].

## 7.2 Convergence analysis

### 7.2.1 The non-preconditioned case

For the purpose of the convergence analysis, we again assume that Assumption 6.2.2 holds. In particular, this means that every subdomain  $\Omega_i$  has a shape-regular simplicial triangulation  $\Xi_i$  which triangulates the elements in  $\mathcal{T}_i$ . i.e., polytopal elements in  $\mathcal{T}_i$  are unions of simplices from  $\Xi_i$ . Its global mesh size will be denoted  $h^F = \max_i \max_{\tau \in \Xi_i} \text{diam } \tau$ . Since, by assumption, every  $T \in \mathcal{T}_i$  consists of only “a few” simplices, we have  $h \leq Ch^F$  with a small uniform constant  $C$ .

On these meshes, we construct standard piecewise linear finite element spaces  $\mathcal{V}_h(\Omega_i)$  as well as  $\mathcal{V}_{h,0}(\Omega_i) := \{w \in \mathcal{V}_h(\Omega_i) : w|_{\Gamma_D} = 0\}$  as in Section 2.5. We note that, by construction, the spaces  $Y_i$  introduced above are just the traces of  $\mathcal{V}_{h,0}(\Omega_i)$  onto  $\partial\Omega_i$ .

Discretizing the usual bilinear form  $\int \alpha \nabla u \cdot \nabla v$  and employing the standard nodal basis for the FE spaces, we obtain the standard FE stiffness matrix  $K_i$  on every subdomain  $\Omega_i$ . Eliminating, as in Section 7.1, the interior unknowns, we obtain the Schur complement

$$S_i^F = K_{i,\Gamma\Gamma} - K_{i,\Gamma I} (K_{i,II})^{-1} K_{i,I\Gamma}$$

which provides a FEM approximation of the Steklov-Poincaré operator.

For the exact Steklov-Poincaré operator for  $\Omega_i$  as introduced in (2.8), we write

$$S_i = S_{\Omega_i} : H^{1/2}(\partial\Omega_i) \rightarrow H^{-1/2}(\partial\Omega_i).$$

It satisfies, for all  $u \in H^{1/2}(\partial\Omega_i)$ ,

$$\langle S_i u, u \rangle = \inf_{\substack{v \in H^1(\Omega_i) \\ v|_{\partial\Omega_i} = u}} \int_{\Omega_i} \alpha \nabla v \cdot \nabla v \, dx = \inf_{\substack{v \in \mathcal{W}(\Omega_i) \\ v|_{\partial\Omega_i} = u}} a_i(v, v) \quad (7.11)$$

where  $\mathcal{W}(\Omega_i) = H^{1/2}(\bigcup_{T \in \mathcal{T}_i} \partial T)$  is the trace space of  $H^1(\Omega_i)$  onto the skeleton of  $\Omega_i$  as defined in Section 4.1, and the bilinear form

$$a_i(u, v) = \sum_{T \in \mathcal{T}_i} \langle S_T u|_{\partial T}, v|_{\partial T} \rangle.$$

is the analogue of  $a(\cdot, \cdot)$  from (4.8) on  $\Omega_i$ . Denoting by

$$\mathcal{H}_i = \mathcal{H}_{\Omega_i} : H^{1/2}(\partial\Omega_i) \rightarrow H^1(\Omega_i)$$

the PDE-harmonic extension operator on the subdomain  $\Omega_i$  as introduced in Definition 2.4.1, we see that  $v = \mathcal{H}_i u$  is just the  $H^1(\Omega_i)$ -function which minimizes the energy in (7.11), and thus it is clear that

$$\langle S_i u, u \rangle = \sum_{T \in \mathcal{T}_i} \alpha_T |\mathcal{H}_i u|_{H^1(T)}^2. \quad (7.12)$$

The FEM Schur complement  $S_i^F$  is spectrally equivalent to the Steklov-Poincaré operator for discrete functions in the following sense.

**Theorem 7.2.1.** *For discrete functions, the FEM Schur complement  $S_i^F$  satisfies the spectral equivalence*

$$\langle S_i u, u \rangle \leq \langle S_i^F u, u \rangle \leq c_{\Pi}^2 \frac{\bar{\alpha}_i}{\underline{\alpha}_i} \langle S_i u, u \rangle \quad \forall u \in \mathcal{W}_h(\partial\Omega_i), \quad (7.13)$$

where  $c_{\Pi}$  is the interpolation constant from Theorem 2.5.2 and thus depends on the shape regularity of the mesh  $\Xi_i$ , and  $\underline{\alpha}_i$  and  $\bar{\alpha}_i$  are lower and upper bounds, respectively, for the diffusion coefficients  $\{\alpha(x) : x \in \Omega_i\}$ .

*Proof.* The proof is quite standard but short, and we give it for completeness.

By the minimizing property of the Schur complement, it is clear that

$$\langle S_i^F u, u \rangle = \inf_{\substack{v \in \mathcal{V}_h(\Omega_i) \\ v|_{\partial\Omega_i} = u}} \int_{\Omega_i} \alpha \nabla v \cdot \nabla v \, dx \quad \forall u \in \mathcal{W}_h(\partial\Omega_i),$$

and the lower bound thus follows from (7.11) and the fact that  $\mathcal{V}_h(\Omega_i) \subset H^1(\Omega_i)$ . On the other hand, with the special choice  $v = \Pi \mathcal{H}_i u$ , where  $\Pi : H^1(\Omega_i) \rightarrow \mathcal{V}_h(\Omega_i)$  is the Scott-Zhang interpolator from Theorem 2.5.2, and using (7.12), we get

$$\langle S_i^F u, u \rangle \leq \sum_{T \in \mathcal{T}_i} \alpha_T |\Pi \mathcal{H}_i u|_{H^1(T)}^2 \leq c_{\Pi}^2 \bar{\alpha}_i |\mathcal{H}_i u|_{H^1(T)}^2 \leq c_{\Pi}^2 \frac{\bar{\alpha}_i}{\underline{\alpha}_i} \langle S_i u, u \rangle. \quad \square$$

In the following we show that the BEM-based FEM Schur complement satisfies a very similar spectral equivalence. Note that this result may be viewed as generalizing a result from the work of Langer and Steinbach on BETI [79]: where the authors therein analyzed the spectral equivalence for a single BEM domain, we consider the case where a subdomain may consist of arbitrarily many BEM domains which are coupled symmetrically.

**Theorem 7.2.2.** *For discrete functions, the BEM-based FEM Schur complement  $\tilde{S}_i$  satisfies the spectral equivalence*

$$\tilde{c}_i \langle S_i u, u \rangle \leq \langle \tilde{S}_i u, u \rangle \leq c_{\Pi}^2 \frac{\bar{\alpha}_i}{\underline{\alpha}_i} \langle S_i u, u \rangle \quad \forall u \in \mathcal{W}_h(\partial\Omega_i), \quad (7.14)$$

where the constant  $\tilde{c}_i = \min_{T \in \mathcal{T}_i} \{\tilde{c}_T\} \in (0, \frac{1}{4}]$  is the smallest BEM contraction constant,  $c_{\Pi}$  is the interpolation constant from Theorem 2.5.2, and  $\underline{\alpha}_i$  and  $\bar{\alpha}_i$  are lower and upper bounds, respectively, for the diffusion coefficients  $\{\alpha(x) : x \in \Omega_i\}$ .

*Proof.* Fix  $u \in \mathcal{W}_h(\partial\Omega_i)$ . Recall that  $\tilde{S}_i$  was defined as the Schur complement of  $A_i$ , and hence, by the minimizing property of the Schur complement, it satisfies

$$\langle \tilde{S}_i u, u \rangle = \inf_{\substack{v \in \mathcal{W}_h(\Omega_i) \\ v|_{\partial\Omega_i} = u}} \tilde{a}_i(v, v). \quad (7.15)$$

Due to  $\mathcal{W}_h(\Omega_i) \subset \mathcal{W}(\Omega_i)$  and then estimating using the element-level spectral equivalences (3.10) and the second equality in (7.11), we have

$$\begin{aligned} \langle \tilde{S}_i u, u \rangle &= \inf_{\substack{v \in \mathcal{W}_h(\Omega_i) \\ v|_{\partial\Omega_i} = u}} \sum_{T \in \mathcal{T}_i} \langle \tilde{S}_T v_T, v_T \rangle \\ &\geq \inf_{\substack{v \in \mathcal{W}(\Omega_i) \\ v|_{\partial\Omega_i} = u}} \sum_{T \in \mathcal{T}_i} \langle \tilde{S}_T v_T, v_T \rangle \\ &\geq \inf_{\substack{v \in \mathcal{W}(\Omega_i) \\ v|_{\partial\Omega_i} = u}} \sum_{T \in \mathcal{T}_i} \tilde{c}_T \langle S_T v_T, v_T \rangle \\ &\geq \min_{T \in \mathcal{T}_i} \{\tilde{c}_T\} \inf_{\substack{v \in \mathcal{W}(\Omega_i) \\ v|_{\partial\Omega_i} = u}} \sum_{T \in \mathcal{T}_i} \langle S_T v_T, v_T \rangle = \tilde{c}_i \langle S_i u, u \rangle. \end{aligned}$$

This proves the lower bound.

For the upper bound, we recall the Scott-Zhang quasi-interpolation operator  $\Pi : H^1(\Omega_i) \rightarrow \mathcal{V}_h(\Omega_i)$  from Theorem 2.5.2, now defined only on  $\Omega_i$  respectively its mesh  $\Xi_i$ . Let  $\phi := \Pi \mathcal{H}_i u \in \mathcal{V}_h(\Omega_i)$  denote the interpolant of the energy-minimizing function. By restricting its values to the skeleton  $\bigcup_{T \in \mathcal{T}_i} \partial T$ , we get a skeletal interpolant  $\psi \in \mathcal{W}_h(\Omega_i)$  as in Theorem 6.3.1. By the properties of the Scott-Zhang interpolator,  $\psi$  matches  $u$  on the boundary  $\partial\Omega_i$ .

From (7.15), it is clear that

$$\langle \tilde{S}_i u, u \rangle \leq \tilde{a}_i(\psi, \psi) = \sum_{T \in \mathcal{T}_i} \langle \tilde{S}_T \psi_T, \psi_T \rangle,$$

and using the spectral equivalence (3.10), the element energy identity (7.1), the energy-minimizing property of the harmonic extension (Theorem 2.4.2), the stability of the Scott-Zhang interpolator, and (7.12), we may further estimate

$$\begin{aligned} \langle \tilde{S}_i u, u \rangle &\leq \sum_{T \in \mathcal{T}_i} \langle S_T \psi_T, \psi_T \rangle = \sum_{T \in \mathcal{T}_i} \alpha_T |\mathcal{H}_T \psi_T|_{H^1(T)}^2 \\ &\leq \sum_{T \in \mathcal{T}_i} \alpha_T |\phi|_{H^1(T)}^2 \leq \bar{\alpha}_i |\phi|_{H^1(\Omega_i)}^2 \leq c_{\Pi}^2 \bar{\alpha}_i |\mathcal{H}_i u|_{H^1(\Omega_i)}^2 \leq c_{\Pi}^2 \frac{\bar{\alpha}_i}{\underline{\alpha}_i} \langle S_i u, u \rangle, \end{aligned}$$

which proves the upper bound.  $\square$

The spectral equivalences from Theorem 7.2.1 and Theorem 7.2.2 allow us to prove similar equivalences for the pseudo-inverses of the BEM-based FEM Schur complement and of the FEM Schur complement. In the following, we will generally assume that the diffusion coefficient  $\alpha(x)$  is constant on each subdomain  $\Omega_i$ , i.e.,  $\alpha(x) = \alpha_i \forall x \in \Omega_i$ , such that the ratio between  $\bar{\alpha}_i$  and  $\underline{\alpha}_i$  is 1. We will then write  $U \cong V$  for a spectral equivalence between the matrices  $U$  and  $V$  where the equivalence constants depend only on mesh regularity parameters.

**Lemma 7.2.3.** *For subdomain-wise constant diffusion coefficient  $\alpha(x)$  and under Assumption 6.2.2, we have*

$$\tilde{S}_i^\dagger \cong (S_i^F)^\dagger.$$

*Proof.* We first observe that, due to Theorem 6.2.4 and Theorem 2.5.2, the constants in Theorem 7.2.1 and Theorem 7.2.2 depend only on mesh regularity and we thus have

$$S_i^F \cong S_i, \quad \tilde{S}_i \cong S_i.$$

Transitivity gives us  $S_i^F \cong \tilde{S}_i$ , and for nonfloating domains  $\Omega_i$ , the statement then follows directly from the general basic result for symmetric and positive definite  $U$  and  $V$ ,

$$U \cong V \iff U^{-1} \cong V^{-1}. \quad (7.16)$$

On floating domains, we recall that the pseudoinverses have the form

$$\tilde{S}_i^\dagger = (\tilde{S}_i + \beta_i R_i R_i^\top)^{-1}, \quad (S_i^F)^\dagger = (S_i^F + \beta_i R_i R_i^\top)^{-1}$$

with some constant  $\beta_i > 0$ . Then clearly

$$\begin{aligned} \langle (\tilde{S}_i + \beta_i R_i R_i^\top) v, v \rangle &= \langle \tilde{S}_i v, v \rangle + \beta_i \langle R_i^\top v, R_i^\top v \rangle, \\ \langle (S_i^F + \beta_i R_i R_i^\top) v, v \rangle &= \langle S_i^F v, v \rangle + \beta_i \langle R_i^\top v, R_i^\top v \rangle, \end{aligned}$$

from which it follows that  $\tilde{S}_i + \beta_i R_i R_i^\top \cong S_i^F + \beta_i R_i R_i^\top$ . The statement follows then again with (7.16).  $\square$

To prove the following condition number estimates, we make use of the standard FETI theory. In particular, we recall that the construction in Section 7.1 is completely analogous to that of a standard FETI method, and we can define the pseudoinverses  $(S_i^F)^\dagger$  analogously to  $\tilde{S}_i^\dagger$ , but based on the FEM Schur complements  $S_i^F$ . Since  $S_i^F$  and  $\tilde{S}_i$  have identical kernels, all algebraic properties remain the same. This gives rise to the block matrix  $(S^F)^\dagger$  and, finally, the standard FETI operator

$$F^F = B(S^F)^\dagger B^\top.$$

The following theorem summarizes known results from the literature on the condition number of the FETI iteration operator  $P^\top F^F$ .

