INAUGURAL-DISSERTATION

zur
Erlangung der Doktorwürde
der
Gesamtfakultät für Mathematik, Ingenieur- und Naturwissenschaften
der
Ruprecht-Karls-Universität
Heidelberg

vorgelegt von

Strehlow, Arne Paul, M.Sc.
aus Heidelberg

Tag der mündlichen Prüfung:

Generalised Elements on Sheaves and Applications in Spectral Domain Decomposition

Betreuer: Prof. Dr. Robert Scheichl
Prof. Dr. Peter Bastian

Abstract

This thesis develops a unified abstract framework for spectral domain decomposition methods, enabling the formulation and analysis of multilevel solvers for a broad class of partial differential equations. Central to this framework is the introduction of suitable presheaves — a generalization of function spaces that captures essential structural properties, including local bilinear forms and zero extensions. Over such presheaves, we define generalized elements, providing a systematic abstraction of spectral coarse spaces akin to the Generalized Finite Element Method.

A key insight is that generalized elements over a suitable presheaf form a suitable presheaf themselves. This recursive structure yields a natural foundation for multilevel methods. In particular, we develop and analyse abstract multilevel spectral elements, including generalized versions of GenEO and MS-GFEM. A comprehensive convergence analysis of both the additive Schwarz and restricted hybrid Schwarz methods is provided within this unified framework.

The construction encompasses continuous as well as discontinuous finite element discretizations for a wide range of PDE applications. It generalizes existing results on multilevel GenEO and recasts the MS-GFEM preconditioner from earlier work as a multilevel method. New variants are introduced and analysed, including MS-GFEM for Discontinuous Galerkin elements and two formulations of multilevel GenEO with the Weighted Symmetric Interior Penalty discretization which avoid the additional constants present in earlier formulations.

The theoretical results are validated through numerical experiments for heterogeneous diffusion and linear elasticity problems, demonstrating robustness as well as strong and weak scaling. The MS-GFEM preconditioner, in particular, is shown to achieve arbitrarily fast convergence in accordance with theoretical predictions, and a parameter study explores the trade-off between convergence rate and computational cost. As an outlook toward indefinite problems, the methodology is further applied to the Helmholtz equation to great effect.

Zusammenfassung

In der vorliegenden Arbeit wird ein einheitlicher abstrakter Rahmen für spektrale Domänendekompositionsmethoden entwickelt, der die Formulierung und Analyse von Mehrgitterlösern für eine breite Klasse partieller Differentialgleichungen ermöglicht. Im Mittelpunkt steht die Einführung geeigneter Prägarben – einer Verallgemeinerung von Funktionenräumen, die zentrale strukturelle Eigenschaften wie lokale Bilinearformen und Nullfortsetzungen abbildet. Über solchen Prägarben definieren wir verallgemeinerte Elemente und liefern damit eine systematische Abstraktion spektraler Grobräume, ähnlich der Generalized Finite Element Method.

Eine zentrale Erkenntnis ist, dass verallgemeinerte Elemente über einer geeigneten Prägarbe wiederum eine geeignete Prägarbe bilden. Diese rekursive Struktur bietet eine natürliche Grundlage für Mehrgittermethoden. Insbesondere entwickeln und analysieren wir abstrakte mehrgittrige spektrale Elemente, einschließlich verallgemeinerter Versionen von GenEO und MS-GFEM. Innerhalb dieses einheitlichen Rahmens wird eine umfassende Konvergenzanalyse sowohl für die additive Schwarz-Methode als auch für die eingeschränkte hybride Schwarz-Methode bereitgestellt.

Die Konstruktion umfasst sowohl stetige als auch unstetige Finite-Elemente-Diskretisierungen für eine Vielzahl von PDE-Anwendungen. Sie verallgemeinert

bestehende Resultate zu Multilevel GenEO und formuliert den MS-GFEM Präkonditionierer aus früheren Arbeiten als Mehrgittermethode. Neue Varianten werden eingeführt und analysiert, darunter MS-GFEM für Discontinuous-Galerkin-Elemente sowie zwei Formulierungen von Multilevel GenEO mit der Weighted Symmetric Interior Penalty Diskretisierung, welche die in früheren Ansätzen auftretende zusätzliche Konstante vermeiden.

Die theoretischen Ergebnisse werden durch numerische Experimente zu heterogener Diffusion und linearer Elastizität validiert, wobei sowohl Robustheit als auch starkes und schwaches Skalierungsverhalten demonstriert werden. Insbesondere zeigt sich, dass der MS-GFEM Präkonditionierer gemäß den theoretischen Vorhersagen beliebig schnelle Konvergenz erreicht. Eine Parameterstudie untersucht zudem den Kompromiss zwischen Konvergenzrate und Rechenaufwand. Als Ausblick auf indefinierte Probleme wird die Methodik darüber hinaus auf die Helmholtz-Gleichung angewandt.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Rob Scheichl, for his continuous support and guidance throughout my doctoral studies. Rob's optimistic style of leadership, constructive feedback, and ability to see the bigger picture have always provided me with motivation and confidence, even when I doubted my own progress. His encouragement to pursue my own ideas and his trust in my abilities have been invaluable to my development as a researcher.

I am also indebted to my co-supervisor, Prof. Peter Bastian, whose mathematical insight and attention to detail have deeply influenced my work. His critical perspective and technical expertise have been essential, both in research discussions and during the writing of our joint publications.

My sincere thanks go to Prof. Laura Grigori for agreeing to act as my co-examiner. I am honoured to have her as part of my examination committee, and I appreciate the time and effort she has devoted to evaluating my thesis.

I am grateful to my brilliant senior, Chupeng Ma, for sharing his experience, advice, and for the fruitful collaborations that shaped many of the results in this thesis.

Many thanks also to Santiago Ospina De Los Ríos, Jean Bezenech and Dominic Kempf, whose help with C++ and Dune made countless technical hurdles much more manageable. Their expertise and willingness to lend a hand have saved me many hours of frustration.

I would like to thank my colleagues, particularly Robert Kutri and Christian Alber, for many interesting discussions that ranged well beyond the immediate scope of my research and frequently sparked new ideas.

Finally, my deepest appreciation goes to my wife. Her patience, support, and understanding have been nothing short of extraordinary — especially during the last stretch, when I was working hundred-hour weeks in a rush to finish this thesis, just after the birth of our daughter. I could not have done this without her by my side.

To all of you, thank you.

Contents

Introduction

Motivation and Literature

Analytical approaches to partial differential equations (PDEs) go back to the 19th century. Famously, Joseph Fourier introduced the Fourier series [30] for solving the diffusion equation. From a modern perspective, the idea is quite simple: to solve a differential equation, it is best to work in a basis of eigenfunctions of the corresponding differential operator. The Fourier series is exactly such a basis for the Laplace operator. Another popular analytical approach, due to Green [34], is to make use of the solution for a point source, called the fundamental solution or Green's function. The convolution of this fundamental solution with the actual source term yields the desired solution.

Such analytical approaches typically exploit symmetries of the given problem, as they rely on closed forms of their ingredients (eigenfunctions/fundamental solutions) and on solving the resulting integrals with pen and paper. As the complexity of physical models and engineering applications increased however, it became clear that many real-world problems lacked such symmetries. Writing down the Fourier series on a rectangle or disk is a manageable exercise for instance, but if we change the domain shape to something irregular, this can quickly become infeasible.

The advent of computers in the mid-20th century marked a turning point in that regard, as it allowed researchers to move beyond closed-form solutions and toward numerical approximations. Early numerical techniques, such as the finite difference method (FDM) and the finite element method (FEM), emerged in the 1940s and 1950s, providing a framework for approximating solutions to PDEs over discrete meshes. The finite element method [19, 36], eventually overtook finite differences in popularity due to its flexibility in handling complex geometries and boundary conditions, as well as its more elegant mathematical theory.

As computational resources expanded, so did the scope of the problems that could be addressed. However, it became clear that classical methods struggled with multiscale problems, where fine-scale features can dramatically influence large-scale behaviour. Examples include groundwater flow through porous media, composite materials in elasticity, and the modelling of complex biological systems. Traditional finite element methods require highly refined meshes to resolve the fine-scale structures in such problems, which in turn leads to excessively large linear systems.

In response to these challenges, the 1970s and 1980s saw the development of multiscale methods. These approaches aim to capture the effects of small-scale features without resolving them directly. One of the earliest methods in this category was the homogenization technique [13], which averages the fine-scale details into effective coarse-scale equations. However, while homogenization works well for periodic structures and similar cases, it fails to address problems that lack scale separation, such as linear elasticity of fibered materials.

The 1990s and early 2000s witnessed the rise of versatile multiscale methods, including the Multiscale Finite Element Method (MSFEM, [35]) and the Generalized Finite Element Method (GFEM, [59]). They introduced adaptive basis functions, tailored to the given problem by solving local PDE problems on a fine mesh. Around the same time, domain decomposition methods became increasingly popular for their ability to parallelize the solution of PDEs over large computational domains. Methods like the two-level Schwarz method [8] partition the computational domain into smaller subdomains, iteratively alternating local solves with solves in a global coarse space. An appropriate choice of this coarse space is critical for convergence. Applying ideas from multiscale methods to its construction has been a rich research field in the 21st century and will be the topic of this thesis.

Problems that lack scale separation have proven difficult in this regard. Their fine scale features cannot simply be averaged away, which challenges methods that inherit ideas from the homogenization technique. An emerging alternative for such problems are spectral methods such as GMsFEM (Generalized Multiscale Finite Element Method, [23]), GenEO (Generalized Eigenproblems on the Overlap, [56]) or its variant MS-GFEM (Multiscale Generalized Finite Element Method [46]), which handle fine-scale features by building their coarse spaces from solutions to local eigenproblems. In a sense, they return to Fourier's original idea but now localized, within a domain decomposition framework. In Chapter 2 we elaborate in more detail on the historical background and intuition behind these methods.

One issue with the scalability of two-level methods lies in the size of the coarse space as this will eventually become the bottleneck when solving large enough problems with large enough machines. A natural solution to this issue is to again construct a coarse space for the existing coarse space. Iterating this procedure leads to a hierarchy of successively coarser spaces of typically rapidly decreasing dimension which can then be used to construct a multilevel method. If such a multilevel scheme scales perfectly with the number of levels, one can split a given problem into ever smaller subproblems, ultimately achieving a total complexity of $n \log n$. A famously successful example of this is the Fast Fourier Transform which can indeed be phrased as a multigrid method [15]. There are two obstacles with this in multilevel schemes for PDE applications, however. First, their convergence rate typically deteriorates with the number of levels. Second, in overlapping methods each split of the domain introduces additional degrees of freedom due to the overlap. Luckily, only a handful of levels and minimal overlap is often sufficient to avoid excessively large coarse spaces, and they are typically employed in this manner in practice.

In the past, a variety of multilevel methods have been developed that converge (almost) independently of the mesh size h as well as of the number of processors $p \sim H^d$, where d is the spatial dimension of the given domain and H is a bound on the diameter of the employed subdomains. Examples include multigrid (MG) and domain decomposition (DD) methods [55, 61, 21].

While many multilevel methods are robust with respect to the mesh size parameters h and H, achieving robustness in the presence of heterogeneous coefficients or anisotropic meshes is more demanding. The development of robust coarse spaces eventually led to the advent of algebraic multigrid (AMG) techniques [64]. For example, [48] establishes rigorous convergence results for aggregation-based AMG when applied to nonsingular symmetric M-matrices with nonnegative row sums.

Another approach to multilevel methods is to employ a hierarchy of spectral coarse spaces, and this will be the focus of this thesis. GenEO, for instance, has been cast as a multilevel method in [4, 10]. As in the two-level case, multilevel spectral methods are typically extremely robust, for instance with respect to heterogeneous coefficients. Their main drawback, however, is a high setup cost.

Much of the literature on these matters is phrased for particular PDE applications, often down to a specific choice of discretization. While it is usually understood that the given method can in principle be generalized to similar problems, details are often left to the reader. Works such as [2] on the other hand phrase their methods algebraically. This has the benefit of being independent of a particular application. However, it comes at the cost of losing touch with the underlying geometry which is crucial for some important results, such as the near-exponential decay of MS-GFEM's eigenvalues [46].

A more abstract approach is taken in [58] and [42], where the theory is phrased for collections of Hilbert spaces. This thesis takes the abstraction a step further in an

attempt to unify notation. It provides a framework that not only fits a wide range of PDE applications and spectral methods, but also casts all of these into a multilevel method. Moreover, the presented framework encompasses the algebraic as well as the geometric setting, formalizing their equivalence (cf. Example 3.9). We investigate in particular, how the abstract approach unifies the author's previous works [58, 10]. As such, the focus will be on GenEO and MS-GFEM, the spectral methods under discussion in these publications.

GenEO is a spectral multiscale domain decomposition method that was specifically developed for problems lacking scale separation. The original work [56] has sparked numerous follow-up publications such as H-GenEO [14], a variant for the Helmholtz equation. In [2] the authors investigate and generalize GenEO-like coarse spaces from an algebraic viewpoint and [4] takes this to a multilevel scheme. In [10] a multilevel version of GenEO is presented from a geometric perspective.

Meanwhile, MS-GFEM makes use of oversampling and restricts the eigenproblems to locally harmonic subspaces to achieve exponential decay of its eigenvalues. It has originally been phrased as an approximation theory but was also shown to be an effective iterative method with arbitrarily fast convergence rate [58]. A very similar preconditioner has been proposed independently by [49] in an algebraic setting. One advantage of the clean theory of MS-GFEM is that it can readily be expanded to other applications such as the Helmholtz equation [44] or mixed finite elements [5]. We demonstrate this in Section 6.4 for the Helmholtz problem using numerical results presented at the ENUMATH Lisbon 2023. This avenue has since then been continued in [43]. Related results have been independently published in [37, 31].

More on the historical background of the spectral methods under discussion can be found in Chapter 2.

Contributions

This thesis develops an abstract multilevel framework that allows us to phrase spectral methods under general assumptions, capturing a wide range of PDE applications. To this end, we define the notion of a suitable presheaf, essentially, a collection of vector spaces behaving like a function space, with exactly the properties we require for spectral domain decomposition techniques. We then define generalized elements over such a suitable presheaf which provides a framework for spectral methods, much like generalized finite elements do in more concrete settings [59, 46]. Moreover, since generalized elements over a suitable presheaf form a suitable presheaf themselves, the construction can be iterated to obtain a multilevel hierarchy.

We then proceed to define multilevel spectral elements, including the GenEO and MS-GFEM coarse spaces, within this abstract setting. This leads to a general recipe for constructing such methods. In particular, this setup provides the foundation for abstract additive multilevel Schwarz, as well as hybrid restricted multilevel Schwarz preconditioners, which we analyse in Chapter 5 and Chapter 6 respectively.

We show that not only H^1 functions and continuous finite elements, but also discontinuous finite elements fit into this framework. This in particular clarifies technical aspects of multilevel GenEO applied to the discontinuous WSIP discretization as found in [10]. Further, the abstract theory allows us to phrase the MS-GFEM preconditioner from [58] for suitable presheaves. This immediately turns it into a multilevel method which is applicable to a wide range of discretizations.

As such, this thesis provides a framework for the research of its author from [10] and [58] whose experiments we present here, while generalizing theoretical

results. Along the way, several byproducts naturally emerge, such as MS-GFEM for discontinuous finite elements and an MS-GFEM multilevel method, analysed both as an additive (non-restricted) Schwarz, and as a hybrid restricted Schwarz scheme. Moreover, we present a total of three variants of multilevel GenEO with the discontinuous WSIP discretization, two of which do not gather an additional constant as the one used in [58].

Numerical experiments for heterogeneous diffusion and linear elasticity demonstrate the practical applicability of the theoretical findings. A Helmholtz problem serves as an outlook towards indefinite problems.

Outline

In the following we offer an overview of the contents of this thesis. A more detailed summary can be found in the introductions of the individual chapters.

Chapter 1 lays the mathematical groundwork for the thesis, covering functional analysis with an emphasis on Hilbert spaces and bilinear forms, discretization via finite elements, and model multiscale PDEs such as diffusion, elasticity, and Helmholtz problems. It also reviews iterative solvers and two-level domain decomposition methods, concluding with a discussion of approximation spaces and their relation to the coarse spaces in two-level Schwarz preconditioners.

Chapter 2 surveys multiscale methods that influenced GenEO and MS-GFEM, recasting them in the framework of Generalized Finite Elements (GFEM) for comparison. These methods construct adaptive coarse spaces for multiscale PDE problems by solving local linear systems or eigenproblems. The chapter outlines in particular the intuition and ideas that led to GenEO and its variants, and provides examples of the methods that the abstract framework of this thesis aims to generalize.

Chapter 3 develops an abstract framework for spectral domain decomposition, introducing the mathematical object of a suitable presheaf — a presheaf with a set of properties generalizing integral bilinear forms and zero extensions. Continuous and Discontinuous Galerkin (DG) function spaces are shown to fit into this framework, illustrating applicability. We then present the notion of generalized elements over a suitable presheaf, and their natural extension to multilevel hierarchies thereof.

Chapter 4 defines spectral elements, a special case of generalized elements. We state a generally applicable approximation result that will be used for convergence analysis in later chapters. The GenEO and MS-GFEM methods are generalized to this setting and discussed in the context of continuous finite elements and Discontinuous Galerkin discretizations. Finally, multilevel spectral elements are introduced by leveraging the natural multilevel structure of generalized elements.

Chapter 5 extends the GenEO analysis from [10] for a multilevel additive Schwarz preconditioner to the abstract presheaf framework, thus covering both GenEO and MS-GFEM, for continuous finite elements as well as Discontinuous Galerkin discretizations. Adapting standard subspace correction theory, the chapter generalizes the typical convergence results to the presheaf setting. Numerical experiments confirm the methods' practical performance and scalability, including strong and weak scaling with up to 30 million degrees of freedom.

Chapter 6 reviews the MS-GFEM preconditioner for the diffusion equation from [58], for which even a simple Richardson iteration yields arbitrarily fast convergence. The method is then generalized to the abstract framework, enabling a multilevel formulation, again with applicability to both continuous and discontinuous discretizations. Numerical experiments confirm the theoretical results for diffusion and

linear elasticity problems with continuous finite elements, and provide an outlook to Helmholtz problems.

In the conclusions, we wrap up and recapitulate the material presented in this thesis. We discuss gaps in our abstract framework and offer a comparison of MS-GFEM and GenEO based on both theoretical considerations and timing measurements from the experiments.

Chapter 1

Mathematical Background

Before getting into the specifics of spectral solvers, here we establish the mathematical setting, recall some important definitions and fix notation. We review the required functional analysis, relevant discretization techniques, introduce model problems and briefly look at iterative solvers for later use. As this is standard material, proofs are mostly omitted.

Throughout this thesis, we will work with partial differential equations (PDEs) in weak formulation, in the setting of bilinear forms on separable Hilbert spaces. To see how to arrive there from a strong formulation, consider the following classic example.

On a given sufficiently regular domain $\Omega \subset \mathbb{R}^d$, find a function u such that

$$-\Delta u = f \quad \text{on } \Omega \tag{1.1}$$
$$u = 0 \quad \text{on } \partial\Omega \tag{1.2}$$

for a given function f. This is called the Poisson equation (with homogeneous Dirichlet boundary conditions). It arises for instance for diffusion problems in homogeneous materials. The way to treat this problem mathematically seems to be to consider the map

$$\Delta : C^2(\Omega) \to C^0(\Omega)$$

where C^i stands for functions with continuous i-th derivatives. For the theory, one would then show that this map is invertible and calculate

$$u = \Delta^{-1} f.$$

However, we are already assuming a lot of regularity here. In practice f is often not continuous, and we will later want to approximate u with piecewise linear functions which are not differentiable. Treating such issues can be easier in a so-called weak formulation.

Assume that u and f are regular enough for the following to make sense, and let v be any regular enough function on Ω. Then we have, using partial integration,

$$\begin{aligned} &\int_\Omega -\Delta u v \,\mathrm{d}x = \int_\Omega f v \,\mathrm{d}x \\ \iff &\int_\Omega \nabla u \nabla v \,\mathrm{d}x = \int_\Omega f v \,\mathrm{d}x \end{aligned}$$

since the boundary term vanishes. So instead of the PDE above, we can also consider the problem: find $u \in V$ such that for all $v \in W$ we have

$$a(u, v) := \int_\Omega \nabla u \nabla v \,\mathrm{d}x = \int_\Omega f v \,\mathrm{d}x =: \ell(v) \tag{1.3}$$

where V, W are appropriate vector spaces. This is called the weak formulation of the Poisson equation. Here, $V = W = H_0^1(\Omega)$ is a convenient choice, which makes the problem well-posed and is general enough for typical applications. Here, $H_0^1(\Omega)$ is the Sobolev space of functions with bounded weak first derivative and homogeneous Dirichlet boundaries. The space V is sometimes called the ansatz space and W the test space. While there are benefits in allowing $V \neq W$, here we treat only applications where $V = W$ is a Hilbert space. Thus, the mathematical background is developed for this simplified setting.

In this thesis, we consider the weak formulation as the basis for discussion. Strong formulations will be treated as motivating examples only. To justify this, note that the quantities modelled by u and f often arise from averages over infinitely small

volume elements — such as density, pressure or temperature. As such, spaces of (in some sense) integrable functions may indeed be a better fit to the physical reality than spaces of continuous functions. Try, for instance, to make sense of a point-evaluation of temperature, the average kinetic energy of a particle ensemble.

This chapter is structured as follows. Section 1.1 reviews key concepts from functional analysis, focusing on separable Hilbert spaces, bilinear forms, and fundamental results such as the Lax-Milgram Theorem, the BNB Theorem, and the Gårding Lemma. Section 1.2 introduces discretization techniques, in particular continuous finite elements, and recalls standard approximation results. In Section 1.3, several multiscale PDEs are discussed that serve as running examples throughout the thesis — heterogeneous diffusion, linear elasticity, and the Helmholtz equation. These examples help motivate the need for advanced solvers and form the basis for the numerical experiments discussed later. Section 1.4 then provides a brief overview of iterative solvers and domain decomposition methods, emphasizing techniques relevant to this work: the Conjugate Gradient method (CG), the Generalized Minimal Residual method (GMRES), and various two-level Schwarz schemes. The chapter concludes with a discussion of the relationship between approximation spaces and the coarse spaces used in two-level preconditioners. This comparison helps clarify the role of coarse spaces in multiscale methods and highlights the shared intuition between approximation theory and domain decomposition methods.

1.1 Essential Concepts from Functional Analysis

In this chapter, we review fundamental results from functional analysis, focusing on separable Hilbert spaces, which provide the foundation for much of this thesis. Specifically, we cover the representation theorem, the spectral theorem for self-adjoint compact operators, the Lax-Milgram Theorem, the BNB Theorem, and the Gårding Lemma. The spectral theorem for self-adjoint compact operators will be important later for understanding the asymptotic decay of eigenvalues in MS-GFEM. The remaining results are standard tools in PDE theory for finite element applications.

1.1.1 Separable Hilbert Spaces

A *Hilbert space* is a complete topological vector space, whose topology is induced by an inner product. Note that Hilbert spaces are allowed to be infinite dimensional. The goal of this subsection is to highlight their most important properties, and in particular similarities and differences to $\mathbb{R}^n$.

A *topological vector space* is a vector space that also carries a topology such that both structures are compatible. More precisely, vector addition and scalar multiplication are required to be continuous. When considering maps between topological vector spaces we will usually silently assume that they are structure-preserving, i.e. linear and continuous. Note that without further assumptions, a topological vector space can lack very fundamental properties such as local convexity.

Fréchet spaces add more structure to the point where many familiar constructions make sense. They are thus popular as a baseline for functional analytic discussions. Fréchet spaces are complete, locally convex topological vector spaces, whose topology is induced by a translation-invariant metric. This admittedly sounds rather unwieldy. Fortunately, for this thesis, it is sufficient to note that any complete *normed vector space* is a Fréchet space. A normed vector space is, unsurprisingly, a vector space whose topology is induced by a norm. Complete normed vector spaces are called *Banach spaces*. A welcome side benefit of working with norms is that addition and scalar multiplication are automatically continuous.

Finally, we are arriving at the definition of a Hilbert space, the mathematical structure of interest in this thesis.

Definition 1.1 (Hilbert spaces)**.** A *Hilbert space* $(V, b(\cdot,\cdot))$ is a Banach space whose norm is induced by an inner product $b(\cdot,\cdot)$. We denote the associated norm by $\|\cdot\|_b$.

In this thesis, we will encounter in particular the following examples of Hilbert spaces.

Example 1.2. Let $\Omega \subset \mathbb{R}^d$ be open. The space $L^2(\Omega)$ of square integrable functions on Ω is a Hilbert space together with the inner product

$$(v, w)_{L^2(\Omega)} := \int_\Omega v(x)w(x)\,\mathrm{d}x.$$

To be precise, to make this product positive definite we need to identify any two functions which agree everywhere except on a set of measure 0. This is usually silently assumed, so whenever we refer to L^2 functions, we actually mean equivalence classes in this sense. We abbreviate $(\cdot,\cdot)_{L^2} = (\cdot,\cdot)_{L^2(\Omega)}$ when there is no risk of confusion.

Example 1.3 (H^1)**.** Let $\Omega \subset \mathbb{R}^d$ open and let $H^1(\Omega) \subset L^2(\Omega)$ be the subspace of square integrable functions with square integrable weak first derivatives. Together

with the product

$$(v, w)_{H^1(\Omega)} := \int_\Omega v(x)w(x)\,\mathrm{d}x + \int_\Omega \nabla v(x) \cdot \nabla w(x)\,\mathrm{d}x$$

this is a Hilbert space. Again, we abbreviate $(\cdot,\cdot)_{H^1} = (\cdot,\cdot)_{H^1(\Omega)}$ when there is no risk of confusion.

Example 1.4. Let $\Omega \subset \mathbb{R}^d$ be open. The Sobolev space $H^2(\Omega) \subset L^2(\Omega)$ consists of functions whose weak derivatives up to second degree are square integrable. That is,

$$H^2(\Omega) := \left\{ v \in L^2(\Omega) \;\middle|\; D^\alpha v \in L^2(\Omega) \text{ for all multi-indices } \alpha \text{ with } |\alpha| \leq 2 \right\}.$$

Together with the inner product

$$(v, w)_{H^2(\Omega)} := \sum_{|\alpha|\leq 2} \int_\Omega D^\alpha v(x)\, D^\alpha w(x)\, dx,$$

this space is a Hilbert space. We again write $(\cdot,\cdot)_{H^2}$ for the inner product when the domain Ω is clear from context.

When different inner products are in play, it is important to clearly distinguish between them. The following two examples highlight this.

Example 1.5. The space $(H^1(\Omega), (\cdot,\cdot)_{L^2})$ is not complete and thus not a Hilbert space.

Example 1.6. Let $a(\cdot,\cdot)$ be the bilinear form arising from the diffusion equation defined in (1.3). Then $(H^1, a(\cdot,\cdot))$ is not a Hilbert space, because $a(\cdot,\cdot)$ is only positive semi-definite. To see this, note that any constant function lies in its kernel. Adding boundary conditions can fix this. Consider for instance

$$H^1_0(\Omega) = \overline{C_c(\Omega)},$$

the closure in $H^1(\Omega)$ of the space of continuous functions on Ω with compact support. Then $(H^1_0(\Omega), a(\cdot,\cdot))$ is a Hilbert space. This is a consequence of the Poincaré inequality and the Lax-Milgram Theorem which we state below, Theorem 1.28. In the applications of this thesis, we only consider domains Ω with Lipschitz boundaries, and for these cases rather define $H^1_0(\Omega) \subset H(\Omega)$ to be the subspace of functions with homogeneous Dirichlet boundary conditions, i.e. with zero trace on $\partial\Omega$. For a detailed treatment of traces, see [27, Chapter 3].

Hilbert spaces carry many properties that we are familiar with from working with $\mathbb{R}^n$. We will make use of them throughout this thesis, often without explicit mention. First, note that a Hilbert space is canonically isomorphic to its dual, as formulated in the representation theorem.

Theorem 1.7 (Riesz Representation Theorem)**.** *A Hilbert space $(V, b(\cdot,\cdot))$ is isomorphic to its dual through the canonical isometry*

$$\begin{aligned} V &\to V^* \\ v &\mapsto b(v,\cdot). \end{aligned}$$

By using this result in L^2, we can now easily show that the "for all v"-statement in the weak formulation of the Poisson equation (1.3) is enough to imply the strong

formulation. To elaborate, assume that $\Delta u, f \in L^2$. Then we have

$$\begin{aligned} &\int_\Omega -\Delta u v \,\mathrm{d}x = \int_\Omega f v \,\mathrm{d}x \text{ for all } v \in L^2 \\ \iff\ &(\Delta u, \cdot)_{L^2} = (f, \cdot)_{L^2} \\ \iff\ &\Delta u = f. \end{aligned}$$

Another familiar notion that works very well in Hilbert spaces are orthogonal projections. The only issue in infinite dimensions is that subspaces are not necessarily closed. This could lead to them not being complete and thus the very point we want to project on could be missing. Consequently, one has to add closedness as an assumption.

Proposition 1.8. *Any closed subspace of a Hilbert space is again a Hilbert space with the inherited inner product.*

Definition 1.9 (Orthogonal projection, orthogonal complement, orthogonal)**.** Let $(V, b(\cdot,\cdot))$ be a Hilbert space and $W \subset V$ closed. Then we call the map $\pi_W : V \to W$ defined by

$$b(\pi_W(v), w) = b(v, w) \quad \text{for all } w \in W$$

the *b-orthogonal projection* onto W. Further, we call the set

$$W^{\perp_b} := \{v \in V \mid b(v, w) = 0 \text{ for all } w \in W\}$$

the *b-orthogonal complement* of W. Two vectors $v, w \in V$ are called *b-orthogonal* if $b(v, w) = 0$.

Remark 1.10. The orthogonal projection above is well-defined by the Riesz representation theorem. Moreover, $W^{\perp_b}$ is a closed subspace of V as the kernel of the continuous map

$$\begin{aligned} V &\to W^*, \\ v &\mapsto b(v, \cdot), \end{aligned}$$

and thus $W^{\perp_b}$ is a Hilbert space.

Given the notion of orthogonality, it seems natural to ask whether Hilbert spaces have orthogonal bases. They do indeed. Moreover, a given Hilbert space is separable if and only if one (and then all) orthonormal bases are countable.

Definition 1.11 (Orthonormal basis)**.** Let V be a Hilbert space. An orthonormal system for V is a set of pairwise orthogonal vectors $\{e_i\}_{i \in N}$ of norm 1 such that their linear hull is dense in V. An orthonormal system is called an *orthonormal basis* if it is linearly independent.

Proposition 1.12. *Any Hilbert space has an orthonormal basis. A Hilbert space is separable (as a topological space) if and only if it has a countable orthonormal basis.*

To sum up, one can in many ways work with a Hilbert space V as with $\mathbb{R}^n$. V is isomorphic to its dual just as row vectors are dual to column vectors. Orthogonal projections behave as expected, and we can write any vector as a weighted sum of orthogonal basis vectors. Despite all this, there are important properties missing. For instance, infinite dimensional Hilbert spaces are never locally compact.[1] This

[1] To be precise, a Hausdorff topological vector space is locally compact if and only if it is finite dimensional.

means for instance that we cannot expect bounded sequences to have converging sub-sequences in a general Hilbert space. Thus, despite the similarities to the finite dimensional setting, one still has to be careful when generalizing notions from linear algebra and calculus to this setting.

1.1.2 Continuous Linear Maps

Whenever considering sets with structure in mathematics, studying the corresponding structure-preserving maps is of key importance. Here we thus look at the properties of continuous linear maps. Again we will focus on differences and similarities to the finite dimensional case. The first thing to note is that a linear map between topological vector spaces is not automatically continuous. Luckily, there is still an easy characterization of continuous maps between normed vector spaces.

Proposition 1.13. *A linear map between normed vector spaces is continuous if and only if it is bounded. Bounded here means that the supremum in* (1.4) *below is finite.*

A typical thing to do when considering structure preserving maps is to investigate sets of such maps. Often these sets themselves carry the very same structure they are preserving. This is also the case here: continuous linear maps between normed vector spaces form a normed vector space themselves. This generalizes the finite-dimensional result that the space of matrices forms a normed vector space under the spectral norm.

Definition 1.14. Let $(V, \|\cdot\|_V), (W, \|\cdot\|_W)$ be normed vector spaces. The vector space of all continuous linear maps $\mathcal{B}(V, W)$ from V to W is, unless otherwise noted, equipped with the norm

$$\|A\| := \sup_{\|v\|_V = 1} \|Av\|_W, \tag{1.4}$$

where $A \in \mathcal{B}(V, W)$.

Note that to be fully precise, we would have to write

$$\mathcal{B}(V, W) = \mathcal{B}((V, \|\cdot\|_V), (W, \|\cdot\|_W)),$$

but the norms are often suppressed in the notation. A question that comes to mind when considering this generalization of the spectral matrix norm, is whether the concept of eigenvalues holds up in infinite dimensions. In principle, it does; however there are a few subtleties to consider. For once, the two following definitions are not equivalent any more.

Definition 1.15 (Spectrum)**.** The *spectrum* of a continuous linear map $T : V \to V$, acting on a normed vector space V, is the set of all complex numbers λ for which $\lambda I - T$ does not have a continuous inverse.

Definition 1.16 (Eigenvalues)**.** The *eigenvalues* of a linear map $T : V \to V$, acting on a normed vector space V, is the set of all complex numbers λ for which there is a $v \in V$ such that

$$Tv = \lambda v.$$

Such a v is called an eigenvector of T.

As an example, consider a Hilbert space V with an orthonormal basis $\{e_n\}_n \in \mathbb{N}$ and the continuous linear map $T : V \to V$ with $T(e_n) := 1/n e_n$. The eigenvalues of this map are given by the set $\{1/n\}_n$ while the spectrum is given by $\{1/n\}_n \cup \{0\}$. This T can also serve as an example of a compact linear map, the next definition.

Definition 1.17 (Compact linear map)**.** A linear map $T : V \to W$ between normed vector spaces is called *compact* if the image of the unit ball is precompact in W. A subset of a topological space is called precompact if its closure is compact.

It is easy to see that any compact linear map must be bounded, and thus continuous. Also, compact linear maps behave very nicely under composition with continuous linear maps.

Proposition 1.18. *Let U, V, W be normed vector spaces and $S : U \to V$, $T : V \to W$ continuous linear maps. Then, if either S or T are compact, so is $T \circ S$.*

Note that any linear map with finite image is compact. In fact, a good intuition for compact linear maps is that their image can be approximated with a finite dimensional vector space, which is a consequence of the theorem below. First however, we need the notion of an adjoint.

Proposition 1.19 (Adjoint)**.** *Let $(V, a(\cdot,\cdot))$, $(W, b(\cdot,\cdot))$ be Hilbert spaces and $T : V \to W$ a continuous linear map. Then there exists a unique continuous linear map $T^* : W \to V$ such that*

$$b(Tv, w) = a(v, T^*w) \text{ for all } v \in V, w \in W.$$

This map is bounded and linear, and called the adjoint *of T.*

Proposition 1.20. *Let $(V, a(\cdot,\cdot))$, $(W, b(\cdot,\cdot))$ be Hilbert spaces and $T : V \to W$ a continuous linear map. Then the nullspace of its adjoint T^* is given by the orthogonal complement of the image of T, i.e.*

$$\ker(T^*) = \mathit{Image}(T)^{\perp}.$$

Definition 1.21 (Self-adjoint, symmetric)**.** Let $(V, b(\cdot,\cdot))$ be a Hilbert space and $T : V \to V$ a continuous linear map. Then T is called *self-adjoint*, or *symmetric* (with respect to b), if $T = T^*$.

Theorem 1.22. *Let T be a compact, self-adjoint, linear map acting on a Hilbert space. Then the orthogonal complement of the nullspace of T admits a countable orthonormal system of eigenvectors of T. The eigenvalues of T are bounded and have 0 as their only accumulation point (if any).*

The above theorem gives a tangible characterization condition for compactness of a linear map $V \to V$. To get a similar result for maps $V \to W$ we need an analogue to the singular values of a rectangular matrix. For Hilbert spaces this is easy to define.

Definition 1.23 (Singular values)**.** The *singular values* of a linear map $T : V \to W$ between Hilbert spaces, is the set of the square roots of the eigenvalues of T^*T, where T^* is the adjoint of T.[2]

With this definition we can formulate the following analogue of Theorem 1.22 for maps $T : V \to W$.

Theorem 1.24. *A continuous linear map $T : V \to W$ between Hilbert spaces is compact if and only if T^*T is compact.*

Compactness of linear maps will play a role in spectral methods later, when we set up eigenproblems that are supposed to single out the most important functions for a given task. The idea there is to construct an operator T such that the important functions of interest are eigenfunctions of T with large eigenvalues, say larger than ϵ. This is only practically useful however, if only finitely many such eigenvalues exist. If T is compact, this will be true for any $\epsilon > 0$.

[2]The notion of singular value can be generalized beyond Hilbert spaces. Without the availability of an adjoint, the definition becomes somewhat cumbersome, however.

1.1.3 Bilinear Forms

A *bilinear form* on vector spaces V, W is a map $a : V \times W \to \mathbb{R}$ which is linear in each argument. If $V = W$ we simply say that a is a bilinear form on V. We will often write $a(\cdot, \cdot)$ to better distinguish from the indefinite article a.