**Assumption 7.2.4.** *We make the following assumptions.*

- *The subdomains  $\Omega_i$  are unions of a few simplices from a shape-regular coarse conforming triangulation.*
- *In the 3D case, there is no subdomain  $\Omega_i$  with a boundary that intersects  $\partial\Omega$  in only one or a few points.*

**Theorem 7.2.5** ([37, 96]). *Under Assumption 7.2.4 and with subdomainwise constant diffusion  $\alpha$ , the classical FETI operator with the choice  $Q = I$  satisfies the condition number estimate*

$$\kappa(P^\top F^F|_{\Lambda_0}) \leq C \frac{\bar{\alpha}}{\underline{\alpha}} \left( \max_{i=1, \dots, N} \frac{H_i}{h_i^F} \right),$$

where  $\bar{\alpha} = \max_{x \in \Omega} \alpha(x)$ ,  $\underline{\alpha} = \min_{x \in \Omega} \alpha(x)$ , and  $C$  depends only on mesh regularity parameters.

**Lemma 7.2.6.** *For subdomain-wise constant diffusion coefficient  $\alpha(x)$  and under Assumption 6.2.2, the operator  $P^\top F$  and its classical FETI analogue  $P^\top F^F$  are spectrally equivalent on the subspace  $\Lambda_0$ , and the equivalence constants depend only on the mesh regularity.*

*Proof.* First note that since  $P$  is a projector onto  $\Lambda_0$ , we have

$$P^\top F \lambda = P^\top F P \lambda \quad \text{and} \quad P^\top F^F \lambda = P^\top F^F P \lambda \quad \forall \lambda \in \Lambda_0.$$

Using the definitions of  $F$  and  $F^F$ , we obtain, for  $\lambda \in \Lambda_0$ ,

$$\begin{aligned} \langle P^\top F P \lambda, \lambda \rangle &= \langle \tilde{S}_i^\dagger B^\top P \lambda, B^\top P \lambda \rangle, \\ \langle P^\top F^F P \lambda, \lambda \rangle &= \langle (S_i^F)^\dagger B^\top P \lambda, B^\top P \lambda \rangle, \end{aligned}$$

and thus the statement follows from Lemma 7.2.3.  $\square$

A condition number estimate of the same quality now follows for our new FETI-type operator for the BEM-based FEM using the spectral equivalence proved before.

**Theorem 7.2.7.** *For subdomain-wise constant diffusion coefficient  $\alpha(x)$ , under Assumptions 6.2.2 and 7.2.4, and with the choice  $Q = I$ , we have the condition number estimate*

$$\kappa(P^\top F|_{\Lambda_0}) \leq C \frac{\bar{\alpha}}{\underline{\alpha}} \left( \max_{i=1, \dots, N} \frac{H_i}{h_i} \right),$$

where  $\bar{\alpha} = \max_{x \in \Omega} \alpha(x)$ ,  $\underline{\alpha} = \min_{x \in \Omega} \alpha(x)$ , and  $C$  depends only on mesh regularity parameters.

*Proof.* The statement follows from Lemma 7.2.6, the standard FETI result given in Theorem 7.2.5 and the bound  $h_i \leq Ch_i^F$  which follows from Assumption 6.2.2.  $\square$

### 7.2.2 The preconditioned case

In a similar fashion, we can transfer the known results on the condition number of the FETI system preconditioned with the Dirichlet preconditioner to our setting.

**Theorem 7.2.8.** *For subdomain-wise constant diffusion coefficient  $\alpha(x)$  and under Assumptions 6.2.2 and 7.2.4, and with the choice  $Q = M_D^{-1}$ , we have the condition number estimate*

$$\kappa(P M_D^{-1} P^\top F|_{\Lambda_0}) \leq C \left( 1 + \log \left( \max_{i=1, \dots, N} \frac{H_i}{h_i} \right) \right)^2,$$

and  $C$  depends only on mesh regularity parameters. In particular,  $C$  does not depend on the values of  $\alpha$ .

*Proof.* The classical FETI analogue to our scaled Dirichlet preconditioner is given by

$$(M_D^F)^{-1} = B_D S^F B_D^\top.$$

The proof for the lower bound of the FETI operator proceeds in a purely algebraic fashion, and we follow the steps in, e.g., the proof of [117, Theorem 6.15], line by line to obtain

$$\langle M_D \lambda, \lambda \rangle \leq \langle F \lambda, \lambda \rangle \quad \forall \lambda \in \Lambda_0.$$

For the upper bound, we start by introducing the projector

$$P_D := B_D^\top B.$$

In [117, 97], it is shown that for any  $y \in Y$ , there exists a unique  $z_y \in \ker S^F = \ker \tilde{S}$  such that

$$B^\top Q B(y + z_y) \perp \ker S^F,$$

and there holds

$$|P_D(y + z_y)|_{S^F}^2 \leq C \left( 1 + \log \left( \max_{i=1, \dots, N} H_i/h_i^F \right) \right)^2 |y|_{S^F}^2. \quad (7.17)$$

Due to the spectral equivalence  $S^F \cong \tilde{S}$  established earlier, the induced norms are equivalent and we obtain

$$|P_D(y + z_y)|_{\tilde{S}}^2 \leq C \left( 1 + \log \left( \max_{i=1, \dots, N} H_i/h_i \right) \right)^2 |y|_{\tilde{S}}^2 \quad (7.18)$$

with another uniform constant  $C$ . Here we also used  $h_i \leq Ch_i^F$ . The estimate (7.17) is the main ingredient in the proof of the upper bound in [117, Theorem 6.15], which apart from this inequality again proceeds purely algebraically. We can thus again repeat the proof line by line, using (7.18) in place of (7.17), in order to obtain

$$\langle F\lambda, \lambda \rangle \leq C \left( 1 + \log \left( \max_{i=1, \dots, N} H_i/h_i \right) \right)^2 \langle M_D\lambda, \lambda \rangle \quad \forall \lambda \in \Lambda_0.$$

The condition number estimate then follows as described in [117, Section 6.3.3].  $\square$

All in all, the results of this chapter show that FETI-style solvers and suitable preconditioners can be easily adapted to the BEM-based FEM, and condition number estimates of the same asymptotic order hold. We will demonstrate the favorable properties of the solver developed here in a numerical example in Section 9.2.3.

# Chapter 8

## Convection-diffusion problems

### 8.1 Stabilized methods: an overview

In this chapter, we turn our attention to convection-diffusion problems of the form

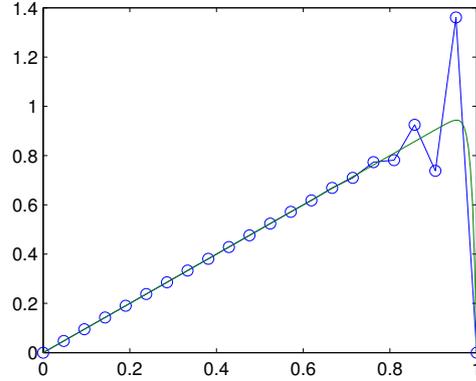
$$\begin{aligned} Lu = -\alpha\Delta u + b \cdot \nabla u &= f && \text{in } \Omega, \\ u &= g && \text{on } \Gamma. \end{aligned}$$

In particular, we are interested in the situation where  $\alpha > 0$  is small, i.e., in convection-dominated problems. Observe that, for the limit case  $\alpha = 0$ , the problem degenerates into a first-order transport problem for which the Dirichlet boundary condition has to be restrained to the inflow boundary, i.e., those points  $x \in \Gamma$  of the boundary where  $b(x) \cdot n(x) < 0$ . The solution of the problem with  $\alpha > 0$  behaves quite differently and converges only in a very weak sense to the solution of the reduced problem as  $\alpha \rightarrow 0$ . The problem is thus referred to as being of singularly perturbed type, and we refer the reader to [104] for details thereon as well as on numerical methods for such types of problems. An important phenomenon in this respect are so-called boundary layers: narrow regions in the vicinity of the outflow boundary (where  $b \cdot n > 0$ ) in which the gradient of the solution is very large. Historically, the theory of boundary layers was first introduced by Ludwig Prandtl in a paper presented at the International Congress of Mathematicians in Heidelberg in 1904.

Standard numerical schemes like the finite element method become unstable when applied to this type of convection-dominated problems. Typically, the issue manifests itself in the form of spurious oscillations near the boundary layers which may spread over the whole computational domain (cf. Figure 8.1). The critical quantity here is the mesh Peclet number,

$$\text{Pe} = \frac{h |b|}{\alpha},$$

which should not be much larger than 1. Here  $h$  is the mesh size and must thus be chosen on the order of  $\alpha/|b|$  in order to obtain a stable approximation of the true solution. This is often infeasible for small diffusion or large convection for reasons of computational effort.



**Figure 8.1:** A convection-dominated 1D problem solved by a standard Courant FEM. Green line: exact solution, blue line: FEM solution.

Due to this phenomenon, the need for stabilized methods has been acknowledged. For 1D finite difference schemes, a well-known technique is the so-called “upwinding”: instead of the central finite difference quotients, one uses asymmetric difference quotients which give precedence to information from the “upwind” direction, i.e., from the left if  $b > 0$  and vice versa. This takes into account the natural transport of information due to the convection  $b$  and allows the construction of stable methods.

It is therefore natural to look for generalizations of this idea to higher dimensions. One of the most successful approaches has been the Streamline Upwind/Petrov-Galerkin (SUPG) method proposed by Brooks and Hughes in 1982 [20]. The first rigorous analysis was given by Johnson, Nävert, and Pitkäranta soon after [72]. This approach is now understood to fall into a larger class of stabilized methods where the standard variational equation (2.4) is augmented by a stabilizing term,

$$\mathcal{L}(u_h, v_h) + \langle R(u_h), W(v_h) \rangle = \langle F, v_h \rangle \quad \forall v_h \in \mathcal{V}_{h,0}, \quad (8.1)$$

where  $R(u_h) = -\alpha \Delta u_h + b \cdot \nabla u_h - f$  is the residual of the partial differential equation, to be understood elementwise such that all terms are well-defined, and  $W(v_h)$  is a suitably chosen weighting operator. Note that the use of the residual in the stabilizing term ensures consistency of the resulting method, i.e., the variational formulation remains valid for the exact solution, which is a major advantage over previously developed upwind schemes. The SUPG method is obtained with the streamline upwind weighting operator

$$W_{\text{SUPG}}(v_h) = \tau(b \cdot \nabla v_h),$$

while the choice

$$W_{\text{GLS}}(v_h) = \tau(-\alpha \Delta v_h + b \cdot \nabla v_h)$$

results in the Galerkin/least-squares method [69] and the choice

$$W_{\text{USFEM}}(v_h) = \tau(+\alpha\Delta v_h + b \cdot \nabla v_h)$$

yields a so-called unusual stabilized finite element method [41]. In all cases,  $\tau > 0$  is a stabilization parameter which is chosen per element,  $\tau(x) \equiv \tau_T$  on  $T$ , and all of these weighting operators are again to be read elementwise. Originally, in [20],  $\tau$  was chosen by comparison with finite difference stencils and was limited to the piecewise linear case. More sophisticated choices have been proposed later (see [41] and the references therein). For an excellent overview of the field of stabilized finite element methods from a modern perspective, see [43] and the references therein.

Another approach for constructing stable numerical methods are element bubble functions, or simply bubbles, originally suggested by Brezzi et al. in 1992 [13]. Here the trial and test spaces are enriched with bubble functions which may be taken from different spaces, but always share the characteristic that they vanish on element boundaries and are thus independent from element to element. This allows static condensation of the bubble components such that in the end the number of degrees of freedom of the global linear system is not increased.

Already in [13], it was shown that the SUPG method in 2D for piecewise linear trial functions can be obtained by a particular choice of bubbles. These results were extended in [5] to show that also stabilized methods of higher polynomial degree and the Galerkin/least-squares method can be obtained by proper choices of bubble functions. Interestingly, in this way, a natural choice for the stabilization parameter  $\tau$  in the stabilized FEM suggests itself.

A particular choice of bubbles are the residual-free bubbles, where the bubbles are chosen as local solutions of the partial differential equation. This is of course close in spirit to the way we have introduced the BEM-based FEM, and we will therefore outline the construction of the residual-free bubbles method later in this chapter in order to show that the underlying constructions are equivalent. Residual-free bubbles have been shown to lead to a stable, quasi-optimal numerical method for convection-dominated problems by Brezzi, Hughes, Marini, Russo, and Süli [16] in the 2D case with piecewise linear trial functions. In fact, in this setting the method is equivalent to the SUPG method with a particular choice of  $\tau$ . The analysis was later also performed for trial functions of higher polynomial degree [17]. We also refer to [11] for a discussion of bubbles in the context of subgrid scales.