The weak formulation of a PDE is typically phrased in terms of a bilinear form, i.e. find u in a Hilbert space V such that

$$a(u, \cdot) = \ell(\cdot)$$

for a bilinear form $a(\cdot, \cdot)$ and a linear form ℓ on V. In other words, the problem is lifted to the dual space of V. From an abstract viewpoint this is unnecessary here, since in our setting V is canonically isomorphic to V^* by the representation theorem, and thus going to the dual space cannot give any theoretical benefit. In fact, Proposition 1.26 below makes that precise by showing that for any continuous bilinear form $a(\cdot, \cdot)$ on V there is an associated continuous linear map $T : V \to V$. If this linear map is invertible, the problem is well-posed.

We keep the historically grown bilinear form notation, however, and so the goal of this chapter is to develop tools that relate properties of a given bilinear form $a(\cdot, \cdot)$ to invertibility of the associated linear map T. Since we are working in topological vector spaces, the first thing to look at are, of course, characterizations of continuity.

Proposition 1.25. *Let V, W be topological vector spaces and $a(\cdot, \cdot)$ a bilinear form on $V \times W$. Then the following statements are equivalent.*

(i) a is continuous,

(ii) a is continuous in each argument individually,

(iii) a is bounded, i.e. there exists a $\Lambda \in \mathbb{R}$ such that $|a(v, w)| \leq \lambda \|v\| \|w\|$ for all $v \in V, w \in W$.

As mentioned above, bilinear forms can be represented by linear maps as in the finite dimensional case. Recall that for any bilinear form a on $\mathbb{R}^n$, there is a matrix $\boldsymbol{A}$ such that $a(u, v) = (\boldsymbol{A}u, v)_2$. This indeed works similarly over general Hilbert spaces.

Proposition 1.26. *Let $a : V \times V \to \mathbb{R}$ be a continuous bilinear form on a Hilbert space V. Then there is a continuous linear map $T : V \to V$ such that $a(v, w) = (Tv, w)$ for all $v, w \in V$. We say that T is* associated *to a.*

So the problem: "given $f \in V$, find $u \in V$ such that $a(u, v) = (f, v)$ for all $v \in V$", is well-posed if and only if the associated linear map has a continuous inverse, since then simply $u = T^{-1}f$. A sufficient condition for this is that the bilinear form is coercive, as is stated in the Lax-Milgram Theorem.

Definition 1.27 (Coercive bilinear form)**.** A bilinear form on a normed vector space V is called coercive if it is bounded from below in the following sense: there exists some $\alpha > 0$ such that $a(v, v) \geq \alpha \|v\|^2$ for all $v \in V$.

Theorem 1.28 (Lax-Milgram)**.** *Let $a(\cdot, \cdot)$ be a continuous and coercive bilinear form on a Hilbert space $(V, (\cdot, \cdot))$. Then the linear map T associated to a is unique, invertible, and it holds that $\|T^{-1}\| \leq 1/\alpha$ with α the bound from below in the coercivity of a.*

Proof. This is a corollary of the BNB Theorem below. □

To get a bit of an intuition for the Lax-Milgram Theorem, consider the special case of a being symmetric. In this case a is an inner product, and since a is assumed to be coercive, its induced norm is equivalent to the existing norm on V. Thus, $(V, a(\cdot,\cdot))$ is complete, and in particular a Hilbert space. The statement on the linear map T associated to a is then essentially the representation theorem. Lax-Milgram generalizes this to the non-symmetric case.

Unfortunately, many bilinear forms in practical applications are not coercive. Take for instance the weak formulation of the Helmholtz equation. The following theorem offers a more conclusive if and only if statement to treat such cases. It comes at the cost of somewhat more obscure conditions, however.

Theorem 1.29 (Banach–Necas–Babuska (BNB))**.** *Let $(V, (\cdot,\cdot))$ be a Hilbert space, $a(\cdot,\cdot)$ a continuous bilinear form, and ℓ a continuous linear form on V. Then the linear map T associated to a is unique, invertible, and it holds that $\|T^{-1}\| \leq 1/\alpha$ if and only if*

(i) $\inf_{v\in V} \sup_{w\in V} \frac{a(v,w)}{\|v\|\|w\|} =: \alpha > 0$, *and*

(ii) for any $w \in V$ we have that $a(\cdot, w) = 0$ implies $w = 0$.

Proof. This is a special case of Theorem 25.9 in [28]. We give a sketch of a proof here. To make a bit more sense of the inf-sup condition, note that it equates to saying that the linear map $V \to V'$, $v \mapsto a(v,\cdot)$ is bounded from below in a certain sense. But then so is T, since we can write it as

$$T : v \mapsto a(v,\cdot) = (Tv,\cdot) \mapsto Tv$$

and the second arrow is an isometry. Now note that a continuous linear map between Banach spaces is bounded from below if and only if it is injective and its image closed.

From assumption (ii) on the other hand we get that T^* is injective since $a(\cdot, w) = (\cdot, T^*w)$. And with that we use the fact that a continuous linear map between Banach spaces which has closed image and an injective adjoint, is also surjective.

So T is injective and surjective, thus bijective. The upper bound on $\|T^{-1}\|$ finally is a consequence of T being bounded from below. □

While the BNB Theorem is more generally applicable than Lax-Milgram, verifying the inf-sup condition can be difficult in practice. The following lemma offers a tool to this end for the Helmholtz equations, see also Section 1.3.3.

Lemma 1.30 (Gårding)**.** *Let $(V, (\cdot,\cdot)_V), (L, (\cdot,\cdot)_L)$ be Hilbert spaces and $V \hookrightarrow L$ an injective, compact linear map. Let $a(\cdot,\cdot)$ be a continuous bilinear form on V and assume that there are $\beta, \gamma > 0$ such that*

(i) $|a(v,v)| + \beta(v,v)_L \geq \gamma(v,v)_V$ *for all $v \in V$, and*

(ii) if $a(v,w) = 0$ for all $w \in V$, then $v = 0$.

Then there is an $\alpha > 0$ such that $\inf_{v\in V} \sup_{w\in V} \frac{|a(v,w)|}{\|v\|_V\|w\|_V} \geq \alpha$.

Proof. [28, Lemma 35.3]. □

To sum up, we have seen the Lax-Milgram Theorem which yields well-posedness for coercive problems, then the more general BNB Theorem, and finally the Gårding Lemma, a tool to verify the assumptions of the BNB Theorem in specific applications. We will use these tools to show well-posedness of the model problems introduced in Section 1.3.

1.2 Discretization with Finite Elements

Throughout this section, we consider an infinite-dimensional Hilbert space V and are interested in the solution $u \in V$ of

$$a(u, v) = \ell(v) \quad \text{for all } v \in V, \tag{1.5}$$

with a bilinear form a and a linear form ℓ on V. In practice, we will have to make do with finite-dimensional approximations. In this exposition we focus on conforming methods while Discontinuous Galerkin elements will be introduced later when needed. Conforming discretizations use a sequence of subspaces $\{W_n\}_{n\in\mathbb{N}}$ with $W_n \subset V$, and approximate solutions $u_n \in W_n$. In principle, we can define W_n and u_n as we like. The aim is to obtain $u_n \to u$, preferably with a known rate of convergence. Achieving this can be split into two separate tasks. First, construct the spaces W_n such that

$$\inf_{v\in W_n} \|u - v\|_V \to 0 \quad \text{as} \quad n \to \infty \tag{1.6}$$

in an appropriate norm. Second, define the $u_n \in W_n$ such that there is a $C > 0$ with

$$\|u_n - u\|_V \leq C \inf_{v\in W_n} \|u - v\|_V \tag{1.7}$$

for all n.

Suppose for the moment that we have found W_n's that satisfy (1.6). A reasonable choice of the $u_n \in W_n$ then seems to be the solution of the restricted problem

$$a(u_n, v) = \ell(v) \quad \text{for all } v \in W_n.$$

This choice of u_n is known as the Galerkin approximation. Under suitable assumptions, it can be shown to satisfy the error estimate in (1.7). When W_n is finite-dimensional, computing u_n reduces to solving a linear system.

What remains to settle now is the choice of the W_n. A theoretically ideal choice would be the linear span of the most important eigenfunctions of the given problem. A famous example are Fourier modes for the Laplace equation. Unfortunately, this approach is only feasible when the eigenfunctions are already known. Otherwise, we would have to numerically compute approximations for them, which leaves us essentially where we started. Still, the general idea of using eigenfunctions is a powerful concept. When we discuss spectral methods in Chapter 2 we will find that they employ this idea locally, in a domain decomposition setting.

Since, for most problems, the theoretically optimal choice of using global eigenfunctions is not feasible, one often looks in the opposite direction: towards spaces that are uninformed about the problem, but straightforward to construct, such as finite element spaces on a uniform mesh. We provide a brief review of continuous piecewise polynomial finite elements in Section 1.2.2. A key advantage of finite elements, besides being inexpensive to set up, is that the resulting linear system is sparse and thus solvable with complexity $O(n)$, where n is the dimension of the linear system. This seemingly linear complexity is a bit misleading, however, since to guarantee a good approximation a small enough mesh size h is needed, and on a structured mesh in dimension d one has $n \propto h^{-d}$ for typical discretizations. Consequently, real world applications often lead to huge linear systems which challenge current and foreseeable computation technology.

To alleviate the issue of too large n, various ways have been proposed to construct spaces which are at least somewhat adapted to the problem. One idea is to use

finite elements on an informed mesh that has smaller cells in regions of the domain where the given problem is particularly difficult (h-refinement). Such a mesh is typically constructed iteratively, starting from a coarse mesh and refining in regions where the local error is largest. As an alternative, one can allow higher polynomial degrees in the difficult regions (p-refinement), or combine both these approaches (hp-refinement). These methods have been studied intensely in the 1990s but have not seen a breakthrough to practical fields due to high computational cost and the fact that not all difficulties can be seen on a coarse mesh, on which the method is initiated.

Without an adaptive mesh, we are left with large linear systems that are too extensive to solve on a single computer. Thus, the problem may have to be parallelized. Parallelization can be employed on the level of solvers, i.e. after W_n is defined, which will be treated in Section 1.4. Alternatively, one can parallelize the construction of W_n itself. Combining both approaches leads to methods such as GenEO or MS-GFEM, which solve local eigenproblems in parallel to construct an informed coarse space for a parallel linear solver.

1.2.1 Galerkin Approximation

In this subsection, we assume to have decided on a closed subspace $W \subset V$ to approximate the solution u of (1.5). To this end we investigate the following problem. Find $w \in W$ such that

$$a(w, v) = \ell(v) \quad \text{for all } v \in W. \tag{1.8}$$

This is sometimes called the Galerkin approximation. We aim to show that, under suitable assumptions, w is well-defined and provides a good approximation to $u \in V$. More precisely, we would like to have

$$\|w - u\| \leq C \inf_{v \in W} \|v - u\|, \tag{1.9}$$

i.e. that w is the best approximation to u in W, up to a multiplicative constant. Such a result is often called *quasi optimality* of the approximate solution w. As usual, the goal is to keep this constant independent of parameters such as the mesh size h.

We begin with the case where a is coercive. For well-posedness note that since W is closed, it is itself a Hilbert space. Thus, w is well-defined for the same reason that u is, namely by the Lax-Milgram Theorem.

Suppose next that a is also symmetric. In this case, w is the a-orthogonal projection of u into W, and thus satisfies (1.9) with $C = 1$ in the $\|\cdot\|_a$-norm. Moreover, the a-norm is equivalent to the norm on V, since $a(\cdot, \cdot)$ is bounded from below and above. Thus, we can get (1.9) also in the desired V-norm if we account for the norm equivalence constants.

Cea's Lemma below states this in detail and more generality: we can drop symmetry, and still get a similar result.

Lemma 1.31 (Cea's Lemma)**.** *Let V be a Hilbert space and $a(\cdot, \cdot)$ a continuous, coercive bilinear form on V, i.e. there are constants $\alpha, \beta > 0$ such that*

$$\begin{aligned} a(v, v) &\geq \alpha \|v\|^2 \quad \textit{for all } v \in V \textit{ (coercive)}, \\ |a(v_1, v_2)| &\leq \beta \|v_1\| \|v_2\| \quad \textit{for all } v_1, v_2 \in V \textit{ (continuous)}. \end{aligned}$$

Let further $W \subset V$ be a closed subspace. Then for u, the solution to (1.5) and w, the solution to the restricted version (1.8), we have

$$\|u - w\| \leq \left(\frac{\beta}{\alpha}\right) \inf_{v \in W} \|u - v\|.$$

If $a(\cdot, \cdot)$ is additionally symmetric, we even get the stronger result

$$\|u - w\| \leq \left(\frac{\beta}{\alpha}\right)^{1/2} \inf_{v \in W} \|u - v\|.$$

Proof. See [28], Lemma 26.13. □

In the non-coercive case, the situation becomes more subtle. The inf-sup condition from the BNB Theorem does not naturally carry over to subspaces, nor do the assumptions of the Gårding Lemma. Nevertheless, if the inf-sup condition happens to hold for a specific subspace $W \subset V$, then a quasi-optimality result can still be obtained, as stated in the following lemma.

Lemma 1.32 (Xu–Zikatanov)**.** *Let V be a Hilbert space and $a(\cdot, \cdot)$ a continuous bilinear form on V, i.e. there exists $\beta > 0$ such that*

$$|a(v_1, v_2)| \leq \beta \|v_1\| \|v_2\| \quad \text{for all } v_1, v_2 \in V.$$

Let further $W \subset V$ be a closed subspace with

$$\inf_{v \in W} \sup_{w \in W} \frac{a(v, w)}{\|v\| \|w\|} =: \alpha > 0.$$

Then for u, the solution to (1.5) and w, the solution to the restricted version (1.8), we have

$$\|u - w\| \leq \left(\frac{\beta}{\alpha}\right) \inf_{v \in W} \|u - v\|.$$

Proof. See [28], Lemma 26.18. □

Note that the restricted inf-sup condition in Xu–Zikatanov depends on the specific choice of W as well as on the given bilinear form $a(\cdot, \cdot)$. It thus has to be checked for each such combination individually.

1.2.2 Conforming Finite Elements

Here we discuss finite elements, a popular choice for a family of finite-dimensional subspaces $W_n \subset V$ that we will use throughout this thesis. Over the last decades, finite elements have become the go-to discretization method for PDE problems in many fields. In principle, they are simply piecewise polynomial functions. For practical use however, there are many important details to consider. These include the nature of the underlying mesh, the choice of basis, and whether discontinuities between elements are permitted.

In this subsection, we focus on the special case of continuous piecewise linear functions defined over a simplicial mesh, which will serve as a running example throughout the thesis. All the discussed methods are applicable to a wider range of discretizations. Phrasing them in a more general setting can be tedious, however. We will thus stick to the special case for now and later address the generalization of

GenEO to the Discontinuous Galerkin method in Section 3.2.2 as well as an abstract version of MS-GFEM for general Hilbert spaces in Section 6.2.

To define simplicial finite elements, we first need the notion of a simplex. At its core, a *k-simplex* K is simply a set containing $k+1$ elements. In our context, we only consider $K \subset \mathbb{R}^d$ and denote by $|K|$ the convex hull of K. If we want to be precise, we call $|K|$ the geometric realization of K, but we will mostly suppress this distinction and identify K and $|K|$.

An i-dimensional *face* of K is any subset of K containing $i+1$ points. Note that the faces of K are thus simplices themselves. The faces of dimension zero of K are exactly the elements of K, and are often called *vertices*. Further, a simplex K of dimension k is called *nondegenerate* if the k-dimensional volume of its geometric realization is non-zero. If K is nondegenerate, then so are all of its faces.

We call a collection of nondegenerate k-simplices $\mathcal{T}$ matching if the intersection of any $K_1, K_2 \in \mathcal{T}$ is either empty, or a $k-1$ dimensional face of both K_1 and K_2. We call such a $\mathcal{T}$ a *matching simplicial mesh* of dimension k.

For the following, we assume that our open and connected domain $\Omega \subset \mathbb{R}^d$ is resolved by a matching simplicial mesh $\mathcal{T}$, i.e. that we have

$$\bar{\Omega} = |\mathcal{T}| := \bigcup_{K \in \mathcal{T}} |K|.$$

A property is said to hold *piecewise* (on Ω) if it holds on each $|K| \in \mathcal{T}$ individually.

The space of *conforming finite elements of degree 1* on $\mathcal{T}$ is now simply the space of piecewise linear, (globally) continuous functions, denoted by W in this subsection. Note that a linear function on a k-dimensional simplex is uniquely determined by its value in $k+1$ points, for instance by the values on its vertices. Thus, we can construct a basis of W as follows. Let $P = p_1, \dots, p_N$ denote the set of all vertices in the mesh $\mathcal{T}$. Further, for any $i = 1, \dots, N$ consider the unique piecewise linear function θ_i that satisfies

$$\theta_i(p_j) = \delta_{ij}$$

and note that these θ_i are continuous (not just piecewise continuous). They form a basis of the N-dimensional space W and are often called *Lagrange basis* or hat functions.

Now recall that we want a convergence result like (1.6). For that we will need to define a whole family of subspaces $\{W_n\}_n$. The idea is, of course, to use finite element spaces over a suitable family of meshes $\{\mathcal{T}_n\}_n$. Obviously, these meshes should become finer as $n \to \infty$, but there is also the more subtle requirement of shape regularity. Essentially, the angles in the mesh have to be bounded away from zero. The following two definitions formalize this.

Definition 1.33 (Diameter)**.** Let K be a k-simplex. We denote by h_K the (outer) diameter of $|K|$, more precisely

$$h_K := \max_{x,y \in |K|} \|x - y\|_2.$$

Further, by ρ_K we denote the inner diameter of $|K|$, that is the diameter of the largest ball that can be contained in $|K|$.

Definition 1.34 (Shape regularity)**.** Let $\{\mathcal{T}_n\}_{n \in \mathbb{N}}$ be a family of matching simplicial meshes with $|\mathcal{T}_n| = \bar{\Omega}$ for all n. We call this family *shape regular* if there is a bound

$\theta \in \mathbb{R}$ such that

$$\frac{h_K}{\rho_K} \leq \theta \text{ for all } K \in \mathcal{T}_n \text{ for all } n \in \mathbb{N}.$$

When formulating approximation results for finite elements over a family of shape regular meshes, the only parameter that cannot be accounted for with a uniform constant is the mesh size. It is thus common to use the following abbreviations.

Definition 1.35. The mesh size h of a given mesh $\mathcal{T}$ is defined as the maximal (outer) diameter of all mesh elements

$$h := h_{\mathcal{T}} := \max_{K \in \mathcal{T}} h_K.$$

Notation 1.36. In the following we will often index families of shape regular meshes by their decreasing mesh size. More precisely, when we write $\{\mathcal{T}_h\}_{h \in \mathcal{H}}$ we silently assume that $\mathcal{H} = \{h_n\}_{n \in \mathbb{N}}$ is a strictly decreasing sequence in $\mathbb{R}_+$ and that $h_{\mathcal{T}_{h_n}} = h_n$. Further, when considering the limit $h \to 0$ we more precisely mean the limit $n \to \infty$ (and thus $h_n \to 0$).

Moreover, when we write V_h and do not specify this further, we mean the space of conforming finite elements of degree 1 over a mesh $\mathcal{T}_h$ which belongs to a shape regular family of meshes as above.

Thus equipped, we can formulate asymptotic approximation properties for the limit $h \to 0$. Before we state the result, let us sketch how to get there.

For the first ingredient, recall that Taylor's Theorem states that polynomials of degree k can approximate $k+1$-times differentiable functions with order k. The *Bramble-Hilbert Lemmas* (see [27], Section 11.3) essentially state that this also works in the setting of weak derivates. In particular, linear functions approximate H^2 functions with order 1.

This settles the question of approximability locally, i.e. on individual mesh elements. What remains to show is that the result carries over to a corresponding global statement. To prove this, one typically investigates a specific interpolation operator

$$\mathcal{I} : V \to V_h$$

and shows that $\|v - \mathcal{I}v\|$ is small in an appropriate sense. As an example for $\mathcal{I}$, consider the *Lagrange interpolation operator* that takes a continuous function f to the finite element functions of degree 1 that agrees with f in all vertices of the given mesh.

Theorem 1.37. *Let $\{\mathcal{T}_h\}_{h \in \mathcal{H}}$ be a family of shape regular simplicial meshes that cover Ω exactly, i.e.*

$$|\mathcal{T}_h| = \bar{\Omega}$$

for all $h \in \mathcal{H}$. For any $h \in \mathcal{H}$ let further V_h be the space of conforming finite elements of degree one on $\mathcal{T}_h$ and let $v \in H^2(\Omega)$. Then we have

$$\inf_{v_h \in V_h} \|v - v_h\|_{H^1(\Omega)} \leq ch|v|_{H^2(\Omega)}.$$

Proof. This is a special case of [27], Corollary 19.8. □

Notably, the result only holds when we approximate a function in H^2, even though we formulate our problem in H^1. Since H^2 is dense in H^1, one can deduce from the

above theorem that also for $v \in H^1$, we have

$$\inf_{v_h \in V_h} \|v - v_h\|_V \to 0 \quad \text{as} \quad h \to 0. \tag{1.10}$$

However, in this procedure the guaranteed rate of convergence is lost.

To sum up, we have so far defined V_h, the space of conforming finite elements of degree 1 on a family of simplicial meshes and stated their approximation property for H^2 functions in the limit $h \to 0$. If we use them to solve (1.5) and $a(\cdot, \cdot)$ is coercive we can use Cea's Lemma to show that the Galerkin approximations $u_h \in V_h$, i.e. the solutions to the restricted problems

$$a(u_h, v) = \ell(v) \text{ for all } v \in V_h,$$

will give a sequence converging to u with order 1. If $a(\cdot, \cdot)$ is not coercive, we will have to check the inf-sup condition for Lemma 1.32, possibly making additional assumptions, to obtain the same result.

The conforming linear elements described above form just one instance of a broader class of finite element methods. We now outline several generalizations. First, note that so far we only discussed so-called affine meshes. To generalize them, consider defining a (generalized) simplex K to be a diffeomorphism from a reference element Σ, say $\Sigma = |\{0, e_1, ..., e_d\}| \subset \mathbb{R}$, to a subset of Ω, i.e.

$$K : |\Sigma| \to |K| \subset \Omega.$$

Then everything works as above as long as we keep track of the derivatives of the K. If all K are affine-linear maps, the mesh is called *affine*, and we recover our case from earlier. An obvious benefit of allowing non-affine K is that it allows curved meshes which can for instance resolve a domain with curved boundaries exactly. This becomes relevant if the error from approximating the domain boundary with a piecewise linear mesh (typically of order h) is the dominating error, such as for finite elements of higher degree.

Even for affine meshes, the notion of a reference element is a popular approach, as it can simplify mathematical notation, as well as the corresponding implementation. Definitions and calculations only have to be written down for the reference element and can then simply be pushed to the actual mesh element via the diffeomorphism K.

As an example of this advantage, consider another generalization: replacing simplices with polygonals of higher vertex count. This runs into some annoying issues if trying to define them directly. Recall that the geometric realization of a simplex is simply the convex hull of its vertices and its faces are exactly all subsets of vertices. These strategies do not have a simple analogue for, say, a quadrilateral. For instance, the convex hull of 2^d points in $\mathbb{R}^d$ does not necessarily have all of these points on its boundary. With a reference element on the other hand, working out these issues is easy. Just define quadrilaterals as diffeomorphisms starting in the unit cube with their geometric realization being the image. Faces are then simply images of faces of the unit cube, which can easily be written down explicitly if needed. Further, and perhaps most importantly, the reference element greatly facilitates writing down a basis for piecewise polynomial spaces, as we demonstrate in the following example.

Example 1.38 (Quadrilateral finite elements of degree 1)**.** Let $Q \subset \mathbb{R}[X_1, ..., X_d]$ be the set of polynomials in d variables with absolute degree 1, that is

$$Q := \{q = \sum_{n \in \{0,1\}^d} a_n X^n \mid a_n \in \mathbb{R}\},$$

where we wrote $X^n = \Pi_i X_i^{n_i}$ for short. Note that the X^n give a basis for this space and that for $x \in \{0,1\}^d$, the corners of the unit cube, we have

$$X^n(x) = \begin{cases} 1 & \text{if } x = n \\ 0 & \text{else.} \end{cases}$$

Let now $\mathcal{T}_h$ be a matching mesh of quadrilaterals as outlined above, i.e.

$$\mathcal{T}_h := \{K : [0,1]^d \to |K| \subset \mathbb{R}^d \mid K \text{ is a diffeomorphism}\}$$

and define the function spaces $Q(K) = \{q \circ K^{-1} \mid q \in Q\}$. These can be glued together to a global space of continuous functions by setting

$$Q(\mathcal{T}_h) = \{f \in C^0(|\mathcal{T}_h|) \mid f|_{|K|} \in Q(K)\}.$$

Note that functions in $Q(K)$ are uniquely determined by the values in the vertices $\{p_1, ..., p_N\}$ of $\mathcal{T}_h$. We can thus define Lagrange hat functions similar to the simplicial case. I.e. there are $\theta_i \in Q(\mathcal{T}_h)$ that satisfy

$$\theta_i(p_j) = \delta_{ij}.$$

These give a basis of $Q(\mathcal{T}_h)$, the space of conforming quadrilateral finite elements of degree 1. If the mesh is affine and belongs to a shape-regular family, the approximation result of Theorem 1.37 holds here as well.

Another obvious generalization is to use higher order polynomials, rather than piecewise linears. This will lead to a higher order of convergence in Theorem 1.37, provided the function we are approximating itself is sufficiently regular. A drawback is, however, that the resulting stiffness matrix, i.e. the matrix representing $a(\cdot,\cdot)$ will be less sparse. We refer to [27] for details.

Another popular variant of finite elements is to allow discontinuous jumps between the mesh elements. This has some benefits such as the corresponding solution exactly satisfying conservation of mass in convection-diffusion problems. However, they involve significantly more degrees of freedom, as each discontinuous finite element function can take different values at the same vertex. Moreover, they integrate less naturally into the standard theory, since $V = H^1$ does not include discontinuous functions. As a result, discontinuous finite element spaces are not subspaces of V, which introduces various technical complications.

1.2.3 Basis Choice

Most discretizations come with a choice of a basis in mind, such as continuous finite elements with the Lagrange basis. However, at least from a theoretical perspective, selecting a finite-dimensional subspace and choosing a basis for it are distinct steps. While we expect that the reader is familiar with the concept, we give a formal definition here to clarify notation.

Definition 1.39 (Basis choice)**.** Let V be an n-dimensional vector space. Then we call a linear isomorphism $\Phi : V \to \mathbb{R}^n$ a *basis choice*. We often suppress Φ in the notation and denote $\Phi(v)$ by boldface letters, i.e. $\boldsymbol{v} := \Phi(v)$.

Definition 1.40 (Matrix representation)**.** Let V, W be finite-dimensional vector spaces with chosen bases $\Phi : V \to \mathbb{R}^n$ and $\Psi : W \to \mathbb{R}^m$. For a linear map $M : V \to W$ we

write $\boldsymbol{M}$ for the $m \times n$ matrix representing $\Psi M \Phi^{-1}$. Further, for a bilinear form $a(\cdot, \cdot)$ on $V \times W$ we write $\boldsymbol{A}$ for the matrix representing $a \circ (\Phi^{-1} \times \Psi^{-1})$.

Note that the matrix $\boldsymbol{A}$ in the definition above is uniquely defined by

$$a(v, w) = \boldsymbol{v}^\top \boldsymbol{A} \boldsymbol{w} \text{ for all } (v, w) \in V \times W.$$

Choosing a basis for the problem in (1.20) thus leaves us with the linear system

$$\boldsymbol{v}^\top \boldsymbol{A} \boldsymbol{u} = \boldsymbol{v}^\top \boldsymbol{f} \text{ for all } v \in V,$$

or equivalently

$$\boldsymbol{A} \boldsymbol{u} = \boldsymbol{f}. \tag{1.11}$$

For finite element discretizations the matrix $\boldsymbol{A}$ corresponding to the bilinear form $a(\cdot, \cdot)$ is called *stiffness matrix*. A matrix is called sparse if each row contains at most a fixed number k of non-zero entries, independent of the overall system size n. For such matrices, key operations like storage and matrix-vector multiplication can be performed in $O(nk)$ complexity. If k can be considered constant, this leads to an overall complexity of $O(n)$. We will return to this point in Section 1.4.

1.3 Examples of Multiscale PDE Problems

In this section we introduce three examples of elliptic PDEs that will serve as model problems in this thesis. The focus is on well-defined weak formulations within a Hilbert space setting. For a more rigorous treatment, in particular of the respective boundary conditions, see [28, Chapter 31 and Chapter 35].

Throughout this chapter, $\Omega \subset \mathbb{R}^d$, $d = 2, 3$ will be a regular enough open and bounded domain. A common assumption in that regard is that the boundaries of Ω are locally the graph of a Lipschitz function. For further details see [27].

1.3.1 Diffusion Equation in Groundwater Simulations

Earlier we introduced the homogeneous diffusion equation which is perhaps the simplest second-order PDE. Many applications are complicated by a heterogeneous diffusion coefficient A however, that enters in the following way.

On a given Lipschitz domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$ we now consider the problem of finding a function u such that

$$-\nabla(A\nabla u) = f \quad \text{on } \Omega \tag{1.12}$$

$$u = 0 \quad \text{on } \partial\Omega \tag{1.13}$$

for some $f \in L^2(\Omega)$ and matrix-valued diffusion coefficient (or permeability field) $A \in \left(L^\infty(\Omega)\right)^{d\times d}$. This coefficient A is assumed to be pointwise symmetric and there exist constants $0 < A_{\min} < A_{\max} < \infty$ with

$$A_{\min}|\zeta|^2 \leq A(x)\zeta \cdot \zeta \leq A_{\max}|\zeta|^2, \quad \forall \zeta \in \mathbb{R}^d, \quad x \in \Omega. \tag{1.14}$$

Introducing heterogeneous A barely makes a difference for the theory. The resulting bilinear form remains symmetric and coercive under standard assumptions, just as in the homogeneous case $A \equiv \boldsymbol{I}$. Computationally, however, the simple addition can render the problem almost arbitrarily difficult. This is because the coefficient may contain small-scale features that have a significant effect on the large-scale solution. When using standard finite elements, these features have to be resolved with a fine mesh which quickly leads to infeasibly large linear systems. Resolving this dilemma is the goal of multiscale methods. The simplicity of the homogeneous case makes its heterogeneous counterpart a convenient model problem for multiscale methods. If an algorithm fails here, it is likely due to its inability to handle the problem's multiscale structure, rather than any inherent complexity of the PDE itself.

To arrive at the weak formulation, we use partial integration as for the Poisson problem. Given the linear functional $\ell(v) := \int fv \,\mathrm{d}x$ on $H_0^1(\Omega)$, we seek $u \in H_0^1(\Omega)$ such that

$$a(u, v) := \int_\Omega (A\nabla u) \cdot \nabla v \, dx = \ell(v) \quad \text{for all } v \in H_0^1(\Omega). \tag{1.15}$$

The bilinear form $a(\cdot, \cdot)$ is bounded and coercive, and hence, the above has a unique solution by the Lax-Milgram Theorem.

For the discretization consider a family of shape-regular affine meshes $\{\mathcal{T}_h\}_h$ and let V_h be the corresponding conforming finite element spaces of degree 1. We assume that the mesh-size h is small enough so that the fine-scale details of the coefficient

A are resolved by $\mathcal{T}_h$.[3] The discretization of problem (1.15) is then given by: find $u_h \in V_h$ such that

$$a(u_h, v_h) = \ell(v_h) \quad \text{for all } v_h \in V_h. \tag{1.16}$$

As we mentioned earlier, coercivity also holds on subspaces and so the discrete problem is likewise well-posed by Lax-Milgram. By Theorem 1.37, the best approximation in V_h converges to u with order 1 as $h \to 0$, and by Cea's Lemma so do the solutions to the restricted problem u_h.

1.3.2 Linear Elasticity for Composite Materials

Linear elasticity provides an interesting field of application for multiscale methods: composite materials. These are made of layers with different properties, combined to produce desirable behaviour under mechanical stress.

Carbon fibre materials, for instance, are used in the construction of aircraft wings and other lightweight constructs. They come out of production as thin mats made of fibres in parallel orientation. These are then very durable under stress along their fibres, but not in the other directions. To obtain an overall robust material, several of these mats are glued together with different orientations. The resulting piece is then highly heterogeneous at small scale and this heterogeneity is globally important for the stability of an entire aeroplane wing. A textbook case of a multiscale problem.

The underlying equation is in principle the same as the heterogeneous diffusion equation, only with more dimensions: here we are solving in $V = (H^1(\Omega))^d$. We assume that Ω represents a piece of deformable material that is subjected to an external force $f \in (L^2(\Omega))^d$. The goal is to find the displacement field $u \in V$ under which the material is in equilibrium. We assume that u is small enough that linear elasticity theory applies. Further, we consider a partition of the boundary $\partial\Omega = \partial\Omega_D \cup \partial\Omega_n$ such that $|\partial\Omega_D| > 0$. We assume that the displacement vanishes on $\partial\Omega_D$, i.e. that the material is fixed in place at that part of the boundary. On the rest of the boundary we impose a normal load $g \in (L^2(\partial\Omega_n))^d$.

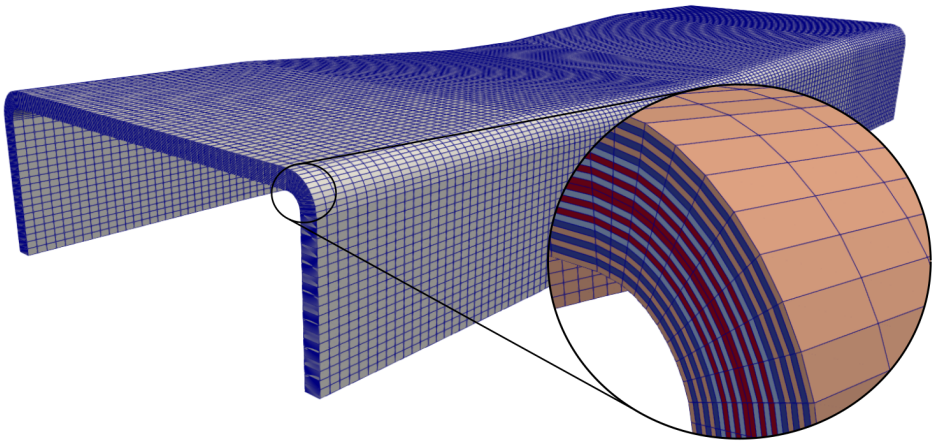


FIGURE 1.1: Model of a C-shaped bar, composed of layers of fibre material, and meshed with hexahedrals.

[3] To make this a bit more precise, we have to be able to evaluate integrals involving A. This is usually done with a quadrature rule, applied to each of the mesh elements individually. Hence, A has to be regular enough within each element for this quadrature to work. Piecewise constant A seems to be the most common case in practice.

To be exact, we are looking for a $u : \Omega \to \mathbb{R}^d$ that satisfies

$$-\frac{1}{2}\nabla\left(\boldsymbol{C}\left(\nabla u + (\nabla u)^\top\right)\right) = f, \tag{1.17}$$

$$u = 0 \quad \text{on } \partial\Omega_D, \tag{1.18}$$

$$\frac{1}{2}\left(\nabla u + (\nabla u)^\top\right)\boldsymbol{n} = g \quad \text{on } \partial\Omega_n, \tag{1.19}$$

where $\boldsymbol{C} \in (L^\infty(\Omega))^{(d\times d)\times(d\times d)}$ is a material property. For each $x \in \Omega$ we view the forth order tensor $\boldsymbol{C}(x)$ as a linear map on matrices, i.e.

$$\boldsymbol{C}(x) : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$$

and assume that $\boldsymbol{C}$ is symmetric, bounded and coercive in a sense we will make precise below.

In the weak formulation we are trying to find $u \in V$ such that

$$a(u, v) = \ell(v) \text{ for all } v \in V$$

with the linear form

$$\ell(v) := \int fv \,\mathrm{d}x + \int_{\partial\Omega_n} gv \,\mathrm{d}x$$

and the bilinear form

$$a(v, w) := \int \frac{1}{2}\left(\boldsymbol{C}\left(\nabla v + (\nabla v)^\top\right)\right) \cdot \left(\nabla w + (\nabla w)^\top\right) \,\mathrm{d}x,$$

where we used the product

$$M \cdot N := \sum_{i,j} M_{ij}N_{ij}$$

for any two matrices $M, N \in \mathbb{R}^{d\times d}$.