We finally mention the variational multiscale method, which was introduced by Hughes in 1995 [68] as another approach for deriving stable methods; see also [70]. The idea here is to split the solution into coarse or resolvable scales and fine or unresolvable scales. One then attempts to determine the effect of the fine scales on the coarse ones analytically such that the unresolvable scales could be eliminated from the coarse problem. Element Green's functions are fundamental in computing these fine-scale effects on the coarse

scales. It was shown by Brezzi, Franca, Hughes, and Russo in 1997 [14] that the variational multiscale method and the residual-free bubbles method are essentially equivalent under the assumption that the unresolvable scales do not cross element boundaries.

## 8.2 Residual-free bubbles and their relation to the BEM-based FEM

In the following, we assume that the mesh  $\mathcal{T}$  of  $\Omega$  is simplicial with standard piecewise linear FE spaces  $\mathcal{V}_h$  and  $\mathcal{V}_{h,0}$  as introduced in Section 2.5. We start from the standard homogenized FEM formulation: find  $u_h \in \mathcal{V}_{h,0}$  such that

$$\mathcal{L}(u_h, v_h) = \langle F, v_h \rangle \quad \forall v_h \in \mathcal{V}_{h,0}.$$

The basic approach of any bubble method is to enrich the trial space by the eponymous “bubbles,” defined locally on each element  $T$ ,

$$\mathcal{V}_E := \mathcal{V}_{h,0} \oplus \mathcal{B}_h, \quad \text{where} \quad \mathcal{B}_h := \bigotimes_{T \in \mathcal{T}} \mathcal{B}_h(T)$$

and  $\mathcal{B}_h(T) \subseteq H_0^1(T)$  is a closed subspace of the  $H_0^1$ -functions on  $T$ . The sum in this construction is indeed direct since the only function in  $\mathcal{V}_{h,0}$  which is 0 on the element boundaries is the zero function.

We now have the enriched variational formulation: find  $u_E \in \mathcal{V}_E$  such that

$$\mathcal{L}(u_E, v_E) = \langle F, v_E \rangle \quad \forall v_E \in \mathcal{V}_E. \quad (8.2)$$

Due to the direct sum in the definition of  $\mathcal{V}_E$ , the functions  $u_E, v_E \in \mathcal{V}_E$  have the unique splittings

$$u_E = u_h + u_B, \quad v_E = v_h + v_B$$

with the finite element components  $u_h, v_h \in \mathcal{V}_h$  and the bubble components  $u_B, v_B \in \mathcal{B}_h$ . In particular, testing only with bubbles,  $v_E = v_B \in \mathcal{B}_h$ , in (8.2), we obtain the so-called *bubble equation*

$$\mathcal{L}(u_B, v_B) = -\mathcal{L}(u_h, v_B) + \langle F, v_B \rangle \quad \forall v_B \in \mathcal{B}_h. \quad (8.3)$$

Since functions in the bubble space  $\mathcal{B}_h$  are composed of element-local bubbles  $\mathcal{B}_h(T)$  which are independent of each other, it is clear that the bubble equation (8.3) can be solved locally on each element. Indeed, writing  $u_{B,T} = u_B|_T$  for the bubble component of  $u_E$  on  $T \in \mathcal{T}$ , equation (8.3) determines  $u_{B,T}$  uniquely in dependence of  $u_{h,T} = u_h|_T$ .

One possible choice for  $\mathcal{B}_h(T)$  are suitable polynomial bubbles. The residual-free bubbles method, on the other hand, is obtained by letting  $\mathcal{B}_h(T) = H_0^1(T)$ , and we will stick to this choice in the following. In this setting, (8.3) is nothing but the weak form of

$$\begin{aligned} Lu_{B,T} &= f - Lu_{h,T} && \text{in } T, \\ u_{B,T} &= 0 && \text{on } \partial T, \end{aligned} \quad (8.4)$$

and we will formally write  $u_{B,T} = L_{B,T}^{-1}(f - Lu_h)$  for the solution of the local bubble equation. Thus,  $u_B = \sum_{T \in \mathcal{T}} u_{B,T}$ . Inserting this into the enriched formulation (8.2) and using (8.3), we get the statically condensed variational formulation

$$\mathcal{L}(u_h, v_h) + \sum_{T \in \mathcal{T}} \mathcal{L}(L_{B,T}^{-1}(f - Lu_h), v_h) = \langle F, v_h \rangle \quad \forall v_h \in \mathcal{V}_h \quad (8.5)$$

where the local bubble unknowns  $u_{B,T}$  have been eliminated. We thus see that the process of enriching the trial space can equivalently be viewed as a local modification of the bilinear form  $\mathcal{L}(\cdot, \cdot)$  (although we point out that this modification is not bilinear for  $f \neq 0$ ). Indeed, (8.5) is nothing but the standard variational formulation augmented by the consistent stabilizing term  $\mathcal{L}(u_B, v_h)$ .

In the case where  $f$ ,  $\alpha$ , and  $b$  are constant on each element  $T$ , it is easy to see that each bubble  $u_{B,T}$  is a scalar multiple of the element bubble function  $\varphi_T \in H_0^1(T)$  given as the solution of

$$\begin{aligned} L\varphi_T &= 1 && \text{in } T, \\ \varphi_T &= 0 && \text{on } \partial T. \end{aligned}$$

To be precise, we have  $u_{B,T} = (f_T - b \cdot \nabla u_{h,T})\varphi_T$ . By an integration by parts argument (cf. [13, 11]), one easily shows that in this case the stabilizing term in (8.5) on each  $T$  has the form

$$\mathcal{L}_T(u_{B,T}, v_{h,T}) = \frac{\int_T \varphi_T}{|T|} \int_T (b \cdot \nabla u_{h,T} - f_T) b \cdot \nabla v_{h,T} dx.$$

Referring back to (8.1) and the discussion thereafter, we see that this is precisely the stabilizing term occurring in the SUPG method with a stabilization parameter  $\tau_T = \frac{\int_T \varphi_T}{|T|}$ . With this choice, the SUPG method and the residual-free bubbles method are thus equivalent in this setting.

If one wishes to realize a method of this type numerically, one could thus set up the standard FE stiffness matrix corresponding to  $\mathcal{L}(u_h, v_h)$ , and, in a second step, solve the local bubble problems and add the bubble contributions to the stiffness matrix. Here one encounters the problem that the bubble problems are impossible to solve exactly since they are in general not simpler than the original boundary value problem. Suitable approximations thus have to be made. One possibility is to introduce fine local finite element meshes on all elements and solve the bubble equations (8.4) locally using a finite element discretization in order to obtain approximations to  $u_{B,T}$ . We refer to [42] for a variation of this approach where a Galerkin/least-squares method is used to approximate the element bubbles. Another approach that has been applied successfully to convection-dominated problems is to set the diffusion coefficient  $\alpha$  to zero in (8.4) and use the analytic solution of the resulting first-order transport problem as an approximation

to  $u_{B,T}$ ; cf. [12]. Finally, the bubble space can be restricted to be finite-dimensional, for instance to polynomials up to a certain degree or to hat functions with a single suitably chosen vertex within  $T$ , and the Galerkin projection of the solution of (8.4) onto this space can be used as an approximation to  $u_{B,T}$ . See [15] for an example.

Let us return to the bubble equation and consider it in terms of the full unknown  $u_E = u_h + u_B$  instead. If we restrict it to a single element  $T \in \mathcal{T}$ , we see that  $u_{E,T}$  satisfies

$$\mathcal{L}_T(u_{E,T}, v_B) = \langle F, v_B \rangle \quad \forall v_B \in \mathcal{B}_h(T),$$

and due to the choice  $\mathcal{B}_h(T) = H_0^1(T)$ , the function  $u_E$  satisfies the PDE weakly on every  $T$ . The corresponding strong formulation reads

$$\begin{aligned} Lu_{E,T} &= f && \text{in } T, \\ u_{E,T} &= u_h && \text{on } \partial T. \end{aligned}$$

We thus see that the original FE space  $\mathcal{V}_h$  provides only values on the element boundaries, while its values within the elements are inconsequential as there the trial function  $u_E$  is the solution of a local boundary value problem. Let us mention that this means that we can generalize the mesh  $\mathcal{T}$  from only simplices to the polytopal meshes treated in the previous chapters of this work, as now only values on the skeleton appear in the variational formulation.

In the case that the partial differential equation has a homogeneous right-hand side, the local trial functions  $u_{E,T}$  are nothing but the  $\mathcal{L}$ -harmonic extensions that we used in Chapter 4 to derive the BEM-based FEM. Recalling Definition 2.4.2, for such functions we have

$$\mathcal{L}_T(u_E, v_E) = \langle \gamma_T^1 u_E, \gamma_T^0 v_E \rangle_{\partial T} = \langle S_T u_h, v_h \rangle_{\partial T}$$

and thus we can write the enriched variational formulation (8.2) in this setting as

$$\sum_{T \in \mathcal{T}} \langle S_T u_{h,T}, v_{h,T} \rangle = \langle F, v_h \rangle \quad \forall v_h \in \mathcal{V}_h. \quad (8.6)$$

We see from this that the bubble-enriched variational formulation is equivalent to the skeletal variational formulation derived in Chapter 4 for discrete skeletal spaces, but with exact realization of the Steklov-Poincaré operator  $S_T$ . As noted before, this is in general impossible to realize numerically. In the BEM-based FEM, we take the approach of replacing the Steklov-Poincaré operator with its BEM approximation  $\tilde{S}_T$ , in contrast to the typical bubble methods in the literature where approximations to  $u_{B,T}$  are made as discussed above.

In the literature, one finds the following error estimate for the residual-free bubbles method.

**Theorem 8.2.1** (Brezzi et al. [16, 17]). *Assume that  $\mathcal{T}$  is a shape-regular triangular mesh of  $\Omega \subset \mathbb{R}^2$ , that  $\alpha$  and  $b$  are constant, and that  $f$  is piecewise constant with respect*

to  $\mathcal{T}$ . Let  $u \in H^s(\Omega) \cap H_0^1(\Omega)$ ,  $1 < s \leq 2$ , be the exact solution of the convection-diffusion problem and  $u_E = u_h + u_B \in \mathcal{V}_E$  the solution of the residual-free bubble equation (8.2). Then we have the error estimates

$$\alpha^{1/2} |u - u_E|_{H^1(\Omega)} \leq C \left( \sum_{T \in \mathcal{T}} \gamma_T h_T^{2s-1} |u|_{H^s(T)}^2 \right)^{1/2},$$

and

$$\left( \alpha |u - u_h|_{H^1(\Omega)}^2 + \sum_{T \in \mathcal{T}} \tilde{h}_T \|b \cdot \nabla(u - u_h)\|_{L_2(T)}^2 \right)^{1/2} \leq C \left( \sum_{T \in \mathcal{T}} \gamma_T h_T^{2s-1} |u|_{H^s(T)}^2 \right)^{1/2},$$

where

$$\gamma_T = |b| \max \left\{ 1, \frac{\alpha}{h_T |b|} \right\}, \quad \tilde{h}_T = \left( \int_T \varphi_T \right) / |T|,$$

$\varphi_T \in H_0^1(T)$  is the element bubble as introduced above, and  $C$  is a uniform constant depending only on mesh regularity.

We point out that the norm  $\alpha^{1/2} |u|_{H^1(\Omega)}$  of the exact solution stays uniformly bounded independently of  $\alpha$ , while the same is not true for  $|u|_{H^1(\Omega)}$ , which may blow up for  $\alpha \rightarrow 0$  if boundary layers are present. The norm appearing on the left-hand side of the second estimate above is very similar to the so-called stability norm appearing in the analysis of the SUPG method. In particular, it guarantees that the solution does not develop excessive oscillations in the streamline direction. We mention however that the residual-free bubbles method, as the SUPG method, cannot completely eliminate oscillations near boundary layers. A method which satisfies a discrete maximum principle and is thus free of spurious oscillations has been proposed by Mizukami and Hughes in 1985 [88], but has the drawback that it is a nonlinear scheme even for linear problems with constant coefficients.

Since we have seen above that the skeletal variational formulation (8.6) is equivalent to the exact formulation of the residual-free bubbles method, we expect that the BEM-based FEM similarly results in a stable method if the approximation  $\tilde{S}_T$  is “good enough” in a certain sense. At present, we have no rigorous proof for a quantitative statement of this type. We have however performed some numerical experiments which substantiate this claim and will present them in Section 9.2.2.

One idea which might enhance the stabilizing properties of the method is to refine the coarse element boundary meshes for a better approximation of the Neumann data. We describe this approach in the following.

Remember that the skeletal triangulation, when restricted to the boundary of a single element  $T$ , gave rise to a boundary triangulation of  $\partial T$  on which we defined the space of piecewise constant boundary functions  $\mathcal{Z}_h(T)$ ; cf. Section 3.3. If we uniformly refine these boundary meshes  $m$  times, we obtain a sequence of nested boundary element spaces

$$\mathcal{Z}_h(T) = \mathcal{Z}_h^{(0)}(T) \subset \mathcal{Z}_h^{(1)}(T) \subset \dots \subset \mathcal{Z}_h^{(m)}(T).$$

Let  $i \in \{0, \dots, m\}$ . As in Section 3.3, we then define, for given Dirichlet data  $u \in H^{1/2}(\partial T)$ , the Galerkin projections of the Neumann data,  $t_{h,T}^{(i)}(u) \in \mathcal{Z}_h^{(i)}(T)$  by the variational problem

$$\langle z_h, V_T t_{h,T}^{(i)}(u) \rangle_{\partial T} = \langle z_h, (\frac{1}{2}I + K_T)u \rangle_{\partial T} \quad \forall z_h \in \mathcal{Z}_h^{(i)}(T).$$

This gives rise to a sequence of improved approximations to  $S_T$  by

$$\begin{aligned} \tilde{S}_T^{(i)} : H^{1/2}(\Gamma) &\rightarrow H^{-1/2}(\Gamma), \\ u &\mapsto D_T u + (\frac{1}{2}I + K'_T) t_{h,T}^{(i)}(u), \end{aligned}$$

and finally the improved discrete skeletal problems

$$\sum_{T \in \mathcal{T}} \langle \tilde{S}_T^{(i)} u_{h,T}, v_{h,T} \rangle = \langle F, v_h \rangle \quad \forall v_h \in \mathcal{W}_h$$

for  $i \in \{0, \dots, m\}$ . We see that the global number of degrees of freedom is not affected by this procedure. Rather, only the solution of the local problems is improved by refining the element boundary meshes. In fact, it is possible to choose different refinement levels  $i_T \in \mathbb{N}_0$  per element  $T \in \mathcal{T}$  depending on the magnitude of the local mesh Peclet number,  $\text{Pe}_T = h_T |b_T| / \alpha_T$ .