What remains is to make the assumed properties of $\boldsymbol{C}$ precise. We call $\boldsymbol{C}$ symmetric if

$$(\boldsymbol{C}(x)M) \cdot N = M \cdot (\boldsymbol{C}(x)N) \qquad \text{for all } x \in \Omega, M, N \in \mathbb{R}^{d\times d}$$

In fact, classical linear elasticity imposes stricter symmetry conditions on $\boldsymbol{C}(x)$, but the definition above is sufficient to ensure symmetry of the bilinear form $a(\cdot,\cdot)$, which is our primary concern. Further symmetries may be useful for optimizing a given implementation, but we will not go further into the details of linear elasticity here. Important for us is to assume that $\boldsymbol{C}$ is bounded and coercive, that is that there are $0 < \alpha < \beta < \infty$ such that

$$\alpha M \cdot M \leq (\boldsymbol{C}(x)M) \cdot M \leq \beta M \cdot M \text{ for all } x \in \Omega, M \in \mathbb{R}^{d\times d}.$$

This second assumption on $\boldsymbol{C}$ is mostly mathematically motivated to make sure that $a(\cdot,\cdot)$ itself is bounded and coercive, and thus the problem well-posed by the Lax-Milgram Theorem.

For the discretization we will again use conforming finite elements of degree 1 on a shape-regular family of meshes that resolve the heterogeneities in $\boldsymbol{C}$. Well-posedness and convergence follows from Lax-Milgram, Theorem 1.37 and Cea's Lemma, just as for the heterogeneous diffusion equation.

1.3.3 Heterogeneous Helmholtz Equation

While for much of this thesis we stay in the coercive definite setting, we will take an outlook to an indefinite case in Section 6.4, where we extend the MS-GFEM preconditioner to the Helmholtz equation. Note that in this case we consider functions spaces with values in $\mathbb{C}$, and we thus consider sesquilinear forms. More precisely, we consider the problem of finding a function u on a Lipschitz domain $\Omega \subset \mathbb{R}^d$ such that

$$\begin{aligned} -\text{div}(A\nabla u) - k^2V^2u &= f, \quad \text{on } \Omega \\ u &= 0, \quad \text{on } \Gamma_D \\ A\nabla u \cdot \boldsymbol{n} - ik\beta u &= g, \quad \text{on } \Gamma_R \end{aligned}$$

where $\partial\Omega = \Gamma_D \cup \Gamma_R$ is a decomposition into Dirichlet and Robin boundaries and with $f \in L^2(\Omega)$, $g \in L^2(\Gamma_R)$, $\beta \in L^\infty(\Gamma_R)$, $\beta > 0$. Further we assume that the coefficients $A \in L^\infty(\Omega)$ and $V \in (L^\infty(\Omega))^{d\times g}$ are strictly positive and bounded from below and above, i.e. (1.14) holds for A, and we have real numbers $0 < V_{\text{max}} \leq V_{\text{min}} < \infty$ such that $V_{\text{min}} \leq V(x) \leq V_{\text{max}}$ for all x.

This equation arises from the wave equation by eliminating time, and is sometimes called the Helmholtz equation in the frequency domain. For the weak formulation we set

$$H^1_{\Gamma_D} := \{v \in H^1(\Omega) \mid v|_{\Gamma_D} = 0\}$$

and find $u \in H^1_{\Gamma_D}$ such that

$$\mathcal{B}(u, v) = F(v) \quad \text{for all } v \in H^1_{\Gamma_D}$$

with

$$\begin{aligned} \mathcal{B}(v, w) &:= \int_\Omega (A\nabla v \cdot \nabla \bar{w} - k^2V^2 v\bar{w})\text{d}x - ik\int_{\Gamma_R} \beta v\bar{w}\text{d}s \\ F(v) &:= \int_{\Gamma_R} g\bar{v}\text{d}s + \int_\Omega f\bar{v}\text{d}x. \end{aligned}$$

The sesquilinear form $\mathcal{B}$ is continuous as the sum of continuous forms and thus for well-posedness we can use the Gårding Lemma (Lemma 1.30) as follows.

Proposition 1.41. *The Helmholtz problem as presented here is well-posed.*

Proof. We have

$$|\mathcal{B}(v, v)|^2 + k^2V_{\text{max}}^2 \int_\Omega k^2V^2 u\bar{v}\text{d}x \geq A_{\text{min}} \int_\Omega \nabla u \cdot \nabla \bar{v}\text{d}x$$

and thus $\mathcal{B}$ fulfill assumption (i) of the Gårding Lemma. We give a sketch of how to show that it also satisfies (ii). Suppose we have $v \in H^1_{\Gamma_D}$ with

$$\mathcal{B}(v, w) = 0 \text{ for all } w \in H^1_{\Gamma_D}.$$

Then, looking only at the imaginary part we must have $|\int_{\Gamma_R} \beta v\bar{w}\text{d}s| = 0$ for all w and thus $v = 0$ on Γ_R. It then follows that $v = 0$ on all of Ω by employing a suitable unique continuation principle. For details see [28], Theorem 35.10. With the assumptions of the Gårding Lemma satisfied, the inf-sup condition holds and well-posedness follows by the BNB Theorem. □

In contrast to the two previous examples there is more to consider here for the discretization. Let again V_h be the space of conforming finite elements over an affine simplicial mesh, belonging to a shape regular family. Then as mentioned in the previous section we have to verify the inf-sup condition again for V_h. We skip the details here, referring to [28], Section 35.4, and merely state the following result.

Theorem 1.42. *There is a constant $c > 0$, independent of the wavenumber k, such that for all $h \leq c/k^2$ the inf-sup condition holds on V_h, i.e.*

$$\inf_{v \in V_h} \sup_{w \in V_h} \frac{\mathcal{B}(v,w)}{\|v\|\|w\|} =: \alpha_{h,k} > 0$$

Further, $\alpha_{h,k}$ can be bounded from below, uniformly in h, but not in k. More precisely there is a constant $c_\alpha > 0$ such that $\alpha_{h,k} \geq c_\alpha / k$.

Proof. See [28, Lemma 35.14 and Example 35.18]. □

So to guarantee a small discretization error $\|u - u_h\|$, the mesh size has to be chosen appropriately small, scaling with k^{-2}. This is often called the *pollution effect*. Still, at least when keeping k fixed, Theorem 1.37 together with Lemma 1.32 yields convergence $u_h \to u$ of order 1 in the limit $h \to 0$.

1.4 Solvers

After choosing a discretization for a PDE problem in weak formulation, we are left with solving a problem of the form: find $u \in V$ such that

$$a(u, v) = (f, v) \text{ for all } v \in V \tag{1.20}$$

in an n-dimensional Hilbert space $(V, (\cdot, \cdot))$. Choosing a basis turns this into a sparse linear system with $O(n)$ non-zero entries as described in Section 1.2.3. This implies that $O(n)$ is a lower bound for the complexity of solving the system.

Direct linear solvers perform basis transformations to bring the given matrix into a more favourable form, such as triangular. If done without great care, this will ruin the sparsity, making it impossible for such methods to reach the desired complexity of $O(n)$. Modern sparse direct solvers attempt to avoid this by exploiting the specific sparsity pattern of a given problem. Although they do not achieve $O(n)$ complexity, they are competitive for surprisingly large systems.

As $n \to \infty$, however, there comes a point where the computational cost becomes unfeasible for any known direct solver. Thus, for very large systems, iterative solvers must be employed. They achieve complexity $O(n)$, although with a large constant, which is why direct solvers typically win out on smaller problems. The general idea is to construct a sequence $\{u_i\}_i$ that converges to u, with u_{i+1} computed from u_i using only sparse matrix-vector products.

We will not delve deeper into iterative methods here but simply present the ones used in this thesis: a Richardson-type iteration, the Conjugate Gradient method (CG) and the Generalized minimal residual method (GMRES). Furthermore, we will provide an overview of various two-level Schwarz preconditioning schemes, some of which will be treated in detail in later chapters. For a more thorough introduction to the field of sparse linear solvers, refer to textbooks such as [53, 61].

We close the section with a discussion of the concepts of approximation space and coarse space whose interplay will be a returning theme throughout this thesis.

1.4.1 Well-Conditioned Systems

The convergence rate of an iterative solver typically depends on whether the given system is well-conditioned. For a general viewpoint on this, consider the linear problem of finding $x \in V$ in a Hilbert space V such that $Mx = b$ for a linear map $M : V \to V$ and some $b \in V$.

The classic way to formalize well-conditioning is the notion of condition number. The reader is probably familiar with the concept but since we will work with different norms in play, we go into some detail here. To get a better understanding, we start in coordinate free notation.

Definition 1.43. Let

$$M : (X, \|\cdot\|_x) \to (Y, \|\cdot\|_y)$$

be a bounded linear map between normed vector spaces with bounded inverse M^{-1}. Then we call

$$\kappa_{x,y}(M) := \|M\|_{x,y}\|M^{-1}\|_{y,x}$$

the *condition* number of M. We often omit the subscripts if there is no risk of confusion.

Notably this notion depends on the norms in use. The condition number of an invertible matrix is then defined as the condition number of the corresponding linear

map $(\mathbb{R}^n, \|\cdot\|_2) \to (\mathbb{R}^n, \|\cdot\|_2)$. It is very important to note however, that the Euclidean norm $\|\cdot\|_2$ depends on the choice of basis and thus, so does the condition number.

To make this dependence more clear, we assume that V is finite dimensional and choose a basis $\Phi : V \to \mathbb{R}^n$. Then the condition number of M's matrix representation is given by

$$\kappa_2(\boldsymbol{M}) = \|\boldsymbol{M}\|_2 \|\boldsymbol{M}^{-1}\|_2 = \|\Phi M \Phi^{-1}\|_2 \|\Phi M^{-1} \Phi^{-1}\|_2.$$

In this form, the dependence on Φ becomes explicit.

Note that if we chose an orthonormal basis, i.e. if Φ was an isometry, we would have $\kappa(\boldsymbol{M}) = \kappa(M)$. Often however, an orthonormal basis is not available or just not as practical as other choices. Consider for instance the diffusion equation from Section 1.3.1. Several norms may be of interest here such as the L^2-norm, the H^1-norm, or the a-norm (energy norm).

Unfortunately the classic Lagrange basis for finite elements is not orthonormal with respect to any of those. Nevertheless, for historical and computational reasons, it remains the most widely used choice. This often leads to a mismatch between the norms that are convenient for theoretical analysis and the Euclidean norm employed in practice. This fact introduces a norm equivalence constant to many theoretical results when they are applied to practical settings — a constant that typically depends on the mesh size h and should be tracked carefully.

In particularly favourable cases, we may have

$$\|\mathrm{Id} - M\| \leq C \tag{1.21}$$

for a constant $C < 1$. This condition makes its dependence on the norm explicit and is easy to interpret.

While this criterion may appear more straightforward than working with condition numbers, its scope is often quite limited. For instance, if $C \geq 1$, the condition becomes practically meaningless — it is satisfied by $M = 0$. On the other hand, requiring $C < 1$ immediately imposes a strong constraint. To illustrate, suppose M is symmetric and positive definite, and that (1.21) holds with $C = 1 - \epsilon$ in the $\|\cdot\|_2$ norm. Then all eigenvalues of $\boldsymbol{M}$ must lie in the interval $[\epsilon, 2 - \epsilon]$. The upper bound in this is often too restrictive and is violated by many common preconditioners.

Nevertheless, this property does hold for the MS-GFEM preconditioner introduced later in this thesis, and it will serve as a key tool for proving convergence of the corresponding iterative solvers.

1.4.2 Iterative Solvers

In this section, we briefly introduce the iterative solvers used throughout this thesis: the Richardson-type iteration, the Conjugate Gradient method (CG), and the Generalized Minimal Residual method (GMRES).

We consider the task of solving a linear system

$$Mx = b,$$

where $M \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n =: V$. As a basic example, consider the Richardson-type iteration:

Example 1.44 (Richardson)**.** Let

$$x_{k+1} := x_k + b - Mx_k$$

with some arbitrary starting value $x_0 \in V$. Note that we have

$$\|x_{k+1} - x\| = \|(\boldsymbol{I} - \boldsymbol{M})(x_k - x)\| \leq \|\boldsymbol{I} - \boldsymbol{M}\|\|x_k - x\|.$$

So if (1.21) holds, one has $x_k \to x$ with rate C.

As discussed in the previous subsection, the condition (1.21) can be verified for the MS-GFEM preconditioner which we introduce in Chapter 6. However, it is often too restrictive, and more sophisticated solvers are required. Two of the most commonly used alternatives are CG and GMRES, both Krylov subspace projection methods. CG applies to symmetric positive definite systems, while GMRES is designed for more general non-symmetric systems. In this thesis, we treat both methods as black boxes and focus on their high-level structure. For more detailed treatments, see [53, 61].

Let $K_m, L_m \subset V$ be m-dimensional subspaces, and let $x_0 \in V$ be an initial guess. A *projection method* seeks an approximation $x_m \in x_0 + K_m$ such that

$$b - Mx_m \perp L_m. \tag{1.22}$$

Here, K_m is the *candidate subspace* and L_m the *restriction subspace*. The following results characterize x_m under certain assumptions.

Proposition 1.45. *Using the notation above, x_m is well-defined if either:*

1. *M is symmetric positive definite and $L_m = K_m$, or*
2. *M is invertible and $L_m = MK_m$.*

Proposition 1.46 (CG optimality)**.** *If M is symmetric positive definite and $L_m = K_m$, then x_m is uniquely characterized by*

$$\|x - x_m\|_M = \min_{v \in x_0 + K_m} \|x - v\|_M,$$

where we write $\|\cdot\|_M := \|M^{1/2}\cdot\|$.

Proposition 1.47 (GMRES optimality)**.** *If M is invertible and $L_m = MK_m$, then x_m is uniquely characterized by*

$$\|b - Mx_m\| = \min_{v \in x_0 + K_m} \|b - Mv\|,$$

i.e. x_m minimizes the residual in the Euclidean norm.

These results form the foundation for Krylov subspace methods. The key is to choose K_m such that the error or residual norms converge to zero as $m \to \infty$.

However, the norms involved in these optimality conditions are often not the ones of theoretical interest. The Euclidean norm $\|\cdot\|_2$ depends on the computational basis and is rarely orthonormal with respect to meaningful physical norms. Likewise, the M-norm may correspond to an energy norm — e.g. when solving $\boldsymbol{A}u = \boldsymbol{f}$ in diffusion problems — but in practice, preconditioning is used. That is, we often solve $\boldsymbol{BA}u = \boldsymbol{Bf}$ for some preconditioner $\boldsymbol{B}$, effectively optimizing in the artificial norm $\|\cdot\|_{BA}$.

Therefore, when analysing iterative solvers, it is essential to keep track of which norm is in use. While all norms are equivalent in finite dimensions, the constants of equivalence typically depend on the mesh size h.

Krylov subspaces. Both CG and GMRES are based on the Krylov subspace

$$K_m := \operatorname{span}\{r_0, Mr_0, \dots, M^{m-1}r_0\},$$

where $r_0 := b - Mx_0$ is the initial residual.

Example 1.48 (Conjugate Gradient)**.** The *Conjugate Gradient* method (CG) is a Krylov subspace method for symmetric positive definite matrices and uses $L_m = K_m$.

Example 1.49 (Generalized Minimal Residual Method (GMRES))**.** GMRES is a Krylov subspace method for general nonsingular matrices and uses $L_m = MK_m$.

Note that both methods are defined with respect to an inner product, which determines the orthogonality in (1.22) and the norms in the optimality results Proposition 1.46 (CG) and Proposition 1.47 (GMRES). Unless otherwise specified, this is assumed to be the Euclidean inner product.

A typical convergence result for CG is given by the following theorem.

Theorem 1.50 (CG convergence)**.** *Let x_m be the m-th CG iterate. Then*

$$\|x - x_m\|_M \leq 2\left(\frac{\sqrt{\kappa(M)} - 1}{\sqrt{\kappa(M)} + 1}\right)^m \|x - x_0\|_M.$$

Proof. See [53], Theorem 6.6 and Equation (6.107). □

In Chapter 6, we will use Proposition 1.47 directly to obtain sharper bounds for the MS-GFEM preconditioner in the context of GMRES.

1.4.3 Parallel Preconditioners

Most known iterative solvers achieve acceptable convergence rates only when applied to well-conditioned systems. This is made precise by theorems like Theorem 1.50. Unfortunately, however, the systems arising from PDE discretizations are typically extremely ill-conditioned.

Consider for instance the stiffness matrix $\boldsymbol{A}$ that arises from discretizing the diffusion equation with piecewise linear, continuous functions on a uniform mesh of size h in the Lagrange basis. It is well-known that in this case $\kappa(\boldsymbol{A}) \propto h^{-2}$.

Typically, linear systems arising from PDE discretizations thus have to be preconditioned. To this end an invertible matrix $\boldsymbol{B}$ is multiplied to the system (1.11) from the left to obtain the equivalent system

$$\boldsymbol{BAu} = \boldsymbol{Bf}.$$

The aim is to choose the preconditioner $\boldsymbol{B}$ such that $\boldsymbol{BA}$ is well-conditioned, while also being reasonably inexpensive to construct and apply.

A simple example is the diagonal preconditioner

$$\boldsymbol{B}_{ij} = \delta_{ij} / \boldsymbol{A}_{ij},$$

that consists of the inverses of the diagonal entries of $\boldsymbol{A}$. This can work reasonably well for diagonally dominant matrices in favourable cases. For the highly heterogeneous PDEs under consideration in this thesis, however, we will need something more sophisticated.

In the following, we define several common overlapping two-level Schwarz preconditioning schemes. Two-level Schwarz methods are an example of subspace correction methods. They alternate between local solves on subdomains, typically computed in parallel, with a coarse correction in a low-dimensional global *coarse space* S.

To keep the presentation general, we formulate all methods in coordinate-free notation over an abstract Hilbert space $(V, (\cdot,\cdot))$. We consider the problem: find $u \in V$ such that

$$a(u, v) = l(v) \quad \text{for all } v \in V, \tag{1.23}$$

for a continuous bilinear form $a(\cdot,\cdot)$ and linear form $l(\cdot)$ on V.

Recall that by Proposition 1.26 there is a linear map $A : V \to V$ such that

$$a(v, w) = (Av, w) \text{ for all } v, w \in V,$$

and by the Riesz representation theorem there is an $f \in V$ such that $l(v) = (f, v)$ for all $v \in V$. Therefore, the problem is equivalent to

$$Au = f,$$

and u is well-defined if A is invertible. Note that in practice we will use these preconditioners for the case $(V, (\cdot,\cdot)) = (\mathbb{R}^n, (\cdot,\cdot)_2)$, and in this case A is simply the stiffness matrix.

Subspace correction methods split the global problem (1.23) into smaller ones by decomposing V into spaces $S, V_1, ..., V_M \subset V$, i.e.

$$V = S + V_1 + ... + V_M. \tag{1.24}$$

The idea is to restrict the global problem to the individual subspaces and use these restrictions as building blocks for a preconditioner. We use the symbol S for the coarse space to account for the special role it will take below. Note that we do not require this decomposition to be a direct sum of vector spaces. Or in other words, while elements $v \in V$ do have a decomposition

$$v = s + v_1 + \cdots + v_M,\ s \in S, v_i \in V_i,$$

this decomposition is not assumed to be unique. Before we continue, consider the following example.