We will demonstrate the stabilizing properties of the BEM-based FEM for convection-diffusion problems in a numerical example in Section 9.2.2.

# Chapter 9

## Implementation and numerical examples

### 9.1 Implementation details

While the focus of this thesis lies on the theoretical analysis of the BEM-based FEM, effort has also gone into implementing the BEM-based FEM and performing numerical tests in order to verify the theoretical results, as well as carrying out numerical studies for cases where a complete analysis is still missing, such as convection-diffusion problems. The implementation was done in C++ and builds upon the `ParMax` framework<sup>1</sup> which was initially developed by Clemens Pechstein and Dylan Copeland at the Institute of Computational Mathematics, Linz, and since used and extended by several other collaborators. This software package is based on object-oriented and generic programming paradigms and includes a host of useful classes and algorithms for numerical programming; for instance,

- a flexible mesh class with a native mesh file format as well as support for importing third party mesh file formats, for instance the `NetGen` format,
- classes for statically and dynamically memory managed vectors,
- classes for fixed-size and variable-sized matrices of dense and sparse type,
- expression templates for many matrix and vector operations,
- bindings to `BLAS`, `LAPACK`, `PARDISO` [108], `UMFPACK` [28], `METIS`, and other established numerical libraries,
- export facilities to the `VTK` and `GMV` visualization toolkits,
- a powerful and flexible implementation of the finite element method for various element types,
- an implementation of a finite element tearing/interconnecting (FETI) solver supporting various preconditioners and local solvers,

---

<sup>1</sup><http://www.numa.uni-linz.ac.at/P19255/software.shtml>

- implementations of iterative solutions methods, e.g., CG [57], MinRes [92], and GMRes [105],

and many more. During the implementation of the BEM-based FEM, the author has contributed several features to the `ParMax` framework himself, extended functionality of several existing classes, and discovered and fixed bugs in some components, thus making useful contributions also for future users of the software package.

An important and non-trivial issue of any scheme making use of boundary element methods is the question of how to compute the occurring boundary integrals. Typically, for the Galerkin-type discretization that we use (as opposed to collocation methods), computing one entry of a BEM matrix requires integration over a pair of triangles, thus leading to a four-dimensional integration problem. Due to the singularity of these integrals, contrary to finite element methods, standard Gauss quadrature formulae fail when applied to this problem. For our implementation of the BEM-based FEM, we have taken two different approaches to this problem. First, for Laplace and diffusion problems, we use the approach of the `OSTBEM` library [111] due to Olaf Steinbach: here, the inner (collocation) integral is computed analytically, while the outer integral is approximated by a Gauss-type quadrature rule for triangles using, e.g., 4 or 7 quadrature points. For more general type of equations incorporating convection or reaction terms as well as matrix-valued diffusion coefficients, no analytic formula for the inner integral is available to the best of our knowledge. Therefore, in these cases, we employ the fully numerical integration scheme described by Sauter and Schwab [106]. Here, the four-dimensional integral is transformed in a well-chosen way to the four-dimensional hypercube such that the resulting integrand is analytic within this hypercube. To this problem, four-dimensional quadrature rules are then applied, for instance tensor products of one-dimensional Gauss rules or rules specifically designed for the hypercube. An encyclopedia of suitable quadrature rules can be found in [23]. We note that we could not obtain an existing implementation of this approach for the case of convection-diffusion equations, and thus the development of BEM quadrature routines for 3D convection-diffusion operators, which might be of more general interest, is also among the results of the thesis.

For the solution of the resulting linear system, we use the Conjugate Gradient (CG) method for symmetric problems and the GMRes method for nonsymmetric problems.

## 9.2 Numerical experiments

### 9.2.1 Example 1: The Laplace equation

In the first example, we consider the Dirichlet boundary value problem for the Laplace equation in the unit cube  $\Omega = (0, 1)^3$ . We prescribe the exact solution

$$u(x, y, z) = \exp(x) \cos(y)(1 + z).$$

We perform computations on two different mesh configurations. The first one is a standard regular tetrahedral mesh obtained by uniform refinement of a coarse mesh. The second one is derived from the first one by unifying some pairs of adjacent tetrahedra. This results in meshes consisting of both tetrahedra and polyhedra having 5 vertices, 9 edges and 6 faces. Some of the latter may be non-convex. Because our method places its degrees of freedom in element vertices, this unification procedure does not change the number of unknowns.

For computing the  $L_2$ -error, we use the representation formula (Theorem 3.2.1) to evaluate the solution at some inner points of the elements and perform quadrature. For computing the  $H^1$ -error, we estimate the gradient as a piecewise constant quantity from the computed Neumann data and again perform quadrature.

The results are shown in Table 9.1, where Table 9.1(a) gives the results for the tetrahedral meshes, while Table 9.1(b) gives the results for the mixed meshes. In each table, the first column gives the mesh size (here calculated as the maximum edge length). The second and third columns give the error in the  $H^1$ -seminorm and the  $L_2$ -norm, respectively. The final columns give the number of tetrahedra and polyhedra in each mesh.

**Table 9.1:** Numerical results

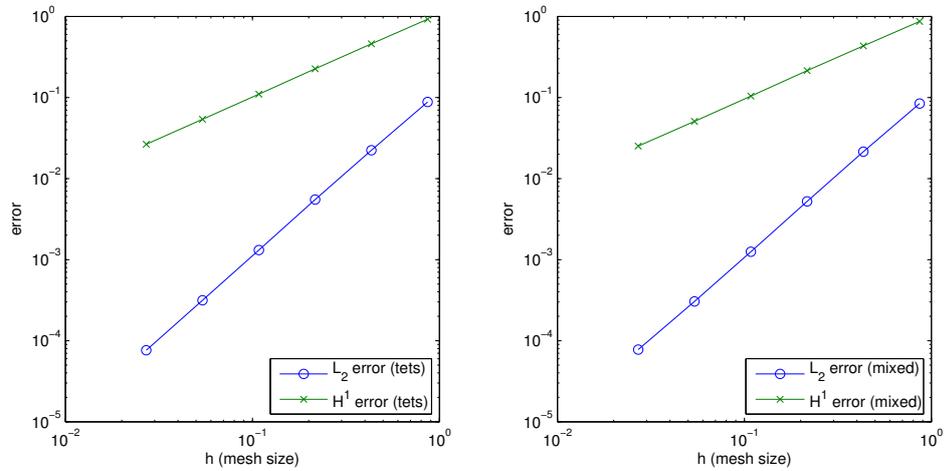
mesh size $h$	$H^1$ -error	$L^2$ -error	#tets
0.866025	0.923507	0.0879679	48
0.433013	0.459565	0.0223147	384
0.216506	0.226186	0.00549834	3,072
0.108253	0.109806	0.00131165	24,576
0.0541266	0.0537825	0.000315016	196,608
0.0270633	0.0264988	7.62441e-05	1,572,864

(a) Results with tetrahedral mesh.

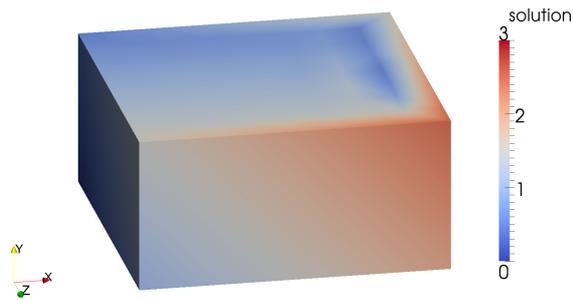
mesh size $h$	$H^1$ -error	$L^2$ -error	#tets	#polys
0.866025	0.867685	0.0842554	40	4
0.433013	0.433557	0.0214242	258	63
0.216506	0.214188	0.00522372	2,044	514
0.108253	0.103955	0.00124863	15,822	4,377
0.0541266	0.0508436	0.000304395	125,350	35,629
0.0270633	0.0251327	7.76704e-05	996,390	288,237

(b) Results with mixed mesh.

In both cases, the  $H^1$ -error decays with  $\mathcal{O}(h)$ , as Theorem 6.5.1 predicts. Also, the  $L_2$ -error decays with  $\mathcal{O}(h^2)$  in both experiments. Figure 9.1 visualizes these results graphically. As can be seen, the errors for the tetrahedral and mixed meshes are virtually indistinguishable.



**Figure 9.1:**  $L_2$ - and  $H^1$ -error for tetrahedral and mixed mesh.



**Figure 9.2:** Solution of the convection-diffusion model problem in Example 2 with  $\alpha = 0.01$  by the BEM-based FEM. Displayed is a cross section through the unit cube at  $y = 0.5$ .

### 9.2.2 Example 2: A convection-diffusion problem

In the unit cube,  $\Omega = (0, 1)^3$ , we consider the boundary value problem

$$\begin{aligned} -\alpha \Delta u + (1, 0, 0)^\top \cdot \nabla u &= 0 && \text{in } \Omega, \\ u(x, y, z) &= x + y + z && \text{on } \partial\Omega \end{aligned}$$

for the convection-diffusion equation with constant diffusion coefficient  $\alpha > 0$  and constant convection  $b = (1, 0, 0)$ . For small  $\alpha$ , boundary layers form which are challenging to reproduce in a stable way, as described in Chapter 8. We discretize  $\Omega$  by a relatively

coarse tetrahedral mesh with 729 vertices and 3072 tetrahedra and compare a standard Courant finite element method, as described in Section 2.5, and the BEM-based FEM, for varying values of  $\alpha$ . In Figure 9.2, we plot the solution using the BEM-based FEM for the case  $\alpha = 0.01$ . Note that, especially for small  $\alpha$ , the chosen mesh is too coarse to accurately resolve the boundary layer near  $x = 1$ . Error norms are thus inappropriate to assess the quality of the numerical solution. Instead, we are only interested in the stability behavior of the numerical schemes in this example. Therefore, we check if a discrete maximum principle is satisfied for the numerical solutions. Observe that the given boundary data has values in the interval  $[0, 3]$ , and thus, by the maximum principle, the same holds true for the exact solution  $u$ . In Table 9.2, we display the range of values taken by the solutions produced by the FEM and the BEM-based FEM for various choices of  $\alpha$ . The resulting linear system was solved using GMRes without restart using a simple geometric row scaling (GRS) preconditioner (see [52]), i.e., a diagonal preconditioner

$$C^{-1} = \text{diag}(1/\|\underline{K}_1\|_p, \dots, 1/\|\underline{K}_n\|_p),$$

where by  $\underline{K}_j$  we mean the  $j$ -th row of the stiffness matrix of either the FEM or the BEM-based FEM discretization, and we choose the vector norm with  $p = 1$  in our example. We also give the iteration numbers taken by the GMRes iteration in order to reduce the norm of the initial residual by a factor of  $10^{-6}$  for both schemes in the table.

$\alpha$	BBF range	FEM range	BBF iter.	FEM iter.
0.1	<b>0–3</b>	<b>0–3</b>	25	26
0.05	<b>0–3</b>	<b>0–3</b>	24	26
0.025	<b>0–3</b>	<b>0–3</b>	25	32
0.01	<b>0–3</b>	–0.553574–3	28	47
0.005	<b>0–3</b>	–1.13544–3	32	67
0.0025	<b>0–3</b>	–1.85352–3.07495	34	100
0.001	<b>0–3</b>	–3.7772–4.08492	36	178
0.0005	<b>0–3</b>	–4.90292–5.7114	41	261
0.00025	–173.73–114.837	–5.73103–6.89838	355	351
0.0001	–40.1773–28.3901	–9.1113–12.551	196	379

**Table 9.2:** Results of the BEM-based FEM and a Courant FEM for the convection-diffusion problem in Example 2 for different  $\alpha$ . Columns, from left to right: diffusion coefficient  $\alpha$ , range of values taken by the BEM-based FEM solution and the FEM solution, number of GMRes iterations in the BEM-based FEM and in the FEM.

The cases where the discrete maximum principle is fulfilled, and thus where we have stability of the discrete solution, are marked in bold in Table 9.2. We see that this is the case for much smaller values of  $\alpha$  for the BEM-based FEM compared to the

FEM. However, while the oscillations in the FEM grow in a predictable manner past  $\alpha = 0.025$ , the solution produced by the BEM-based FEM becomes dramatically worse from  $\alpha = 0.0005$  to  $\alpha = 0.00025$ . Refinements of the element boundary meshes in order to improve the approximation of the Neumann data as described at the end of Chapter 8 do not noticeably mitigate the oscillations in this case. It is clear that further analysis is needed in order to explain these phenomena. We conjecture that the quadrature rules of fully numeric type [106] which we use to compute the entries of the element BEM matrices become unstable for convection-dominated problems. Indeed, their analysis is, to the best of our knowledge, only known for potential equations, and further work is needed on this topic.

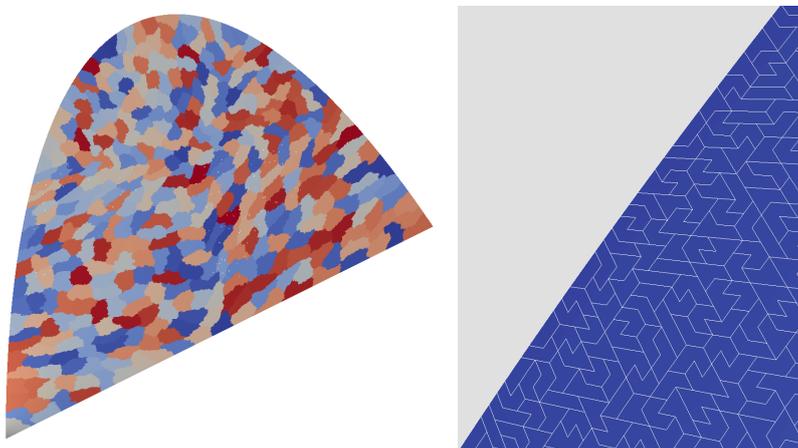
We also point out that the number of GMRes iterations is smaller in almost all cases for the BEM-based FEM than it is for the FEM.

### 9.2.3 Example 3: Domain decomposition

We solve a problem by the use of the FETI-type domain decomposition procedure for the BEM-based FEM presented in Chapter 7. As a model problem, we solve the Laplace equation with pure Dirichlet boundary conditions

$$-\Delta u = 0 \quad \text{in } \Omega, \quad u(x) = -\frac{1}{2\pi} \log |x - x^*| \quad \text{on } \partial\Omega,$$

on a two-dimensional domain  $\Omega$  (Figure 9.3, left) which is discretized by an irregular polygonal mesh. The source point  $x^* = (-1, 1)^\top$  lies outside of  $\Omega$ .



**Figure 9.3:** *Left:* Computational domain with a sample partitioning of  $N = 400$  subdomains. *Right:* Detail of polygonal mesh at domain boundary.

The polygonal mesh is constructed by first setting up a standard triangular mesh of the computational domain and then applying the graph partitioning software METIS [74] to this base mesh with a very high choice for the desired number of domains. This results in METIS creating partitions which contain only few triangles per subdomain. In our example, the base mesh had 524,288 triangles, and after applying the step described above, we obtain a mesh with 99,970 polygonal elements, most of which are unions of 5 or 6 triangles, and approximately 263,000 vertices, corresponding to the number of degrees of freedom. A few of these elements are shown in the closeup in Figure 9.3, right. This is the mesh  $\mathcal{T}$  on which the BEM-based FEM is set up.

Next, we decompose the domain  $\Omega$  into  $N$  subdomains  $\Omega_i$ , each triangulated by subsets  $(\mathcal{T}_i)_{i=1}^N$  of the polygonal mesh. These subsets are constructed by applying METIS a second time on top of the mesh constructed above. The result of this step is shown in Figure 9.3, left, for the case of  $N = 400$  subdomains. On this partitioning, we set up the FETI-type solver described in Chapter 7. We use the Dirichlet preconditioner described therein and use multiplicity scaling as well as a suitable diagonal matrix for  $Q$  as described in [76] for ease of implementation. We have to solve the preconditioned equation

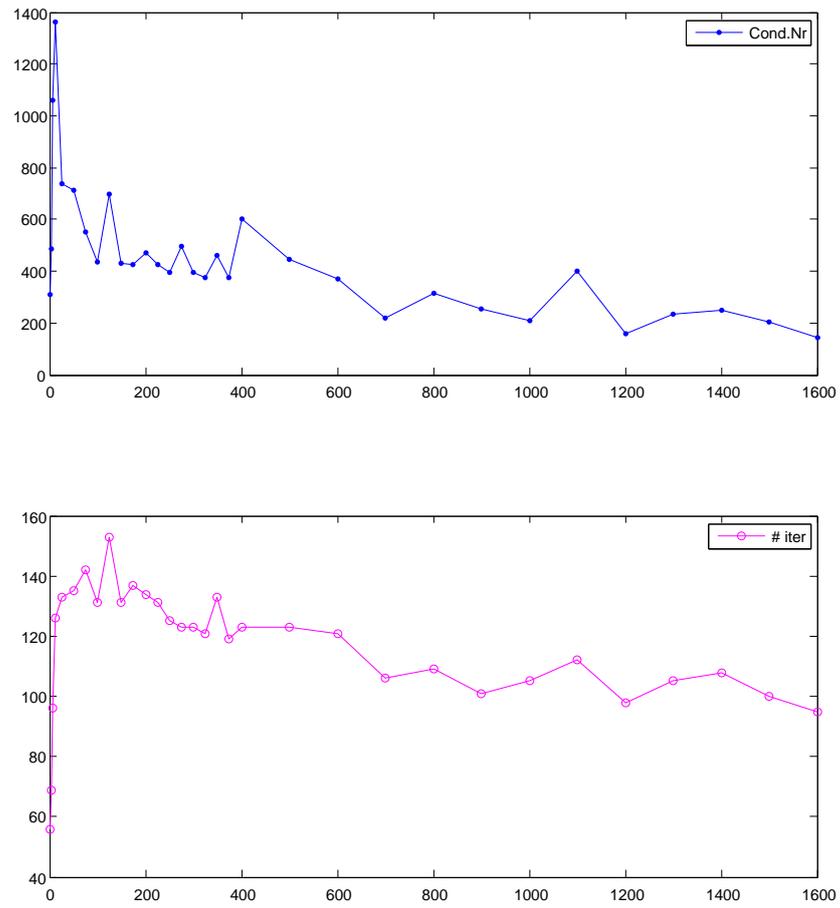
$$PM_D^{-1}P^\top F\lambda_0 = PM_D^{-1}\tilde{g}$$

for  $\lambda_0 \in \Lambda_0$ , which we do by preconditioned Conjugate Gradient (PCG) iteration. In the following, we give the number of PCG iterations required to achieve reduction of the initial residual by a factor of  $10^{-8}$  for varying number  $N$  of subdomains. We also compare these results to the non-preconditioned equation

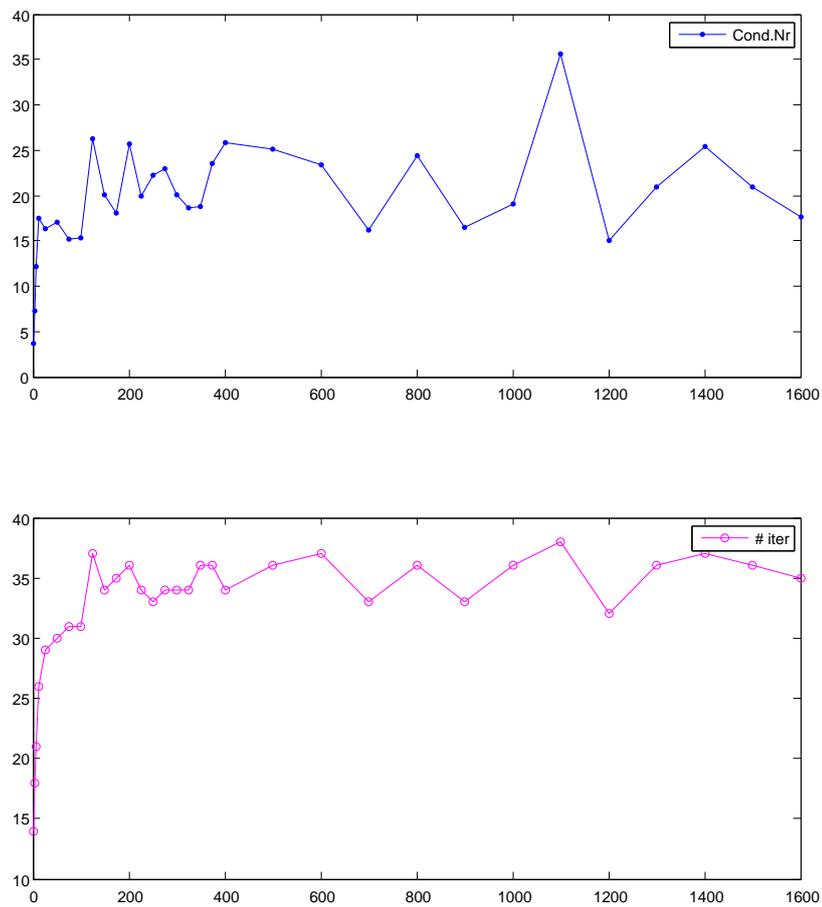
$$P^\top F\lambda_0 = \tilde{g}$$

solved by standard CG iteration. The estimated condition numbers and iteration numbers for the non-preconditioned and the preconditioned case, respectively, are shown in Figures 9.4 and 9.5. Additional data on these two cases can be found in Tables 9.3 and 9.3, respectively.

First, we point out that the jagged nature of the plots in Figures 9.4 and 9.5 is due to the domain decompositions created by METIS for varying  $N$  not being nested. The non-preconditioned case, Figure 9.4, shows a decay in the condition number which roughly correlates to the theoretical estimate from Theorem 7.2.7,  $\kappa = \mathcal{O}(H/h) = \mathcal{O}(N^{-1/2})$ . The condition numbers for the preconditioned case in Figure 9.5 show no clear tendency, which may be due to the problem size being too small. Most importantly, they stay uniformly bounded, and therefore so do the iteration numbers. We have compared the condition and iteration numbers to an analogous FETI method for a Courant FEM on the underlying triangular mesh, and the numbers are comparable, indicating that the behavior of the condition number is not particular to the BEM-based FEM.



**Figure 9.4:** Non-preconditioned FETI-type solver for the BEM-based FEM: estimated condition numbers and CG iteration numbers in dependence of  $N$



**Figure 9.5:** Preconditioned FETI-type solver for the BEM-based FEM, Dirichlet preconditioner: estimated condition numbers and PCG iteration numbers in dependence of  $N$

$N$	total time	avg. loc. time	#iter	error	# Lagrange
2	22.9563	2.70074	14	6.45469e-06	709
3	23.2247	1.51811	18	6.54676e-06	1143
6	21.7471	0.54795	21	6.39847e-06	2168
12	22.1529	0.213242	26	6.3791e-06	3592
25	20.4929	0.0759496	29	4.77774e-06	5875
50	19.1048	0.0310428	30	5.90121e-06	8962
75	18.4645	0.0184307	31	4.81877e-06	11148
100	17.7017	0.013063	31	4.75217e-06	13012
125	18.9504	0.00994315	37	4.72053e-06	14674
150	17.5015	0.00799038	34	4.72264e-06	16219
175	17.4271	0.00671432	35	5.39282e-06	17715
200	17.4147	0.00573496	36	5.14293e-06	19056
225	16.5907	0.00498166	34	5.14425e-06	20082
250	16.0321	0.00440914	33	4.75428e-06	21372
275	16.3442	0.00400218	34	5.13166e-06	22440
300	16.1877	0.00360337	34	4.73872e-06	23460
325	16.1579	0.00329481	34	4.75269e-06	24544
350	16.6816	0.00303931	36	4.736e-06	25459
375	16.5968	0.00281004	36	4.74014e-06	26248
400	16.1319	0.00263458	34	4.73785e-06	27324
500	16.7581	0.00205011	36	4.72341e-06	30703
600	17.2983	0.00169811	37	4.76958e-06	33808
700	16.5087	0.00143421	33	4.76179e-06	36822
800	17.6753	0.00125147	36	4.77933e-06	39304
900	17.2409	0.00110415	33	5.19262e-06	41693
1000	18.3749	0.000990824	36	4.76959e-06	44467
1100	19.4621	0.000898355	38	4.78112e-06	46611
1200	18.1828	0.000819669	32	4.82913e-06	48813
1300	19.5805	0.000753359	36	4.84528e-06	50727
1400	20.4094	0.000699564	37	5.21988e-06	53107
1500	20.6657	0.000650639	36	4.81754e-06	54925
1600	20.9631	0.000613087	35	4.813e-06	56632

**Table 9.3:** Results of the preconditioned solver. Columns: number of subdomains, total CPU time for solution, averaged time for solution of local problems, number of iterations, residual error, number of Lagrange multipliers.

$N$	total time	avg. loc. time	#iter	error	# Lagrange
2	24.503039	2.738385	56	0.000006	709
3	26.543736	1.528688	69	0.000007	1143
6	30.922103	0.559905	96	0.000006	2168
12	35.764327	0.216757	126	0.000006	3592
25	32.226731	0.077579	133	0.000005	5875
50	30.192864	0.031712	135	0.000006	8962
75	29.748844	0.019081	142	0.000005	11148
100	26.638113	0.013545	131	0.000005	13012
125	28.644818	0.010270	153	0.000005	14674
150	24.588265	0.008301	131	0.000005	16219
175	24.746298	0.006899	137	0.000005	17715
200	23.694787	0.005947	134	0.000005	19056
225	23.043046	0.005181	131	0.000005	20082
250	21.909531	0.004601	125	0.000005	21372
275	21.473827	0.004131	123	0.000005	22440
300	21.365941	0.003765	123	0.000005	23460
325	20.811414	0.003414	121	0.000005	24544
350	22.405915	0.003150	133	0.000005	25459
375	20.679575	0.002942	119	0.000005	26248
400	21.062826	0.002720	123	0.000005	27324
500	21.282017	0.002164	123	0.000005	30703
600	21.107394	0.001756	121	0.000005	33808
700	19.643515	0.001498	106	0.000005	36822
800	20.233496	0.001295	109	0.000005	39304
900	19.592730	0.001145	101	0.000005	41693
1000	20.692352	0.001033	105	0.000005	44467
1100	21.770149	0.000929	112	0.000005	46611
1200	20.503277	0.000848	98	0.000005	48813
1300	21.781626	0.000782	105	0.000005	50727
1400	22.920965	0.000732	108	0.000005	53107
1500	22.128856	0.000679	100	0.000005	54925
1600	22.188883	0.000636	95	0.000005	56632

**Table 9.4:** Results of the non-preconditioned solver. Columns: number of subdomains, total CPU time for solution, averaged time for solution of local problems, number of iterations, residual error, number of Lagrange multipliers.