Example 1.51 (Domain decomposition)**.** Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary, and let $\mathcal{T}$ be a simplicial mesh of Ω. Define V as the space of continuous piecewise linear functions on $\mathcal{T}$ that vanish on the boundary $\partial\Omega$.

Suppose that Ω is covered by a collection of open subdomains $\omega_1, \ldots, \omega_M$ such that $\Omega = \bigcup_{i=1}^{M} \omega_i$. For each i, define

$$V_i := \{v \in V \mid \operatorname{supp}(v) \subset \omega_i\}.$$

Then the decomposition (1.24) holds with $V = V_1 + \cdots + V_M + S$, for any choice of coarse space $S \subset V$. The selection of an appropriate S is crucial for the convergence of the Schwarz methods discussed below. Preconditioners based on this type of setting are commonly referred to as domain decomposition methods.

For any subspace $W \subset V$ we write $\iota_W : W \hookrightarrow V$ for the inclusion. We assume here that W is closed, making $(W, (\cdot,\cdot))$ a Hilbert space and ι_W an isometry. If we restrict

$a(\cdot,\cdot)$ to W, we have for all $v, w \in W$ that

$$a(v,w) = a(\iota_W v, \iota_W w) = (A\iota_W v, \iota_W w) = (\iota_W^* A \iota_W v, w) = (A_W v, w)$$

where we abbreviated

$$A_W := \iota_W^* A \iota_W.$$

For invertible A_W, we can define the maps

$$\pi_W := A_W^{-1} \iota_W^* A.$$

Notation 1.52. For the remainder of this section, we assume that A_W is invertible for $W \in \{V, S, V_1, \dots V_M\}$. We abbreviate $\pi_i := \pi_{V_i}$ as well as $A_i := A_{V_i}$ and $\iota_i := \iota_{V_i}$.

Note that if π_W is well-defined, we have for any $v \in V$ that

$$a(\pi_W v, w) = (A_W \pi_W v, w) = (\iota_W^* A v, w) = (Av, w) = a(v,w) \text{ for all } w \in W,$$

which leads to the following example.

Example 1.53. If $a(\cdot,\cdot)$ is symmetric and coercive, then for any closed subspace $W \subset V$ the map A_W is invertible and π_W is the a-orthogonal projection onto W.

We can now formulate (overlapping) two-level Schwarz methods for the Hilbert space setting. In this thesis, we will look at additive Schwarz, as well as a particular hybrid variant that is multiplicative in the levels and additive between subdomains. The key for the success of these methods is that most of the work (applying the maps π_i) can be done in parallel. They are thus well-suited to exploit modern hardware architectures.

We will phrase each method as a Richardson-type iteration before we extract the corresponding preconditioner that can be used in any iterative solver. Note that for these methods to be effective, a suitable S has to be chosen. This choice will be a major topic in this thesis.

Definition 1.54 (Additive Schwarz method)**.** For a dampening factor $\chi \in \mathbb{R}$, we call the iteration

$$u^{k+1} = u^k + \chi \left(\pi_S + \sum_{i=1}^{M} \pi_i \right) (u - u^k), \tag{1.25}$$

the additive Schwarz Richardson-type iteration. The corresponding preconditioner is given by

$$B = \iota_S A_S^{-1} \iota_S^* + \sum_{i=1}^{M} \iota_i A_i^{-1} \iota_i^*.$$

Note that we allowed for a dampening factor $\chi \in \mathbb{R}$. This is to account for the fact that the decomposition of V is not a direct sum. To illustrate this, set $M = 2$ and $S = V_1 = V_2 = V$. In this case the additive Schwarz method will converge in one step with $\chi = 1/3$, but diverge for $\chi = 1$.

A variation of the above is to alternate between corrections in the coarse space S, and corrections the local spaces V_i as follows.

Definition 1.55 (Hybrid (two-level) Schwarz)**.** For a dampening factor $\chi \in \mathbb{R}$, we call the iteration

$$u^{k+1} = u^k + \pi_S(u - u^k) + \chi \left(\sum_{i=1}^{M} \pi_i \right) (u - u^k - \pi_S(u - u^k)), \tag{1.26}$$

the hybrid two-level Schwarz Richardson-type iteration. The corresponding preconditioner is given by

$$B = \left(\sum_{i=1}^{M} \iota_i A_i^{-1} \iota_i^* \right) + \iota_S A_S^{-1} \iota_S^* \left(\mathrm{Id} - A \sum_{i=1}^{M} \iota_i A_i^{-1} \iota_i^* \right).$$

Note that this method is multiplicative between the levels, hence the name hybrid.

Restricted Schwarz methods introduce a partition of unity which can be seen as a more sophisticated version of a dampening factor. The resulting preconditioner looses symmetry which is unfortunate as this forces the use GMRES instead of the more efficient CG. However, as we will demonstrate in Chapter 6, this can be well worth it due to much faster convergence.

Definition 1.56 (Restricted additive Schwarz)**.** For a dampening factor $\chi \in \mathbb{R}$ and a partition of unity $\chi_i : V \to V_i$, $i = 1, ..., M$, we call the iteration

$$u^{k+1} = u^k + \chi \left(\pi_S + \sum_{i=1}^{M} \chi_i \pi_i \right) (u - u^k), \tag{1.27}$$

the restricted additive Schwarz Richardson-type iteration. The corresponding preconditioner is given by

$$B = \iota_S A_S^{-1} \iota_S^* + \sum_{i=1}^{M} \chi_i \iota_i A_i^{-1} \iota_i^*.$$

Definition 1.57 (Restricted hybrid (two-level) Schwarz)**.** For a partition of unity $\chi_i : V \to V_i$, $i = 1, ..., M$, we call the iteration

$$u^{k+1} = u^k + \pi_S(u - u^k) + \chi \left(\sum_{i=1}^{M} \chi_i \pi_i \right) (u - u^k - \pi_S(u - u^k)), \tag{1.28}$$

restricted hybrid (two-level) Schwarz Richardson-type iteration. The corresponding preconditioner is given by

$$B = \left(\sum_{i=1}^{M} \chi_i \iota_i A_i^{-1} \iota_i^* \right) + \iota_S A_S^{-1} \iota_S^* \left(\mathrm{Id} - A \sum_{i=1}^{M} \chi_i \iota_i A_i^{-1} \iota_i^* \right).$$

Note that without the choice of a coarse space S, these preconditioning schemes are relatively bland templates. In the next chapter we will discuss a whole class of such choices: spectral coarse spaces. Before that however, let us look at S from a more general perspective.

Suppose we are solving a problem arising from PDE discretization. More precisely let $\{\mathcal{T}_h\}_h$ be a family of shape regular meshes and $\{V_h\}_h$ the corresponding family of finite element spaces. In this scenario, we want to find $u_h \in V_h$ such that

$$a(u_h, v) = \ell(v) \text{ for all } v \in V_h$$

where the bilinear form $a(\cdot,\cdot)$ and the linear form ℓ arise from the weak PDE formulation. Choosing the Lagrange basis we obtain the linear system

$$\boldsymbol{A}\boldsymbol{u} = \boldsymbol{f}$$

for which we would like to find a suitable coarse space $S \subset V_h$ such that at least one of the two-level Schwarz preconditioning schemes leads to a robust preconditioner $\boldsymbol{B}$. "Robust" is typically taken to mean the following in this setting.

Definition 1.58. In the setting of a PDE discretization as outlined above, an invertible matrix $\boldsymbol{B}$ is called a robust preconditioner for $\boldsymbol{A}$, if

$$\kappa_2(\boldsymbol{B}\boldsymbol{A}) \leq C \tag{1.29}$$

as $h \to 0$, with a constant $C > 0$ independent of h.

In Chapter 5, we will give a detailed analysis of the properties of a multilevel version of the additive Schwarz method for spectral coarse spaces. Then, in Chapter 6, we will use the MS-GFEM approximation space as coarse space in a multilevel restricted hybrid Schwarz preconditioner, and show arbitrarily fast convergence.

1.4.4 Approximation Space vs. Coarse Space

Returning to the choice of the coarse space S, a classical candidate is the same finite element space, defined instead on a coarser mesh. This approach can be effective, at least for homogeneous problems. The observation that the same construction can serve as both an approximation space and a coarse space naturally raises the question: what precisely is the relationship between these two concepts?

Consider, for instance, the problem of solving the diffusion equation (1.15) in a finite element space V defined on a very fine mesh. Formally, we seek $u \in V$ satisfying

$$a(u,v) = \ell(v) \quad \text{for all } v \in V, \tag{1.30}$$

where $a(\cdot,\cdot)$ is the energy inner product, and $\ell(\cdot)$ is a linear functional on V. After choosing appropriate bases, this leads to a linear system

$$\boldsymbol{A}\boldsymbol{u} = \boldsymbol{f}.$$

Assume we are seeking a way to either reduce the dimension of this problem by restricting it to an approximation space, or to parallelize it with a two-level Schwarz preconditioner. By an *approximation space*, we mean a sequence of subspaces $W_n \subset V$ for which the approximation error vanishes in the limit:

$$\inf_{v \in W_n} \|u - v\|_V \to 0 \quad \text{as} \quad n \to \infty. \tag{1.31}$$

The aim here is to choose the spaces W_n with a low dimension, while ensuring a guaranteed rate of convergence in the approximation property above.

Remark 1.59. Note that in classical approximation theory, property (1.31) is typically formulated in infinite-dimensional function spaces, such as $V = H_0^1(\Omega)$, rather than their finite-dimensional discretizations. Here we adopt the discrete setting to draw a clearer parallel to coarse spaces arising in domain decomposition methods.

Following the notation of the previous chapter, let

$$B_n = A_{W_n}^{-1} \iota_{W_n}^*.$$

Then we have $B_n A = \pi_{W_n}$, the a-orthogonal projection onto W_n. Due to the approximation property of W_n, we thus have

$$\|u - B_n A u\| \to 0 \quad \text{as} \quad n \to \infty. \tag{1.32}$$

The intended application of an approximation space is to apply $\boldsymbol{B}_n$ once to $\boldsymbol{f}$, producing an approximate solution to $\boldsymbol{u}$. This is sometimes informally referred to as *using the space W_n as a solver*. At first glance, one might imagine that such a construction must be based on an even stronger property than what is required from a preconditioner — after all, preconditioners typically must be applied iteratively to yield accurate approximations. Thus, one might suspect $\boldsymbol{B}_n$ itself would serve as a good preconditioner. However, this intuition is misleading, as we illustrate below.

The first and most straightforward observation is that $\boldsymbol{B}_n$ is not invertible, unless $W_n = V$. Consequently, the preconditioned system

$$\boldsymbol{B}_n \boldsymbol{A} \boldsymbol{u} = \boldsymbol{B}_n \boldsymbol{f}$$

is not equivalent to the original one. Another viewpoint is that, as a projection, $\boldsymbol{B}_n \boldsymbol{A}$ is idempotent:

$$(\boldsymbol{B}_n \boldsymbol{A})^2 = \boldsymbol{B}_n \boldsymbol{A}.$$

Therefore, iterative methods employing such a preconditioner stagnate immediately after the first step.

Clearly, condition (1.32) alone does not suffice to guarantee a good preconditioner. The fundamental issue is that this convergence property is not uniform in u — or equivalently, not uniform with respect to arbitrary right-hand sides. Indeed, as long as $W_n \neq V$, there always exists some $u \in W_n^{\perp_a}$ that W_n completely fails to approximate. A robust preconditioner, on the other hand, provides uniform bounds on the condition number independently of the right-hand side.

Thus, an approximation space alone does not inherently yield a suitable preconditioner. Nevertheless, approximation spaces often do serve as effective coarse spaces within a two-level Schwarz scheme. Historically, the development of approximation spaces and coarse spaces has significantly influenced each other. Still, the suitability of an approximation space W_n as a coarse space crucially depends on the specific structure of the chosen Schwarz method, and cannot be deduced solely from the approximation property (1.32).

We will repeatedly encounter the interplay of these two concepts throughout this thesis.

Chapter 2

Spectral Domain Decomposition

In this chapter we review a selection of spectral domain decomposition methods and their predecessors. For comparability, they are presented in the framework of generalized finite elements (GFEM) where possible. The goal is to build intuition and provide examples for the abstract spectral elements developed in the following chapters. A review with a stronger focus on model order reduction methods can be found in [6].

For ease of exposition we consider the setting of solving the heterogeneous diffusion equation and use the notation given in Section 1.3.1. This means we consider a finite element space $V_h \subset H_0^1(\Omega)$ in which we are aiming to find $u_h \in V_h$ such that

$$a(u_h, v_h) = \ell(v_h) \quad \text{for all } v_h \in V_h,$$

with the coercive bilinear form $a(\cdot,\cdot)$ and linear form ℓ as defined there. Choosing a basis leads to a linear system that can be solved on a computer as discussed in the previous chapter. Since we require the mesh to resolve the fine-scale coefficient, this linear system will often be too large to solve on a single machine. The aim of the methods presented here is to remedy this problem by constructing a space V_H of lower dimension

$$V_H \subset V_h \subset H_0^1(\Omega)$$

that is tailored to the given problem. This V_H is then used either as an approximation space or, alternatively, as the coarse space of a two-level Schwarz preconditioner. Section 1.4.3 elaborates on the similarities and differences of these two approaches.

To construct this V_H, the methods discussed here roughly follow two approaches. Earlier ones look at local linear systems that essentially solve the given PDE subject to artificial boundary conditions. Their solutions then offer building blocks for a global solution and are tailored to the given problem. Later efforts to find a general way to define purposeful local boundary conditions led to solving local eigenproblems rather than just local linear systems. This approach gives the corresponding methods the name *spectral*.

This raises a natural question. Why should it be cheaper to construct an informed V_H rather than solving in V_h directly? The answer is that it is not. In fact, it will very likely be more expensive in total computation time. However, the vast majority of the work is done locally, i.e. can be performed in parallel. Moreover, there are scenarios where setting up V_H may be cheaper than expected, or where some, or even all, of the work of setting up V_H can be recycled. Consider, for instance, a periodic coefficient. In this case it may be enough to perform the local computations for setting up V_H only once, since all local problems are identical. Alternatively, consider a case where the same PDE has to be solved multiple times with different right-hand sides, which allows multiple uses for the same V_H. Further possible applications arise from nonlinear problems, and from uncertainty quantification, where typically many similar problems have to be solved.

Throughout this chapter, for any sufficiently regular subdomain $\omega \subset \Omega$ we consider the local function space

$$V_h(\omega) = \{v_h|_\omega \,:\, v_h \in V_h\}, \quad V_{h,0}(\omega) = \{v_h \in V_h(\omega) \,:\, \operatorname{supp}(v_h) \subset \omega\},$$

and the local bilinear form

$$a_\omega(u, v) := \int_\omega (A\nabla u) \cdot \nabla v \, dx \quad \text{for all } \; u, v \in H^1(\omega). \tag{2.1}$$

Moreover, we define

$$\|u\|_{a_\omega} := \sqrt{a_\omega(u,u)} \quad \text{for all} \;\; u \in H^1(\omega),$$

and when $\omega = \Omega$, we simply write $\|u\|_a$.

2.1 Approximation with Locally Harmonic Functions

There are numerous methods that build a coarse space from locally harmonic functions. In this section, we explore two methods that were particularly influential for the development of the Generalized Finite Element Method (GFEM) and the Multiscale Finite Element Method (MSFEM). Interestingly, both methods follow similar ideas, even though GFEM initially focused on complex domain shapes with re-entrant corners, a popular topic in the 1990s. In contrast, MSFEM, which emerged in the 2000s, primarily addresses highly heterogeneous coefficients, a field made computationally feasible by advances in computer technology.

Locally harmonic, in this context, refers to solving the respective PDE on a subdomain $\omega_i \subset \Omega$, with either an open cover (GFEM) or a non-overlapping decomposition (MSFEM). A fundamental issue with this approach is that, without further restrictions (such as boundary conditions), a typical PDE may have infinitely many solutions. To form a coarse space, a finite number of solutions must be selected. For example, one might consider polynomials up to a certain degree (GFEM), or enforce a handful of artificial boundary conditions (MSFEM). These restrictions can seem somewhat arbitrary, which eventually led to the development of spectral methods as discussed in Section 2.2.

With a finite selection of locally harmonic functions, a remaining issue is that these functions do not generally vanish at the boundary of their respective subdomains. To address this, an additional step is required to extend them to global, continuous functions. In GFEM, this is achieved using a partition of unity, while in MSFEM, artificial boundary conditions are selected in such a way that harmonic pieces from neighbouring subdomains can be fit together, much like the hat functions in conforming finite elements.

The intuition behind using piecewise locally harmonic functions is straightforward: we aim to build a global solution by combining locally valid pieces. Formalizing this idea is more difficult, and neither method provides fully satisfactory theory. MSFEM does offer explicit bounds under certain conditions at least, such as periodicity of the coefficients or the assumption of isotropic random coefficients with low autocorrelation length. Efforts to improve these results led to spectral methods, which we describe later in this chapter.

2.1.1 Generalized Finite Elements

The Generalized Finite Element Method (GFEM) [7, 59] provides a framework for constructing PDE solutions. Originally introduced as an approximation theory at the continuous level, typically in the H^1 setting, we present GFEM here in the discrete setting of the heterogeneous diffusion equation described earlier. This aligns the presentation with the notation and context of the current chapter. Similar concepts have become prevalent in localized model order reduction and domain decomposition methods, and we will employ these ideas frequently throughout this thesis.

Let $\omega_i \subset \Omega$ be open subdomains such that $\Omega = \cup_{i=1}^{M} \omega_i$, each resolved by the given mesh. The maximum number of subdomains overlapping at any given point, known as the *colouring constant*, will be important later and is denoted by

$$\xi := \max_{x \in \Omega} |\{i \mid x \in \omega_i\}|. \tag{2.2}$$

A key ingredient in GFEM is a partition of unity. Specifically, we consider functions $\{\chi_i\}_{i=1}^M$ on Ω that satisfy:

$$\begin{aligned} &\operatorname{supp}(\chi_i) \subset \overline{\omega_i}, \quad 0 \le \chi_i \le 1, \quad \sum_{i=1}^M \chi_i = 1 \ \text{ on } \ \Omega, \\ &\chi_i \in W^{1,\infty}(\omega_i), \ \ \|\nabla \chi_i\|_{L^\infty(\omega_i)} \le \frac{C_\chi}{\operatorname{diam}(\omega_i)}. \end{aligned} \tag{2.3}$$

Using this partition of unity, we define the operators

$$\begin{aligned} \chi_{i,h} : V_h(\omega_i) &\to V_{h,0}(\omega_i), \\ v_h &\mapsto I_h(\chi_i v_h), \end{aligned}$$

where $I_h : H_0^1(\Omega) \to V_h$ is the standard Lagrange interpolation operator. Interpolation is necessary since finite element spaces are generally not closed under pointwise multiplication. Conveniently, this interpolation step preserves the partition of unity property, so that we still have

$$\sum_{i=1}^M \chi_{i,h}(v_h|_{\omega_i}) = v_h \quad \text{for all} \ \ v_h \in V_h.$$

In this chapter, we refer to $\{\chi_{i,h}\}_{i=1}^M$ as the partition of unity operators, or often just the partition of unity.

The central idea of GFEM is to construct a coarse global space $S \subset V_h$ by gluing suitably selected local functions together using the partition of unity. On each subdomain ω_i, let a local particular function $u_{h,i}^p \in V_h(\omega_i)$ and a local approximation space $S_{h,m_i}(\omega_i) \subset V_h(\omega_i)$ of dimension m_i be given. Their precise constructions will be discussed below. Then, the global particular function u_h^p and the global approximation space S of dimension $m = m_1 + \cdots + m_M$ are defined by

$$u_h^p = \sum_{i=1}^M \chi_{i,h}(u_{h,i}^p) \in V_h, \quad S = \sum_{i=1}^M \chi_{i,h}\left(S_{h,m_i}(\omega_i)\right) \subset V_h. \tag{2.4}$$

The GFEM approximation of the fine-scale finite element problem (1.16) is then defined as the unique $u_h^G \in u_h^p + S$ such that

$$a(u_h^G, v_h) = \ell(v_h) \quad \text{for all } v_h \in S. \tag{2.5}$$

By Cea's lemma, the GFEM approximation u_h^G is a quasi optimal approximation for u_h in $u_h^p + S$, i.e.

$$\|u_h - u_h^G\|_a \le C \inf_{\psi \in u_h^p + S} \|u_h - \psi\|_a, \tag{2.6}$$

where C depends only on $a(\cdot,\cdot)$. Hence, the accuracy of the GFEM approximation is determined by the quality of the global approximation space S, which, in turn, hinges on the accuracies of the local approximations spaces $S_{m,i}(\omega_i)$. This is made precise in the following theorem.

Theorem 2.1. *Let $v_h \in V_h$. Assuming that for each $i = 1, ..., M$,*

$$\inf_{\psi \in u_{h,i}^p + S_{h,m_i}(\omega_i)} \left\|\chi_{i,h}(v_h|_{\omega_i} - \psi)\right\|_{a,\omega_i} \le e_i,$$

then,

$$\inf_{\psi \in u_h^p + S} \|v_h - \psi\|_a \leq \left(\xi \sum_{i=1}^{M} e_i^2\right)^{1/2}.$$

Note that this means if we let e_i approach 0 by increasing the dimensions of the local approximation spaces $S_{h,m_i}(\omega_i)$, we achieve the approximation property (1.7) with $W = u_h^p + S$. In the context of two-level Schwarz preconditioners on the other hand, we will consider the coarse space $V_H = S$.

The core challenge in applying GFEM is finding suitable constructions for the local particular functions $u_{h,i}^p$ and local approximation spaces $S_{h,m_i}(\omega_i)$. In the original works, the local particular functions were primarily introduced to enforce global boundary conditions, typically independent of the source term (the right-hand side ℓ). Numerous approaches have been proposed for the local approximation spaces, generally involving locally harmonic functions — for example, local PDE solutions for polynomial source terms.

Spectral methods provide alternative choice for the local approximation spaces, as we will elaborate in Section 2.2. In this context, GFEM is sometimes employed with oversampling.

Definition 2.2. In the context of GFEM, we say that *oversampling* is used if the support of the partition of unity functions is strictly contained within their associated subdomains; that is, if

$$\overline{\operatorname{supp}(\chi_i)} \subset \omega_i \quad \text{for all } i.$$

The *amount of oversampling* is then defined as

$$\min_{1 \leq i \leq M} d\big(\partial\omega_i \setminus \partial\Omega, \operatorname{supp}(\chi_i)\big),$$

where d denotes the distance between subsets of Ω.

2.1.2 Multiscale Finite Elements

The Multiscale Finite Element Method (MSFEM) constructs approximation spaces V_H by solving localized PDE problems. Unlike GFEM, MSFEM was specifically developed to handle heterogeneous problems with strongly varying coefficients. A comprehensive discussion of this method can be found in [25].

MSFEM utilizes a non-overlapping coarse grid $\mathcal{T}_H$, in contrast to the overlapping decomposition employed in GFEM. Instead of using a partition of unity to assemble global basis functions, MSFEM imposes artificial local boundary conditions. These conditions allow neighbouring locally harmonic functions to be seamlessly combined into globally continuous "hat"-type basis functions.

More concretely, let $\mathcal{T}_H$ be a conforming coarse mesh that is resolved by the underlying fine mesh. Additionally, let $\{\psi_i^0\}_{i=1}^N$ denote the standard finite element hat-functions associated with the $\mathcal{T}_H$. Then, define for $i = 1, \ldots, N$ the multiscale basis function $\psi_i \in V_h$ as the local PDE solution sharing Dirichlet boundary conditions with the ψ_i^0, i.e.

$$\begin{aligned} a(\psi_i, v) &= 0 \quad \text{for all } v \in V_{h,0}(\mathring{K}) \text{ and all } K \in \mathcal{T}_H, \\ \psi_i &= \psi_i^0 \quad \text{on } \partial K \text{ for all } K \in \mathcal{T}_H. \end{aligned}$$

In other words, the basis functions ψ_i coincide with standard coarse-grid hat functions on the coarse-grid faces, and are harmonic within each coarse-grid element.

The approximate PDE solution u_H in MSFEM is then obtained by projecting the fine-scale solution u_h onto the coarse space $V_H := \text{span}\left(\{\psi_i\}_{i=1}^N\right)$ via the a-orthogonal projection π_{V_H}, i.e.

$$u_H = \pi_{V_H}(u_h).$$

The analysis assumes a periodic coefficient and can be extended to related settings, such as a random coefficient with small enough autocorrelation length. If the period is greater than the coarse mesh size H, we recover the standard result for finite elements.

Theorem 2.3 ([25], Theorem 6.4)**.** *Assume that the diffusion coefficient is periodic with period $\epsilon > H$, that the solution to (1.16) satisfies $u \in H^2(\Omega)$, and that Ω is a convex polygon. Then we have*

$$\|u - u_H\|_{H^1} \leq CH(|p|_{H^2} + \|f\|_{L^2}).$$

If, on the other hand, the period is smaller, one obtains the more interesting result below. Note that the two error terms on the right can be balanced by choosing an appropriate H.

Theorem 2.4. *Let the coefficient A be periodic with period $\epsilon < H$. Assume that the solution to (1.16) satisfies $u \in H^2(\Omega)$ and that Ω is a convex polygon Then we have*

$$\|u - u_H\|_{H^1} \leq C(H + \epsilon)\|f\|_{L^2} + C(\epsilon/H)^{1/2}\|p_0\|_{1,\infty}$$

where $p_0 \in H_0^2(\Omega) \cap W^{1,\infty}$ is the solution to the associated homogenized problem, see [25], Chapter 6.

While MSFEM was originally phrased as an approximation theory that seeks an approximate solution in the coarse space only, there is follow-up work that investigates the use of V_H in a two-level Schwarz preconditioner, for instance in [38].

2.1.3 Energy Minimization: Towards Spectral Methods

Theorem 2.4 crucially relies on scale separation in the coefficient, specifically on the assumption $\epsilon < H$ or a similar condition. This requirement is somewhat restrictive, as it excludes many important cases, particularly the multiscale problems of interest in this thesis, such as the layered carbon fibre structure shown in Figure 1.1.

Subsequent research has gradually clarified how the quality of the coarse space V_H depends explicitly on the *energy* (the a-norm) of its basis functions. Here, "energy" refers simply to the a-norm; thus, zero energy is equivalent to harmonicity.

For example, in [38], the authors incorporate the MSFEM coarse space V_H into an overlapping two-level Schwarz preconditioner. They allow subdomains to differ from the coarse mesh and employ a partition of unity to maintain global continuity of the coarse basis functions. Their final convergence result explicitly depends on the energies of the coarse basis functions and the partition of unity, and otherwise remains independent of the diffusion coefficient.

This approach is further refined in [54], where the authors consider various constraints for constructing a low-energy basis. Their motivating application is a problem involving channels of large coefficient spanning multiple subdomains. In such scenarios, imposing local boundary conditions that are not carefully adapted to the coefficient may lead to basis functions with high energy. To see why, note that a function exhibits large energy if its gradient is substantial in regions where the coefficient is large. Consider, for instance, prescribing Dirichlet boundary conditions

of 1 at one end of a high-conductivity channel and 0 at the other. In this case, the resulting function inevitably possesses a large gradient along the channel, thus resulting in high energy.

To overcome this difficulty, [54] explores various strategies for constructing low-energy bases through localized computations. These approaches lead naturally to local saddle-point or eigenvalue problems. Thus, their work represents an early example of spectral methods within the broader context of multiscale analysis.

2.2 Spectral Methods

In the late 2000s, the idea of using eigenfunctions of local eigenproblems as a coarse basis gained popularity. The aim of such methods is to provide a more principled and consistent way of selecting the most relevant local functions for constructing the global coarse space in domain decomposition methods. In the following, we list several spectral methods, with particular focus on GenEO and MS-GFEM.

Generally, spectral methods tend to offer strong theoretical guarantees regarding preconditioner robustness and approximation quality. However, solving the associated local eigenproblems makes the construction of the coarse space computationally expensive. Whether a spectral method is worthwhile therefore depends on the specific application. Spectral methods are particularly attractive in scenarios where some or all eigenfunctions can be reused in subsequent solves. Conversely, if only a single system needs to be solved, heuristic approaches such as algebraic multigrid are typically faster — provided they are applicable to the problem at hand.

2.2.1 GMsFEM

The Generalized Multiscale Finite Element Method (GMsFEM) [32, 24, 23] is a spectral method designed for solving heterogeneous PDE problems using finite element techniques. In what follows, we review a special case of GMsFEM for the heterogeneous diffusion equation and present it within the GFEM framework. For a broader treatment and additional applications, see [26].

In the previous section, we discussed how collections of locally harmonic functions can form effective coarse spaces. However, this raised the important question of which locally harmonic functions to select — or equivalently, which artificial local boundary conditions to impose. GMsFEM addresses this question in a systematic way.

We begin by defining the space of all locally harmonic functions over a subdomain ω_i as

$$\widetilde{V}_i := V_{h,0}(\omega_i)^{\perp_a} \subset V_h(\omega_i),$$

that is, as the a-orthogonal complement of $V_{h,0}(\omega_i)$ in $V_h(\omega_i)$.

Since $\widetilde{V}i$ is typically too large to include in the local approximation space in full, GMsFEM selects a few of its most important modes via a local eigenproblem: find eigenvalues $\lambda_{i,1} < \cdots < \lambda_{i,m_i} \in \mathbb{R} \cup +\infty$ and corresponding eigenfunctions $\psi_{i,1}, \ldots, \psi_{i,m_i} \in V_h(\omega_i)$ such that

$$\int_{\omega_i} A(x)\psi_{i,j}(x)v(x)\,dx = \lambda_{i,j}\, a_{\omega_i}(\psi_{i,j}, v) \quad \text{for all } v \in \widetilde{V}_i. \tag{2.7}$$

The local approximation space is then defined by the span of the first m_i eigenfunctions:

$$S_{h,m_i}(\omega_i) = \operatorname{span}\{\psi_{i,1}, \ldots, \psi_{i,m_i}\}. \tag{2.8}$$

These local spaces are assembled into a global approximation space S using a partition of unity, as in GFEM (cf. (2.4)). The resulting global space S can be used, for example, as a coarse space in a two-level Schwarz preconditioner.

Many variants of GMsFEM have been proposed, including modifications to the eigenproblem and strategies for reducing the size of $\widetilde{V}_i$ through preliminary "offline" computations. Theoretical convergence analyses depend on these choices and can be found in, for example, [1, 65, 62].

2.2.2 GenEO

Generalized Eigenproblems in the Overlap (GenEO) [56, 57] is a spectral domain decomposition method. As with GMsFEM, we here provide a brief overview using the GFEM framework for consistency. We will later present a generalization to our abstract setting of multilevel spectral elements in Section 4.2, and provide detailed convergence theory for the resulting multilevel additive Schwarz method in Chapter 5.

GenEO was specifically developed for heterogeneous problems involving alternating layers or channels of high and low coefficients. A typical application is linear elasticity in fibre-reinforced materials. A closely related predecessor is the Dirichlet-to-Neumann (DtN) approach [50, 22].

In GenEO, the local approximation spaces are selected by solving the following eigenproblem: find eigenvalues $\lambda_{i,1} < \cdots < \lambda_{i,m_i} \in \mathbb{R}$ and corresponding eigenfunctions $\psi_{i,1}, \ldots, \psi_{i,m_i} \in V_h(\omega_i)$ such that

$$a_{\omega_i^{\text{ovlp}}}(\chi_{i,h}\psi_{i,j}, \chi_{i,h}v) = \lambda_{i,j}\, a_{\omega_i}(\psi_{i,j}, v) \quad \text{for all } v \in V_h(\omega_i), \tag{2.9}$$

where the overlap region $\omega_i^{\text{ovlp}} \subset \omega_i$ is defined as

$$\omega_i^{\text{ovlp}} := \{x \in \omega_i \mid x \in \omega_j \text{ for some } j \neq i\}.$$

This is the portion of ω_i that overlaps with at least one neighbouring subdomain.

The local approximation space is then defined by

$$S_{h,m_i}(\omega_i) = \operatorname{span}\{\psi_{i,1}, \ldots, \psi_{i,m_i}\}, \tag{2.10}$$

and the global coarse space S is assembled using the GFEM-style gluing from (2.4).

Using the GenEO coarse space S in a two-level additive (non-restricted) Schwarz preconditioner yields the following condition number bound.

Theorem 2.5. *Let $\boldsymbol{B}^{GenEO}$ denote the additive two-level Schwarz preconditioner using the GenEO coarse space. Then the condition number of the preconditioned system satisfies*

$$\kappa(\boldsymbol{B}^{GenEO}\boldsymbol{A}) \leq C\,(1 + \lambda_{\max}), \tag{2.11}$$

where $\boldsymbol{A}$ is the stiffness matrix, $\lambda_{\max}$ the largest eigenvalue whose eigenvector was not added to a local approximation space and

$$C = (1 + \xi)\,(2 + \xi(2\xi + 1)).$$

Proof. See [56, Theorem 3.22]. □

Several variants of GenEO are explored [57]. First, it is proposed to replace $a_{\omega_i^{\text{ovlp}}}$ on the left-hand side of (2.9) with a_{ω_i}. We refer to this variant as the *Full GenEO eigenproblem*.

Second, the use of the resulting coarse space in a hybrid Schwarz method is considered. The combination of Full GenEO with hybrid Schwarz leads to the following improved condition number estimate.

Theorem 2.6. *Let $\boldsymbol{B}_{hy}^{GenEO}$ denote the hybrid two-level Schwarz preconditioner using the Full GenEO coarse space. Then:*

$$\kappa(\boldsymbol{B}_{hy}^{GenEO}\boldsymbol{A}) \leq \xi\,(1 + \xi\lambda_{\max}), \tag{2.12}$$

where $\boldsymbol{A}$ is the stiffness matrix and λ_{max} the largest eigenvalue whose eigenvector was not added to a local approximation space.

Proof. See [57, Theorem 4.40]. □

2.2.3 MS-GFEM

The Multiscale Spectral Generalized Finite Element Method (MS-GFEM) [46] is another spectral method for constructing local approximation spaces in the GFEM framework. A key advantage of MS-GFEM is that it exhibits near-exponential decay in the spectrum of the local eigenproblems, leading to a soft bound on the dimension of the global approximation space S needed to achieve a desired accuracy. This sets it apart from many similar methods, which often lack theoretical guarantees on eigenvalue decay.

To achieve this, MS-GFEM solves the Full GenEO eigenproblems, but restricted to the locally harmonic spaces

$$\widetilde{V}_i := V_{h,0}(\omega_i)^{\perp_a},$$

as in GMsFEM. MS-GFEM is usually employed with oversampling, as this is a requirement to guarantee the above-mentioned eigenvalue decay.

The *local particular functions* $u^p_{h,i} \in V_{h,0}(\omega_i)$ in MS-GFEM are defined as solutions of

$$a_{\omega_i}(u^p_{h,i}, v) = \ell(v) \quad \text{for all } v \in V_{h,0}(\omega_i). \tag{2.13}$$

Note that the remaining error after subtracting $u^p_{h,i}$ lies in the a-orthogonal complement of $V_{h,0}(\omega_i^*)$, i.e.

$$u_h|_{\omega_i^*} - u^p_{h,i} \in \widetilde{V}_i,$$

which motivates the restriction of the eigenproblems to this space.

The local eigenproblems are defined as: find $\lambda_{i,1} < \cdots < \lambda_{i,m_i} \in \mathbb{R}$ and $\psi_{i,1}, \ldots, \psi_{i,m_i} \in \widetilde{V}_i$ such that

$$a_{\omega_i}(\chi_{i,h}\psi_{i,j}, \chi_{i,h}v) = \lambda_{i,j}\, a_{\omega_i}(\psi_{i,j}, v) \quad \text{for all } v \in \widetilde{V}_i. \tag{2.14}$$

Then, the local approximation spaces are set to be

$$S_{h,m_i}(\omega_i) = \operatorname{span}\{\psi_{i,1}, \ldots, \psi_{i,m_i}\}, \tag{2.15}$$

which satisfy the following local approximation property.

Lemma 2.7. *Let u_h be the solution of (1.16), and let $u^p_{h,i}$ and $S_{h,m_i}(\omega_i)$ be defined by (2.13) and (2.15), respectively. Then,*

$$\inf_{\psi \in u^p_{h,i} + S_{h,m_i}(\omega_i)} \|\chi_{i,h}(u_h|_{\omega_i} - \psi)\|_{a,\omega_i} \leq \lambda^{1/2}_{i,m_i+1} \|u_h\|_{a,\omega_i^*}.$$

Proof. See [45, Theorem 2.3]. □

The global approximation space S is constructed as in GFEM, see (2.4). A global error bound follows from Lemma 2.7 and Theorem 2.1:

Theorem 2.8. *Let the local particular functions $u^p_{h,i}$ and local approximation spaces $S_{h,m_i}(\omega_i)$ be defined as above. Then,*

$$\inf_{\psi \in u^p_h + S_{h,m}(\Omega)} \|u_h - \psi\|_a \leq \Lambda \|u_h\|_a,$$

with the constant Λ is given by

$$\Lambda := \left(\xi\xi^*\lambda_{\max}\right)^{1/2},$$

where ξ is the colouring constant for the cover $\{supp(\chi_i)\}_i$, ξ^* a colouring constant of the cover $\{\omega_i\}_i$, and $\lambda_{\max}$ is the largest eigenvalue whose eigenfunction was not included in a local approximation space.