# Chapter 10

## Conclusion and outlook

### 10.1 Conclusion

We conclude the thesis by summarizing the main results.

We have derived the skeletal variational formulation (4.5) for general elliptic partial differential equations and arbitrary polytopal meshes and have shown it to be equivalent to the standard variational formulation of these problems. In order to derive a numerical method from the skeletal variational formulation, we discretized the involved skeletal function spaces by prescribing piecewise linear data on a skeletal triangulation; see Section 4.3.1. Furthermore, in order to make the scheme computable, we have made use of an approximation of the element-level Steklov-Poincaré operators  $S_T$  by boundary element methods as described in Section 3.3, where local spaces of piecewise constant boundary functions were used to approximate the Neumann traces. The resulting stiffness matrix is sparse and, for symmetric elliptic problems, positive definite and symmetric. The element stiffness matrices can be assembled using existing BEM quadrature routines.

After developing new analytical tools for polytopal meshes in Chapter 5, we employed them to perform a rigorous error analysis for a model problem in Chapter 6. An error estimate in the  $H^1$ -norm of optimal order, Theorem 6.5.1, was obtained by a Strang-type lemma. Due to the approximation of the bilinear form by BEM techniques, no Galerkin orthogonality holds for the primal formulation, and we passed to a mixed formulation in order to be able to apply the Aubin-Nitsche trick and thus obtain a quasi-optimal error estimate in the  $L_2$ -norm, Theorem 6.6.3.

An efficient solver for the resulting linear system based on the ideas of the one-level FETI approach was derived in Chapter 7, and the Dirichlet preconditioner was adapted to our setting. After establishing a spectral equivalence between the BEM-based FEM Schur complement and the exact Steklov-Poincaré operator in Theorem 7.2.2, the condition numbers of both the non-preconditioned (Theorem 7.2.7) and the preconditioned solver (Theorem 7.2.8) could be analyzed by transferring known results from the FETI literature.

We also considered convection-diffusion equations, and in particular convection-dominated problems, in Chapter 8 and showed that the BEM-based FEM for such problems is closely linked to the stabilized finite element method of residual-free bubbles, and thus in turn also to the SUPG method.

Finally, we gave some details on an implementation of the BEM-based FEM and discussed some numerical examples in Chapter 9.

## 10.2 Outlook and further work

We mention several avenues along which the study of the BEM-based FEM could be advanced.

### Generalized error analysis

We note that the error analysis presented in this thesis is limited to the case of the three-dimensional Laplace equation. The analysis could be extended as follows.

**More general partial differential operators.** The analysis was performed for the Laplace operator,  $L = -\Delta$ . Most tools are in place to treat the diffusion equation with elementwise constant scalar diffusion coefficient, and the analysis generalizes in a straightforward manner to this case. The analysis of the convection-diffusion or general convection-diffusion-reaction problem seems more challenging. We note that the corresponding boundary integral operators introduced in Section 3.1 are less well studied for this case, and in particular the results on explicit constants from Section 3.4 need to be extended to this case. Since these partial differential equations are not formally self-adjoint, no energy arguments can be used in the analysis, which further complicates some estimates. Finally, we point out that vector-valued PDEs like the equations of linearized elasticity or the Maxwell equations are beyond the scope of the present work, but certainly of interest. The latter have been treated numerically by a BEM-based FEM with some success in [24], although the full analysis is still an open problem.

**Inhomogeneous partial differential equations.** Since boundary integral operators were used heavily in the construction of the method, we have restricted ourselves to the case of homogeneous partial differential equations for the sake of simplicity. Non-zero right-hand sides could be incorporated by the use of Newton potentials. We expect that this would not create major difficulties in the analysis since only the right-hand side of the variational problem is affected. For the numerical realization, a simplicial mesh of the elements, thus far only used as an analytical tool, could be introduced also into the numerical scheme for approximating the volume integrals occurring in the Newton potentials.

**Analysis for the two-dimensional case.** It may seem curious that the two-dimensional case offers additional challenges over the three-dimensional one that we have studied. Indeed, the analytical tools for polytopal elements that we derived in Chapter 5 have already been generalized to be valid for both  $d = 2$  and  $d = 3$ . The difficulties lie in the explicit estimates for boundary integral operators from Section 3.4. A crucial point in obtaining upper bounds was that the capacity,  $\lambda = Vw_{\text{eq}}$ , could be bounded in an explicit way. This is relatively easy to do in the 3D case (see [95, Proof of Lemma

6.7)), but more difficult in the 2D case. We point to the original article by Pechstein [95, Remark 4] for a discussion on how this technical difficulty could be tackled.

### Study of the stability for convection-diffusion problems

We have seen in Chapter 8 that the BEM-based FEM, when realized exactly, is equivalent to the exactly realized method of residual-free bubbles, and thus also to the SUPG method with a particular choice of the stabilization parameter. What is lacking is a rigorous analysis of the situation where the Steklov-Poincaré operators are approximated, which occurs in practice. Here again a careful analysis of the involved boundary integral operators is required, and the literature on these operators for the convection-diffusion equation is sparse compared to the well-studied case of potential equations. Furthermore, the numerical results in Section 9.2.2 indicate that the fully numeric quadrature rule for the boundary integrals may be unstable for convection-dominated problems, and this issue needs further analysis.

### A posteriori error estimators and adaptive refinement

We have not touched upon this issue in the present work, but we point out that Steffen Weißer has made significant contributions to this topic in [123] and his dissertation [124]. He adapted the residual error estimator known from the adaptive FEM to the exact version of the BEM-based FEM and proved reliability of the new estimator. Optimal convergence for non-smooth solutions of an adaptive scheme based on this estimator was demonstrated numerically. Further work could be done by proving this optimal convergence, studying how the approximation of the Steklov-Poincaré operators influences the quality of the a posteriori estimator, and adapting other types of error estimators, e.g., so-called functional error estimators which are based on methods of duality theory (cf., e.g., [98] and the monograph [99]).

### Higher-order trial functions

The skeletal trial functions we used in this thesis were of piecewise linear type, and the convergence rate for smooth functions is limited by this choice. It is therefore natural to think about generalizing the BEM-based FEM approach to higher order trial functions. We point to a recent publication by Rjasanow and Weißer [103] where the method was successfully generalized to quadratic trial functions, with an according increase of the rate of convergence by one order of  $h$ . The approach taken there was to enrich the trial space by bubble functions which are quadratic on the element edges, vanish in the vertices, and are harmonic within the elements. For inhomogeneous problems, also element bubbles defined like  $\varphi_T$  in the residual-free bubbles method, cf. Section 8.2, were introduced. A further generalization to higher-order polynomials seems worth pursuing.

### **Alternative solution techniques**

The fast solver presented in Chapter 7 was based directly on the classical one-level finite element tearing/interconnecting (FETI) approach. Methods that were developed later, like the dual-primal FETI method (FETI-DP, [38, 84]) or balancing domain decomposition by constraints (BDDC, [29, 82, 85]) could be adapted in a similar fashion, and thanks to the spectral equivalences shown in Section 7.2, we expect the analysis of these methods to transfer to the case of the BEM-based FEM in a relatively simple way.

As an alternative solution approach, one might also consider investigating an algebraic multigrid solver for the BEM-based FEM, which early numerical experiments have shown to be promising [25].

## Bibliography

- [1] R. A. Adams and J. J. F. Fournier. *Sobolev Spaces*, volume 140 of *Pure and Applied Mathematics*. Academic Press, Amsterdam, Boston, second edition, 2003.
- [2] D. N. Arnold, F. Brezzi, B. Cockburn, and L. D. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J. Numer. Anal.*, 39(5):1749–1779, 2001/02. ISSN 0036-1429. doi: 10.1137/S0036142901384162.
- [3] I. Babuška and J. E. Osborn. Generalized finite element methods: Their performance and their relation to mixed methods. *SIAM Journal on Numerical Analysis*, 20(3):510–536, 1983. URL <http://www.jstor.org/stable/2157269>.
- [4] I. Babuška. Courant element: Before and after. In M. Křížek, P. Neittaanmäki, and R. Stenberg, editors, *Finite Element Methods: Fifty years of the Courant element*, volume 164 of *Lecture notes in Pure and Applied Mathematics*, pages 37–51, New York, 1994. Marcel Dekker, Inc.
- [5] C. Baiocchi, F. Brezzi, and L. P. Franca. Virtual bubbles and Galerkin-least-squares type methods. *Comput. Meth. Appl. Mech. Engrg.*, 105:125–141, 1993. doi: 10.1016/0045-7825(93)90119-I.
- [6] M. Bebendorf. A note on the Poincaré inequality for convex domains. *Z. Anal. Anwendungen*, 22(4):751–756, 2003.
- [7] M. Bebendorf. *Hierarchical Matrices*. Springer-Verlag, Berlin, Heidelberg, 2008.
- [8] J. Bergh and J. Löfström. *Interpolation Spaces: An Introduction*, volume 223 of *Grundlehren der mathematischen Wissenschaften*. Springer, Heidelberg, 1976.
- [9] S. Börm, L. Grasedyck, and W. Hackbusch. Hierarchical matrices. Lecture Note 21, Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig, 2003. URL <http://www.mis.mpg.de/de/publications/andere-reihen/ln/lecturenote-2103.html>. Revised 2006.
- [10] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, second edition, 2002.

- [11] F. Brezzi. Interacting with the subgrid world. *Numerical analysis*, 2000. URL <http://www.imati.cnr.it/brezzi/papers/sgw.pdf>.
- [12] F. Brezzi and A. Russo. Choosing bubbles for advection-diffusion problems. *Math. Models Methods Appl. Sci.*, 4:571–587, 1994.
- [13] F. Brezzi, M.-O. Bristeau, L. P. Franca, M. Mallet, and G. Rogé. A relationship between stabilized finite element methods and the Galerkin method with bubble functions. *Comput. Meth. Appl. Mech. Engrg.*, 96(1):117–129, 1992. ISSN 0045-7825. doi: 10.1016/0045-7825(92)90102-P.
- [14] F. Brezzi, L. P. Franca, T. J. R. Hughes, and A. Russo.  $b = \int g$ . *Comput. Meth. Appl. Mech. Engrg.*, 145(3–4):329–339, 1997. doi: 10.1016/S0045-7825(96)01221-2.
- [15] F. Brezzi, D. Marini, and A. Russo. Applications of the pseudo residual-free bubbles to the stabilization of convection-diffusion problems. *Comput. Meth. Appl. Mech. Engrg.*, 166(1–2):51–63, 1998. doi: 10.1016/S0045-7825(98)00082-6.
- [16] F. Brezzi, T. J. R. Hughes, L. D. Marini, A. Russo, and E. Süli. A priori error analysis of residual-free bubbles for advection-diffusion problems. *SIAM J. Numer. Anal.*, 36(6):1933–1948, 1999.
- [17] F. Brezzi, D. Marini, and E. Süli. Residual-free bubbles for advection-diffusion problems: the general error analysis. *Numerische Mathematik*, 85:31–47, 2000. doi: 10.1007/s002110050476.
- [18] F. Brezzi, K. Lipnikov, and M. Shashkov. Convergence of the mimetic finite difference method for diffusion problems on polyhedral meshes. *SIAM J. Numer. Anal.*, 43(3):1872–1896, 2005.
- [19] F. Brezzi, K. Lipnikov, and V. Simoncini. A family of mimetic finite difference methods on polygonal and polyhedral meshes. *Math. Models Methods Appl. Sci.*, 15(10):1533–1551, 2005. ISSN 0218-2025. doi: 10.1142/S0218202505000832.
- [20] A. N. Brooks and T. J. R. Hughes. Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations. *Comput. Meth. Appl. Mech. Engrg.*, 32(1–3):199–259, 1982. doi: 10.1016/0045-7825(82)90071-8.
- [21] A.-P. Calderón. Lebesgue spaces of differentiable functions and distributions. In *Proc. Sympos. Pure Math., Vol. IV*, pages 33–49. American Mathematical Society, Providence, R.I., 1961.
- [22] P. G. Ciarlet. *The finite element method for elliptic problems*, volume 4 of *Studies in Mathematics and its Applications*. North-Holland, Amsterdam, 1978.

- [23] R. Cools. An encyclopaedia of cubature formulas. *Journal of Complexity*, 19(3): 445–453, 2003. doi: 10.1016/S0885-064X(03)00011-6.
- [24] D. M. Copeland. Boundary-element-based finite element methods for Helmholtz and Maxwell equations on general polyhedral meshes. *Int. J. Appl. Math. Comput. Sci.*, 5(1):60–73, 2009.
- [25] D.M. Copeland, U. Langer, and D. Pusch. From the boundary element method to local Trefftz finite element methods on polyhedral meshes. In M. Bercovier, M. J. Gander, R. Kornhuber, and O. Widlund, editors, *Domain Decomposition Methods in Science and Engineering XVIII*, volume 70 of *Lecture Notes in Computational Science and Engineering*, pages 315–322. Springer-Verlag, Heidelberg, Berlin, 2009.
- [26] M. Costabel. Symmetric methods for the coupling of finite elements and boundary elements. In C.A. Brebbia, W.L. Wendland, and G. Kuhn, editors, *Boundary Elements IX*, pages 411–420. Springer, Berlin, Heidelberg, New York, 1987.
- [27] R. Courant. Variational methods for the solution of problems of equilibrium and vibration. *Bull. Amer. Math. Soc.*, 49:1–23, 1943.
- [28] T. A. Davis. Algorithm 832: UMFPACK v4.3—an unsymmetric-pattern multifrontal method. *ACM Trans. Math. Softw.*, 30(2):196–199, June 2004. ISSN 0098-3500. doi: 10.1145/992200.992206.
- [29] C. R. Dohrmann. A preconditioner for substructuring based on constrained energy minimization. *SIAM J. Sci. Comput.*, 25(1):246–258, January 2003. doi: 10.1137/S1064827502412887.
- [30] V. Dolejší, M. Feistauer, and V. Sobotiková. Analysis of the discontinuous Galerkin method for nonlinear convection-diffusion problems. *Comput. Methods Appl. Mech. Engrg.*, 194(25-26):2709–2733, 2005. ISSN 0045-7825. doi: 10.1016/j.cma.2004.07.017.
- [31] J. Droniou. Non-coercive linear elliptic problems. *Potential Analysis*, 17:181–203, 2002. doi: 10.1023/A:1015709329011.
- [32] Y. Efendiev and T. Y. Hou. *Multiscale Finite Element Methods: Theory and Applications*, volume 4 of *Surveys and Tutorials in the Applied Mathematical Sciences*. Springer, 2009.
- [33] L. C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.
- [34] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In P.G. Ciarlet and J.L. Lions, editors, *Solution of Equation in  $\mathbb{R}^n$  (Part 3)*, *Techniques of Scientific*

- Computing (Part 3)*, volume 7 of *Handbook of Numerical Analysis*, pages 713–1018. Elsevier, 2000. doi: 10.1016/S1570-8659(00)07005-8.
- [35] C. Farhat and F.-X. Roux. A method of finite element tearing and interconnecting and its parallel solution algorithm. *Int. J. Numer. Meth. Engrg.*, 32:1205–1227, 1991.
- [36] C. Farhat and F.-X. Roux. An unconventional domain decomposition method for an efficient parallel solution of large-scale finite element systems. *SIAM J. Sci. Stat. Comput.*, 13(1):379–396, January 1992. doi: 10.1137/0913020.
- [37] C. Farhat, J. Mandel, and F. X. Roux. Optimal convergence properties of the FETI domain decomposition method. *Computer Methods in Applied Mechanics and Engineering*, 115(3–4):365–385, 1994. ISSN 0045-7825. doi: 10.1016/0045-7825(94)90068-X. URL <http://www.sciencedirect.com/science/article/pii/004578259490068X>.
- [38] C. Farhat, M. Lesoinne, P. Le Tallec, K. Pierson, and D. Rixen. FETI-DP: A dual-primal unified FETI method I: A faster alternative to the two-level FETI method. *Internat. J. Numer. Methods Engrg.*, 50:1523–1544, 2001.
- [39] H. Federer and W. H. Fleming. Normal and integral currents. *Ann. of Math.*, 2 (72):482–520, 1960.
- [40] D. Feuchter, M. Heisig, and G. Wittum. A geometry model for the simulation of drug diffusion through the stratum corneum. *Computing and Visualization in Science*, 9:117–130, 2006. ISSN 1432-9360. doi: 10.1007/s00791-006-0017-x.
- [41] L. P. Franca, S. L. Frey, and T. J. R. Hughes. Stabilized finite element methods: I. Application to the advective-diffusive model. *Comput. Meth. Appl. Mech. Engrg.*, 95(2):253–276, 1992. doi: 10.1016/0045-7825(92)90143-8.
- [42] L. P. Franca, A. Nesliturk, and M. Stynes. On the stability of residual-free bubbles for convection-diffusion problems and their approximation by a two-level finite element method. *Comput. Meth. Appl. Mech. Engrg.*, 166(1–2):35–49, 1998. doi: 10.1016/S0045-7825(98)00081-4.
- [43] L. P. Franca, G. Hauke, and A. Masud. Revisiting stabilized finite element methods for the advective-diffusive equation. *Comput. Meth. Appl. Mech. Engrg.*, 195(13–16):1560–1572, 2006. doi: 10.1016/j.cma.2005.05.028. A Tribute to Thomas J.R. Hughes on the Occasion of his 60th Birthday.
- [44] M. J. Gander and G. Wanner. From Euler, Ritz and Galerkin to modern computing. *SIAM Review*, 2012. URL <http://www.unige.ch/~gander/Preprints/Ritz.pdf>. To appear.

- [45] I. Georgieva and C. Hofreither. Interpolation of harmonic functions based on Radon projections. Technical Report 2012-11, DK Computational Mathematics Linz Report Series, 2012. <https://www.dk-compmath.jku.at/publications/dk-reports/2012-10-17/view>. To be submitted.
- [46] I. Georgieva and C. Hofreither. An algebraic method for reconstruction of harmonic functions via Radon projections. *AIP Conference Proceedings*, 1487(1):112–119, 2012. doi: 10.1063/1.4758948.
- [47] I. Georgieva, C. Hofreither, C. Koutschan, V. Pillwein, and T. Thanatipanonda. Harmonic interpolation based on Radon projections along the sides of regular polygons. *Central European Journal of Mathematics*, 2012. To appear. Also available as Technical Report 2011-12 in the series of the DK Computational Mathematics Linz, <https://www.dk-compmath.jku.at/publications/dk-reports/2011-10-20/view>.
- [48] I. Georgieva, C. Hofreither, and R. Uluchev. Least squares fitting of harmonic functions based on Radon projections. In *Proceedings of Curves & Surfaces 2012, Oslo*, 2012. Submitted.
- [49] I. Georgieva, C. Hofreither, and R. Uluchev. Interpolation of mixed type data by bivariate polynomials. In G. Nikolov and R. Uluchev, editors, *Constructive Theory of Functions, Sozopol 2010: In memory of Borislav Bojanov*, pages 93–107. Prof. Marin Drinov Academic Publishing House, Sofia, 2012. Also available as Technical Report 2010-14 in the series of the DK Computational Mathematics Linz, <https://www.dk-compmath.jku.at/publications/dk-reports/2010-12-10/view>.
- [50] C. J. Gittelsohn, R. Hiptmair, and I. Perugia. Plane wave discontinuous Galerkin methods: analysis of the  $h$ -version. *M2AN Math. Model. Numer. Anal.*, 43(2): 297–331, 2009. doi: 10.1051/m2an/2009002.
- [51] M. Gnewuch and S. A. Sauter. Boundary integral equations for second order elliptic boundary value problems. Technical Report 55, Max-Planck-Institut Leipzig, Leipzig, Germany, 1999.
- [52] D. Gordon and R. Gordon. Row scaling as a preconditioner for some nonsymmetric linear systems with discontinuous coefficients. *J. Computational Applied Mathematics*, 234(12):3480–3495, 2010. doi: 10.1016/j.cam.2010.05.021.
- [53] N. I. M. Gould, J. A. Scott, and Y. Hu. A numerical evaluation of sparse direct solvers for the solution of large sparse symmetric linear systems of equations. *ACM Trans. Math. Softw.*, 33(2), June 2007. doi: 10.1145/1236463.1236465.
- [54] I.G. Graham, T.Y. Hou, O. Lakkis, and R. Scheichl, editors. *Numerical Analysis of Multiscale Problems*, volume 83 of *Lecture Notes in Computational Science and Engineering*. Springer, Heidelberg, 2011.

- [55] C. Großmann and H.-G. Roos. *Numerische Behandlung partieller Differentialgleichungen*. Teubner Studienbücher Mathematik. Springer, Berlin, Heidelberg, third edition, 2005.
- [56] W. Hackbusch. *Hierarchische Matrizen: Algorithmen und Analysis*. Springer, 2009.
- [57] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Research Nat. Bur. Standards*, 49(6):409–436, December 1952.
- [58] R. Hiptmair, A. Moiola, and I. Perugia. Plane wave discontinuous Galerkin methods for the 2d Helmholtz equation: Analysis of the  $p$ -version. *SIAM J. Numer. Anal.*, 49(1):264–284, 2011. doi: 10.1137/090761057.
- [59] C. Hofreither.  $L_2$  error estimates for a nonstandard finite element method on polyhedral meshes. *J. Numer. Math.*, 19(1):27–39, 2011.
- [60] C. Hofreither and C. Pechstein. A rigorous error analysis of coupled FEM-BEM problems with arbitrary many subdomains. In T. Apel and O. Steinbach, editors, *Advanced Finite Element Methods and Applications*, volume 66 of *Lecture Notes in Applied and Computational Mechanics*, pages 109–130. Springer, 2012.
- [61] C. Hofreither, U. Langer, and C. Pechstein. Analysis of a non-standard finite element method based on boundary integral operators. *Electronic Transactions on Numerical Analysis*, 37:413–436, 2010. <http://etna.mcs.kent.edu/vol.37.2010/pp413-436.dir>.
- [62] C. Hofreither, U. Langer, and C. Pechstein. A non-standard finite element method for convection-diffusion-reaction problems on polyhedral meshes. In M. D. Todorov and C. I. Christov, editors, *Proceedings of the Third Conference of the Euro-American Consortium for Promoting the Application of Mathematics in Technical and Natural Sciences*, volume 1404, pages 397–404. AIP Conference Proceedings, 2011. doi: 10.1063/1.3659943.
- [63] C. Hofreither, U. Langer, and C. Pechstein. A non-standard finite element method based on boundary integral operators. In I. Lirkov, S. Margenov, and J. Wasniewski, editors, *Large-Scale Scientific Computing*, volume 7116 of *Lecture Notes in Computer Science*, pages 28–39. Springer Berlin / Heidelberg, 2012. doi: 10.1007/978-3-642-29843-1\_3.
- [64] G. C. Hsiao and W. L. Wendland. A finite element method for some integral equations of the first kind. *Journal of Mathematical Analysis and Applications*, 58(3):449–481, 1977. doi: 10.1016/0022-247X(77)90186-X.

- [65] G. C. Hsiao and W. L. Wendland. Domain decomposition in boundary element methods. In R. Glowinski, Y. A. Kuznetsov, G. Meurant, J. Périaux, and O. B. Widlund, editors, *Proceedings of the Fourth International Symposium on Domain Decomposition Methods for Partial Differential Equations, Moscow, May 21–25, 1990*, pages 41–49. SIAM, Philadelphia, 1991.
- [66] G. C. Hsiao and W. L. Wendland. *Boundary Integral Equations*, volume 164 of *Applied Mathematical Sciences*. Springer, Heidelberg, 2008.
- [67] G. C. Hsiao, O. Steinbach, and W. L. Wendland. Domain decomposition methods via boundary integral equations. *Journal of Computational and Applied Mathematics*, 125(1–2):521–537, 2000.
- [68] T. J. R. Hughes. Multiscale phenomena: Green’s functions, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods. *Comput. Meth. Appl. Mech. Engrg.*, 127(1):387–401, 1995. doi: 10.1016/0045-7825(95)00844-9.
- [69] T. J. R. Hughes, L. P. Franca, and G. M. Hulbert. A new finite element formulation for computational fluid dynamics: VIII. The Galerkin/least-squares method for advective-diffusive equations. *Comput. Meth. Appl. Mech. Engrg.*, 73(2):173–189, 1989. ISSN 0045-7825. doi: 10.1016/0045-7825(89)90111-4.
- [70] T. J. R. Hughes, G. R. Feijóo, L. Mazzei, and J.-B. Quinicy. The variational multiscale method – A paradigm for computational mechanics. *Comput. Meth. Appl. Mech. Engrg.*, 166(1–2):3–24, 1998. doi: 10.1016/S0045-7825(98)00079-6.
- [71] T. J. R. Hughes, J. A. Cottrell, and Y. Bazilevs. Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement. *Computer Methods in Applied Mechanics and Engineering*, 194(39–41):4135–4195, 2005. doi: 10.1016/j.cma.2004.10.008.
- [72] C. Johnson, U. Nävert, and Juhani Pitkäranta. Finite element methods for linear hyperbolic problems. *Comput. Meth. Appl. Mech. Engrg.*, 45(1–3):285–312, 1984. ISSN 0045-7825. doi: 10.1016/0045-7825(84)90158-0.
- [73] P. W. Jones. Quasiconformal mappings and extendability of functions in Sobolev spaces. *Acta Math.*, 147:71–88, 1981.
- [74] G. Karypis and V. Kumar. A fast and highly quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on Scientific Computing*, 20(1):359–392, 1999.
- [75] W. A. Kirk and B. Sims. *Handbook of Metric Fixed Point Theory*. Kluwer Academic Publishers, Dordrecht, 2001.

- [76] A. Klawonn and O. B. Widlund. FETI and Neumann-Neumann iterative substructuring methods: connections and new results. *Comm. Pure Appl. Math.*, 54(1): 57–90, 2001.
- [77] Y. Kuznetsov, K. Lipnikov, and M. Shashkov. The mimetic finite difference method on polygonal meshes for diffusion-type problems. *Comput. Geosci.*, 8(4):301–324, 2004.
- [78] U. Langer. Parallel iterative solution of symmetric coupled FE/BE-equation via domain decomposition. In A. Quarteroni, J. Periaux, Y.A. Kuznetsov, and O.B. Widlund, editors, *Domain decomposition methods in science and engineering (Como, 1992)*, volume 157 of *Contemporary Mathematics*, pages 335–344. Amer. Math. Soc., Providence, Rhode Island, 1994.
- [79] U. Langer and O. Steinbach. Boundary element tearing and interconnecting methods. *Computing*, 71(3):205–228, 2003.
- [80] U. Langer and O. Steinbach. Coupled finite and boundary element domain decomposition methods. In M. Schanz and O. Steinbach, editors, *Boundary Element Analysis: Mathematical Aspects and Application*, volume 29 of *Lecture Notes in Applied and Computational Mechanics*, pages 29–59. Springer, Berlin, 2007.
- [81] X. S. Li. An overview of SuperLU: Algorithms, implementation, and user interface. *ACM Trans. Math. Softw.*, 31(3):302–325, September 2005. doi: 10.1145/1089014.1089017.
- [82] J. Mandel and C. R. Dohrmann. Convergence of a balancing domain decomposition by constraints and energy minimization. *Numer. Lin. Alg. Appl.*, 10:639–659, 2003.
- [83] J. Mandel and R. Tezaur. Convergence of a substructuring method with Lagrange multipliers. *Numer. Math.*, 73:473–487, 1996.
- [84] J. Mandel and R. Tezaur. On the convergence of a dual-primal substructuring method. *Numer. Math.*, 88:543–558, 2001.
- [85] J. Mandel, C. R. Dohrmann, and R. Tezaur. An algebraic theory for primal and dual substructuring methods by constraints. *Appl. Numer. Math.*, 54:167–193, 2005.
- [86] V. G. Maz’ya. Classes of domains and imbedding theorems for functions spaces. *Soviet Math. Dokl.*, 1:882–885, 1960.
- [87] W. McLean. *Strongly Elliptic Systems and Boundary Integral Equations*. Cambridge University Press, Cambridge, UK, 2000.

- [88] A. Mizukami and T. J. R. Hughes. A Petrov-Galerkin finite element method for convection-dominated flows: An accurate upwinding technique for satisfying the maximum principle. *Comput. Meth. Appl. Mech. Engrg.*, 50(2):181–193, 1985. doi: 10.1016/0045-7825(85)90089-1.
- [89] A. Moiola, R. Hiptmair, and I. Perugia. Plane wave approximation of homogeneous Helmholtz solutions. *ZAMP*, 62(5):809–837, 2011. doi: 10.1007/s00033-011-0147-y.
- [90] T. Oden. Some historic comments on finite elements. In *Proceedings of the ACM conference on History of scientific and numeric computation*, HSNC '87, pages 125–130, New York, NY, USA, 1987. ACM. doi: 10.1145/41579.41592.
- [91] L. A. Oganessian and V. Rivkind. On the FEM research in St. Petersburg. In M. Křížek, P. Neittaanmäki, and R. Stenberg, editors, *Finite Element Methods: Fifty years of the Courant element*, volume 164 of *Lecture notes in Pure and Applied Mathematics*, pages 335–343, New York, 1994. Marcel Dekker, Inc.
- [92] C. C. Paige and M. A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal.*, 12:617–629, 1975.
- [93] L. E. Payne and H. F. Weinberger. An optimal Poincaré inequality for convex domains. *Arch. Rat. Mech. Anal.*, 5:286–292, 1960.
- [94] C. Pechstein. *Finite and boundary element tearing and interconnecting methods for multiscale elliptic partial differential equations*. PhD thesis, Johannes Kepler University, Linz, Austria, December 2008. <http://www.numa.uni-linz.ac.at/Teaching/PhD/Finished/pechstein>.
- [95] C. Pechstein. Shape-explicit constants for some boundary integral operators. *Applicable Analysis*, 2011. doi: 10.1080/00036811.2011.643781.
- [96] C. Pechstein. *Finite and Boundary Element Tearing and Interconnecting Solvers for Multiscale Problems*, volume 90 of *Lecture Notes in Computational Science and Engineering*. Springer, 2013. In press.
- [97] C. Pechstein and R. Scheichl. Weighted Poincaré inequalities. NuMa-Report 2010-10, Institute of Computational Mathematics, Johannes Kepler University Linz, Austria, 2010. To appear in *IMA J. Numer. Anal.*
- [98] S. Repin. A posteriori error estimates for approximate solutions of variational problems with functionals of power growth. *Journal of Mathematical Sciences*, 101: 3531–3538, 2000. doi: 10.1007/BF02680150.
- [99] S. Repin. *A Posteriori Estimates for Partial Differential Equations*, volume 4 of *Radon Series on Computational and Applied Mathematics*. de Gruyter, Berlin, 2008.

- [100] W. Ritz. Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik. *Journal für die reine und angewandte Mathematik*, 135, 1909.
- [101] D. Rixen and C. Farhat. A simple and efficient extension of a class of substructure based preconditioners to heterogeneous structural mechanics problems. *International Journal for Numerical Methods in Engineering*, 44(4):489–516, 1999. ISSN 1097-0207. doi: 10.1002/(SICI)1097-0207(19990210)44:4<489::AID-NME514>3.0.CO;2-Z.
- [102] S. Rjasanow and O. Steinbach. *The Fast Solution of Boundary Integral Equations (Mathematical and Analytical Techniques with Applications to Engineering)*. Springer-Verlag, New York, 2007. ISBN 0387340416.
- [103] S. Rjasanow and S. Weißer. Higher order BEM-based FEM on polygonal meshes. *SIAM Journal on Numerical Analysis*, 50(5):2357–2378, 2012. doi: 10.1137/110849481.
- [104] H.-G. Roos, M. Stynes, and L. Tobiska. *Numerical Methods for Singularly Perturbed Differential Equations*, volume 24 of *Springer Series in Computational Mathematics*. Springer, Berlin, Heidelberg, 1996.
- [105] Y. Saad and M. Schultz. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*, 7(3):856–869, 1986. doi: 10.1137/0907058.
- [106] S. A. Sauter and C. Schwab. *Boundary Element Methods*, volume 39 of *Springer Series in Computational Mathematics*. Springer, Berlin, Heidelberg, 2011.
- [107] K. Schellbach. Probleme der Variationsrechnung. *Journal für die reine und angewandte Mathematik*, 41:293–363, 1852.
- [108] O. Schenk and K. Gärtner. Solving unsymmetric sparse systems of linear equations with PARDISO. *Future Generation Computer Systems*, 20(3):475–487, 2004. doi: 10.1016/j.future.2003.07.011.
- [109] L. R. Scott and S. Zhang. Finite element interpolation of non-smooth functions satisfying boundary conditions. *Math. Comp.*, 54:483–493, 1990.
- [110] E. M. Stein. *Singular Integrals and Differentiability Properties of Functions*, volume 30 of *Princeton Math Series*. Princeton University Press, Princeton, New Jersey, 1970.
- [111] O. Steinbach. OSTBEM. A boundary element software package. Technical report, University of Stuttgart, Germany, 2000.

- [112] O. Steinbach. *Stability estimates for hybrid coupled domain decomposition methods*, volume 1809 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2003.
- [113] O. Steinbach. *Numerical Approximation Methods for Elliptic Boundary Value Problems. Finite and Boundary Elements*. Springer-Verlag, New York, 2008.
- [114] O. Steinbach and W. L. Wendland. On C. Neumann's method for second-order elliptic systems in domains with non-smooth boundaries. *J. Math. Anal. Appl.*, 262:733–748, 2001.
- [115] G. Strang and G. Fix. *An Analysis of The Finite Element Method*. Prentice Hall, 1973.
- [116] A. Thrun. Über die Konstanten in Poincaréschen Ungleichungen. Master's thesis, Ruhr-Universität Bochum, Bochum, 2003. [http://www.ruhr-uni-bochum.de/num1/files/theses/da\\_thrun.pdf](http://www.ruhr-uni-bochum.de/num1/files/theses/da_thrun.pdf).
- [117] A. Toselli and O. Widlund. *Domain Decoposition Methods - Algorithms and Theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer, Berlin, Heidelberg, 2004.
- [118] E. Trefftz. Ein Gegenstück zum Ritzschen Verfahren. In *Verh. d. 2. Intern. Kongr. f. Techn. Mech.*, pages 131–137, Zürich, 1926.
- [119] H. Triebel. *Interpolation theory, function spaces, differential operators*. Johann Ambrosius Barth Verlag, Heidelberg, second edition, 1995.
- [120] U. van Rienen. *Numerical Methods in Computational Electrodynamics*. Springer, 2001.
- [121] A. Veeseer and R. Verfürth. Poincaré constants for finite element stars. *IMA Journal of Numerical Analysis*, 32(1):30–47, 2012. doi: 10.1093/imanum/drr011. URL <http://imajna.oxfordjournals.org/content/32/1/30.abstract>.
- [122] T. Weiland. A discretization model for the solution of Maxwell's equations for six-component fields. *Archiv Elektronik und Uebertragungstechnik*, 31:116–120, March 1977. In German.
- [123] S. Weißer. Residual error estimate for BEM-based FEM on polygonal meshes. *Numerische Mathematik*, 118(4):765–788, 2011. doi: 10.1007/s00211-011-0371-6. URL <http://www.springerlink.com/content/r0510th5qx0015r6/>.
- [124] S. Weißer. *Finite Element Methods with local Trefftz trial functions*. PhD thesis, Universität des Saarlandes, Saarbrücken, Germany, 2012.





# Curriculum Vitae

## Personal details

**Full name** Clemens Hofreither  
**Title** Diplomingenieur (DI)  
**Date of birth** January 30, 1984  
**Place of birth** Linz, Austria  
**Nationality** Austria

## Career

2002 High School Diploma (Matura) in Seitenstetten, Lower Austria.  
2002–2003 Studies in Mathematics at Johannes Kepler University Linz, Austria.  
2003–2004 Civilian service at Lower Austrian Red Cross, Stadt Haag.  
2004–2008 Continued studies in Mathematics.  
2008 Graduated as Diplomingenieur.  
**Diploma thesis:** *Boundary Element Solvers for Water Wave Simulation in a Model Basin* (Supervisor: U. Langer)  
2008– PhD studies at the Doctoral Program “Computational Mathematics” at Johannes Kepler University Linz, Austria.

## Teaching

WS 2006/07 Tutor: Numerical Mathematics, Diploma program Mechatronics, Johannes Kepler University Linz, Austria.  
WS 2007/08 Tutor: Numerical Mathematics, Diploma program Mechatronics, Johannes Kepler University Linz, Austria.  
SS 2012 Proseminar “Mathematische Modelle in der Technik” and supervision of bachelor theses, Johannes Kepler University Linz, Austria.

## Research stays

Oct 2010–Feb 2011	Universität des Saarlandes, Saarbrücken, Germany. Visiting the group of Prof. Sergej Rjasanow.
Jun 2011–Jul 2011	Todor Kableshkov University of Transport, Sofia, Bulgaria.
Mar 2012–Apr 2012	Universität des Saarlandes, Saarbrücken, Germany. Visiting the group of Prof. Sergej Rjasanow.
May 2012–Jun 2012	Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria.

## Selected Publications

- [1] C. Hofreither, U. Langer, and S. Tomar. Boundary element simulation of linear water waves in a model basin. In I. Lirkov, S. Margenov, and J. Wasniewski, editors, *Large-Scale Scientific Computing: Proceedings of LSSC 2009*, Volume 5910 of Lecture Notes in Computer Science, pages 132–139, Springer Verlag, 2010.
- [2] C. Hofreither, U. Langer, and C. Pechstein. Analysis of a non-standard finite element method based on boundary integral operators. *Electron. T. Numer. Ana.*, 37:413–436, 2010.
- [3] C. Hofreither.  $L_2$  error estimates for a nonstandard finite element method on polyhedral meshes. *J. Numer. Math.*, 19(1):27–39, 2011.
- [4] C. Hofreither, U. Langer, and C. Pechstein. A non-standard finite element method for convection-diffusion-reaction problems on polyhedral meshes. In Michail D. Todorov and Christo I. Christov, editors, *Application of Mathematics in Technical and Natural Sciences: 3rd International Conference AMiTaNS'11*, volume 1404, pages 397–404. AIP, 2011.
- [5] I. Georgieva, C. Hofreither, and R. Uluchev. Interpolation of mixed type data by bivariate polynomials. In G. Nikolov and R. Uluchev, editors, *Constructive Theory of Functions, Sozopol 2010: In memory of Borislav Bojanov*. Prof. Marin Drinov Academic Publishing House, Sofia, 2011.
- [6] C. Hofreither and C. Pechstein. A rigorous error analysis of coupled FEM-BEM problems with arbitrary many subdomains. In T. Apel and O. Steinbach, editors, *Advanced Finite Element Methods and Applications*, volume 66 of *Lecture Notes in Applied and Computational Mechanics*, pages 109–130. Springer, 2012.
- [7] I. Georgieva, C. Hofreither, C. Koutschan, V. Pillwein, and T. Thanatipanonda. Harmonic interpolation based on Radon projections along the sides of regular polygons. *Cent. Eur. J. Math.* To appear.
- [8] I. Georgieva and C. Hofreither. Interpolation of harmonic functions based on Radon projections. Technical Report 2012-11 in the DK Computational Mathematics Linz Report Series, 2012. Submitted. <https://www.dk-compmath.jku.at/publications/dk-reports/2012-10-17/view>.



## **Eidesstattliche Erklärung**

Ich erkläre an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe. Die vorliegende Dissertation ist mit dem elektronisch übermittelten Textdokument identisch.

Linz, am 30. Oktober 2012

Clemens Hofreither