Proof. See [45, Theorem 3.4]. □

As mentioned earlier, MS-GFEM comes with an explicit result on the asymptotic decay of eigenvalues.

Theorem 2.9. *Let $\delta_i = \operatorname{dist}(supp(\chi_i), \partial\omega_i \setminus \partial\Omega) > 0$. Then there exist constants $k_i > 0$ and $b_i > 0$, independent of h, such that for all $k > k_i$, if*

$$h \leq \frac{\delta_i}{4b_i k^{1/(d+1)}},$$

then

$$\lambda_{i,k}^{1/2} \leq (1 + C_\chi) \exp\left(-b_i k^{1/(d+1)}\right),$$

where C_χ is the constant from (2.3).

Proof. See [45, Theorem 4.8]. □

Remark 2.10. The constants k_i and b_i in Theorem 2.9 are given explicitly in [45]. In particular, k_i decreases and b_i increases with increasing oversampling size δ_i.

2.2.4 Discussion

MS-GFEM constructs its local eigenproblems from the local bilinear forms of Full GenEO, but restricted to the locally harmonic space from GMsFEM. Theorem 2.9 is what specifically sets MS-GFEM apart from the other two methods, which do not offer a comparable result. In this sense, the MS-GFEM eigenproblems can be viewed as the theoretically correct choice, representing a natural conclusion to two decades of work on locally harmonic approximation spaces. The method guarantees arbitrarily accurate approximation — or, as we will see in Section 6.1, arbitrarily fast convergence when used as a preconditioner.

However, the development of increasingly powerful methods has come with a corresponding increase in the computational cost of the coarse space construction. In particular, there is a notable increase in cost when moving from locally harmonic basis functions, as in MSFEM, to spectral methods that require solving local (generalized) eigenproblems.

Among spectral methods, MS-GFEM is particularly expensive for two reasons. First, the use of oversampling increases the size of the local problems, especially in three dimensions. Fortunately, our experiments in Chapter 6 indicate that minimal oversampling is usually sufficient. Second, even without oversampling, the restricted eigenproblems are — ironically — almost twice as large as their unrestricted counterparts. This is due to the introduction of a Lagrange multiplier that enforces orthogonality with respect to the local spaces $V_{h,0}(\omega_i)$. To mitigate this, dimension reduction techniques such as block elimination have been proposed [45], though these come with their own challenges as we investigate in Section 6.3.2.

Because of the high initial cost, spectral methods are particularly attractive in scenarios where the coarse space can be reused. In contrast, for single PDE solves, it is often difficult to justify the computational effort required for solving the local eigenproblems.

Beyond these practical considerations, MS-GFEM remains valuable due to its strong theoretical foundation, which makes it easier to extend to more challenging PDEs. For example, we successfully use it to precondition the Helmholtz equation in Section 6.4, and [42] discusses a wide range of PDE applications.

Chapter 3

Generalized Elements

From an algebraic perspective, the aim of a multilevel method is to decompose a Hilbert space (V, a) into a hierarchy of closed subspaces

$$V_0 \subset V_1 \subset ... \subset V_L = V,$$

that separate a given problem into different scales. The index l and the space V_l are both called the l-th level, where small l are called coarse levels, and large l fine levels. To parallelize the method, the individual levels are then themselves decomposed into closed subspaces

$$V_l = \sum_{i=1}^{M_l} V_{l,i,0}$$

with $V_{l,i,0} \subset V_l$. With such a decomposition, multilevel versions of the Schwarz methods discussed in Section 1.4.3 can be defined and analysed, as we do for the additive case in Chapter 5 and for the hybrid version in Chapter 6.

In this chapter, the goal is to provide an abstract framework for the construction of the spaces $V_l, V_{l,i,0}$. To this end, we develop the notion of a suitable presheaf, an object that is general enough to encompass a wide range of PDE applications, while offering exactly the properties that will be needed later.

After discussing how H^1 functions, continuous finite elements, and discontinuous finite elements fit into this framework in Section 3.2, we go on to define generalized elements on a suitable presheaf, an abstract version of the generalized finite elements. Here however, they come with a key property: Generalized elements on a suitable presheaf are themselves a suitable presheaf. Thus, this construction can be iterated to obtain successively coarser generalized elements — a multilevel hierarchy.

Generalized elements make use of local approximation spaces. We leave these arbitrary for now for the sake of clearer presentation. Their construction with abstract spectral methods is then detailed in the following chapter.

Readers who prefer to have a concrete example in mind may think of continuous piecewise linear finite elements on a simplicial mesh over a regular enough open domain $\Omega \subset \mathbb{R}^d$ with homogeneous Dirichlet boundary conditions in the following, and with the bilinear form arising from the weak formulation of the diffusion equation, i.e.

$$a(u, v) := \int_\Omega (A\nabla u) \cdot \nabla v \, dx = \ell(v) \quad \text{for all } v \in H_0^1(\Omega). \tag{3.1}$$

The following framework will go far beyond this setting, however.

3.1 Suitable Presheaves

The goal of this section is to develop an abstract framework for domain decomposition methods. To this end, we must formalize several core notions: function spaces, bilinear forms defined via integration, extensions of a function to a larger domain with zero, and partitions of unity. Our aim is to construct a setting where these notions are defined with only the minimal properties necessary for the analysis to follow. This abstraction enables us to isolate and understand the exact assumptions required for spectral domain decomposition methods.

In practical applications, we usually work with separable Hilbert spaces, such as the Sobolev space H^1 or their finite-dimensional discretizations. However, our framework already involves multiple bilinear forms. Introducing yet another inner product — namely, that associated with the underlying Hilbert space — would unnecessarily complicate the notation, especially since it plays no direct role in this thesis. Therefore, we instead adopt a setting based on separable Banach spaces, equipped with an underlying norm that we will rarely need to reference. While some of the results could be stated more generally, this level of abstraction is sufficient for a wide range of PDE applications. Within this setting, we adopt the following conventions:

Notation 3.1. Throughout this section, all vector spaces are assumed to be separable Banach spaces, and all maps between them are assumed to be linear and continuous, unless explicitly stated otherwise.

Before we continue, recall the definition of a topological space.

Definition 3.2. A *topological space* is a pair $(\Omega, \mathcal{O})$, where Ω is a set and $\mathcal{O}$ is a collection of subsets of Ω such that:

(i) $\emptyset \in \mathcal{O}$ and $\Omega \in \mathcal{O}$,

(ii) The union of any collection of sets in $\mathcal{O}$ is also in $\mathcal{O}$,

(iii) The intersection of any finite number of sets in $\mathcal{O}$ is also in $\mathcal{O}$.

The collection $\mathcal{O}$ is called a topology on Ω, and the sets in $\mathcal{O}$ are called open.

The *open interior* of a subset $X \subset \Omega$ is denoted by $\mathring{X}$ and is given by the union of all open sets contained by X. Further, the *boundary* ∂X of X is defined as the set of all points $x \in \Omega$ for which each open neighbourhood intersects X as well as $X^c = \Omega \backslash X$. Finally, the *closure* of X is defined as $\overline{X} = X \cup \partial X$.

3.1.1 Presheaves and Sheaves

The notion of a function space is often understood as discussing vector spaces with nice enough properties, such as separable Banach spaces — or preferably separable Hilbert spaces — as we did in Chapter 1. In reality, however, there is more structure to them: functions can be restricted to smaller subsets, and small pieces of them can potentially be glued together to form a function over a larger domain. These mechanisms come into play whenever we switch between local and global settings which is precisely the case for domain decomposition methods. The mathematical object that formalizes such behaviour is a sheaf, and the preliminary concept thereof a presheaf.

Definition 3.3 (Presheaf)**.** Let Ω be a topological space and $\mathcal{O}$ the set of all open subsets of Ω. A *presheaf* (of separable Banach spaces) H on Ω consists of a collection of separable Banach spaces $\{H(\omega)\}_{\omega\in\mathcal{O}}$ and of maps $R^H_{\omega^*,\omega} : H(\omega^*) \to H(\omega)$ for any open $\omega \subset \omega^* \subset \Omega$ such that:

(i) For each open $\omega \subset \Omega$ the map $R^H_{\omega,\omega}$ is the identity.

(ii) For all open $\omega \subset \omega^* \subset \omega^{**} \subset \Omega$ we have $R^H_{\omega^*,\omega} \circ R^H_{\omega^{**},\omega^*} = R^H_{\omega^{**},\omega}$.

The maps $R^H_{\omega^*\omega}$ are called *restrictions*, and we write

$$\cdot|_\omega := R_\omega(\cdot) := R^H_{\omega^*,\omega}(\cdot)$$

for short. Elements of $H(\omega)$ are called *sections* of H on ω.

There is an arguably more elegant definition of a presheaf as a contravariant functor. This is of little use to us here however, as we will not delve into category theory. An interested reader can find a thorough introduction to the topic in textbooks such as [60]. In this thesis, we use presheaves simply as a framework to formalize functions.

Notation 3.4. When working with a presheaf H and a section $v \in H(\omega^*)$, we say that v has a property on $\omega \subset \omega^*$ if the restriction $v|_\omega$ has that property. Further, we will occasionally omit explicitly defining open subsets ω. In essence, anything we plug into a presheaf H is assumed to be an open set, and typically $\omega \subset \omega^* \subset \Omega$.

Two further very basic properties are expected of functions. First, if functions agree locally everywhere, then they are identical globally (locality). Second, two functions can be glued together as long as they agree in any regions where both are defined (gluing). A presheaf with these properties is called a sheaf. In this thesis however, we will mostly only need (and have) locality, and thus introduce the intermediate term of a local presheaf.

Definition 3.5 (Local presheaf, sheaf)**.** Let H be a presheaf on a topological space Ω and consider the following two properties in which I is an arbitrary index set.

(i) Locality: Suppose $\{\omega_i\}_{i\in I}$ is a collection of open sets, and $v, w \in H(\cup_i\omega_i)$ with $v|_{\omega_i} = w|_{\omega_i}$ for all $i \in I$. Then $v = w$.

(ii) Gluing: Suppose $\{\omega_i\}_{i\in I}$ is a collection of open sets, and $v_i \in H(\omega_i)$ such that $v_i|_{\omega_i\cap\omega_j} = v_j|_{\omega_i\cap\omega_j}$ for all $i, j \in I$. Then there exists a $v \in H(\cup_i\omega_i)$ such that $v|_{\omega_i} = v_i$ for all i.

We say that a presheaf H is *local*, if it satisfies (i). H is called a *sheaf* if it satisfies both (i) and (ii).

Note that a sheaf assumes that any number of functions can be glued, even infinitely many. This is not true for all types of functions, as the following examples illustrate.

Example 3.6 (Sheaves of functions over $\mathbb{R}$)**.** Let Ω be a regular enough subset of $\mathbb{R}^d$.

(i) Functions on Ω form a sheaf.

(ii) Continuous functions on Ω form a sheaf.

(iii) Bounded functions on Ω form a local presheaf. This is a sheaf if Ω is compact.

(iv) Equivalence classes of square integrable functions form the local presheaf L^2 on Ω. This is a sheaf if Ω has finite measure.

(v) H^1 functions form a local presheaf on Ω. This is a sheaf if Ω has finite measure.

(vi) Continuous piecewise polynomials (on a given mesh) form a sheaf on Ω.

In the given classical notation, some of these symbols are overloaded. The L^2-presheaf over $\mathbb{R}$ differs from that over $\mathbb{R}^2$, for instance.

Another important example is the discrete sheaf. It covers in particular the algebraic viewpoint of typical PDE discretizations, as these come with a finite number of degrees of freedom, associated to a discrete set of interpolation points and spread throughout Ω.

Example 3.7 (Discrete sheaf)**.** Let I be a countable set equipped with the discrete topology. Then functions from subsets of I to $\mathbb{R}$ form a sheaf N. We will call this the discrete sheaf on I. Note that for $J \subset I$ we have $N(J) = \mathbb{R}^J \cong \mathbb{R}^{\#J}$ after choosing a basis. However, this isomorphism ignores the sheaf-theoretic information about how restrictions between subsets of I interact. Thus, typical finite element implementation keep track of local and global indices separately.

To facilitate work with presheaves, we will need the notion of a morphism between them.

Definition 3.8 (Maps between presheaves)**.** A map or morphism $\mathcal{F} : H \to L$ between presheaves H and L over Ω consists of maps $\mathcal{F}_\omega : H(\omega) \to L(\omega)$ for each open $\omega \subset \Omega$ such that they are compatible with restrictions. More precisely, such that

$$\begin{array}{ccc} H(\omega^*) & \xrightarrow{\mathcal{F}_{\omega^*}} & L(\omega^*) \\ \downarrow{\scriptstyle R^H_{\omega^*\omega}} & & \downarrow{\scriptstyle R^L_{\omega^*\omega}} \\ H(\omega) & \xrightarrow{\mathcal{F}_\omega} & L(\omega) \end{array}$$

commutes for any open $\omega \subset \omega^* \subset \Omega$.

With a notion of maps we immediately also have a notion of isomorphism — a map that has an inverse. With this we can formalize the equivalence between the geometric and the algebraic setting.

Example 3.9 (Algebraic setting)**.** This example highlights how geometric objects (finite element functions) are translated into algebraic data (elements of $\mathbb{R}^n$) through evaluation at mesh nodes. Let I be the set of vertices of a simplicial mesh resolving a domain $\Omega \subset \mathbb{R}^d$. Let N be the discrete sheaf on I and let H be the sheaf of piecewise linear, continuous functions on Ω. Then the maps

$$\begin{aligned} \Phi_\omega : H(\omega) &\to N(\overline{\omega} \cap I) \\ f &\mapsto (i \mapsto f(i)) \end{aligned}$$

are isomorphisms that map a finite element function to its representation in the Lagrange basis. The resulting map $H \to N$ is an isomorphism of sheaves. We sometimes refer to working with N as the algebraic setting, as opposed to the geometric setting H.

As a further example of maps between presheaves, consider pointwise multiplications with a function. These often come into play, for instance, when defining partitions of unity.

Example 3.10 (Pointwise multiplication)**.** Let C be the sheaf of continuous functions over a topological space Ω and let $f \in C(\Omega)$. Then pointwise multiplication with f defines a map of sheaves $C \to C$. Note that this does in general not restrict to a map $H \to H$ if we take H to be a sheaf of finite element functions. We will come back to this when discussing abstract partitions of unity.

Among maps of presheaves, embeddings play a particularly important role, as they formalize inclusion relationships.

Definition 3.11 (Embedded sheaves / subsheaves)**.** For two presheaves H, L over Ω, we say that H is *embedded* in L and write $H \subset L$ if there is a map $\mathcal{F} : H \hookrightarrow L$ such that all $\mathcal{F}_\omega$ are embeddings of separable Banach spaces. In this case, we call $\mathcal{F}$ an embedding and say that H is a subsheaf of L. The embedding is called compact if all the individual maps $\mathcal{F}_\omega$ are compact.

Example 3.12. The sheaf H^1 is embedded in L^2 on any $\Omega \subset \mathbb{R}^d$. If Ω is bounded and has Lipschitz boundary, then $H^1(\Omega) \to L^2(\Omega)$ is compact. However, this does not hold true over all open sets $\omega \subset \Omega$. This is unfortunate, as this embedding being compact is an important tool for many results in PDE theory.

We can address this with a trick as we do not need to be able to plug all open sets into the given sheaves for domain decomposition purposes. We actually only need to consider a finite number of $\{\omega_i\}_{i \in I}$ that are open in the standard topology and cover Ω — as well as possibly their unions, intersections, and complements. The easiest way to achieve this is to assume that the ω_i and Ω are resolved by a mesh. Then we can consider the topology $\mathcal{O}$ of all open (in the standard topology) sets that are likewise resolved by it. All of these will have Lipschitz boundary, so the embedding $H^1 \hookrightarrow L^2$ will be compact — as long as we consider these sheaves over the topological space $(\Omega, \mathcal{O})$.

Note that in this procedure we use the topology $\mathcal{O}$ exclusively for determining what we want to plug into the (pre-)sheaves under consideration. All other properties and constructs — such as continuity, compactness and integration — are still understood with respect to the standard topology.

Later we will construct hierarchies of successively coarser presheaves. The following proposition will be a key ingredient in this procedure.

Proposition 3.13 (Canonical subsheaf)**.** *Let L be a presheaf over Ω, and let $V \subset L(\Omega)$ be a finite-dimensional subspace. Then we can define a presheaf H with embedding $H \hookrightarrow L$ by setting $H(\omega)$ to be the image of V under the restriction. More precisely, for any open ω we set*

$$H(\omega) := R_{\Omega,\omega}(V)$$

and define the restriction maps in H as

$$R^H_{\omega^*,\omega} = R^L_{\omega^*,\omega}|_{H(\omega^*)}.$$

If L is local, so is H. We call H the canonical subsheaf of L with respect to V.

Proof. First we show that the restrictions in H are well-defined. Let $w \in H(\omega^*)$. Then by definition there is a $v \in V$ such that $R^L_{\Omega,\omega^*}(v) = w$. Thus, we have

$$R^H_{\omega^*,\omega}(w) = R^L_{\omega^*,\omega}(w) = R^L_{\omega^*,\omega} \circ R^L_{\Omega,\omega^*}(i(v)) = R^L_{\Omega,\omega}(i(v)) \in H(\omega).$$

The required properties of the $R^H_{\omega^*,\omega}$ — i.e. (i) and (ii) in Definition 3.3 — follow directly from the same properties of $R^L_{\omega^*,\omega}$. Similarly, locality of H follows directly from the locality of L. Finally, since V is finite dimensional, so are all $H(\omega)$, which are thus separable Banach spaces. □

3.1.2 Bilinear Forms on Sheaves

Next, we formalize the notion of integrals to the presheaf setting. We are looking for a generalization of bilinear forms like

$$a_\omega(v,w) = \int_\omega A\nabla v \nabla w \, \mathrm{d}x.$$

On a presheaf, we require a bilinear form for each open set ω, as made precise below.

Definition 3.14 (Local bilinear form)**.** Let H be a presheaf of vector spaces over Ω. A bilinear form $a(\cdot,\cdot)$ on H consists of a bilinear form a_ω on $H(\omega)$ for each open $\omega \subset \Omega$. Further:

(i) If X is a property of bilinear forms, we say that $a(\cdot,\cdot)$ has X if all a_ω have it individually.

(ii) We will often suppress restrictions in the arguments of $a(\cdot,\cdot)$ if they are clear from the context. For instance, for open $\omega \subset \omega^*$ and sections $v, w \in H(\omega^*)$, we write
$$a_\omega(v,w) := a_\omega(R_{\omega^*,\omega}v,\ R_{\omega^*,\omega}w).$$

(iii) We call $a(\cdot,\cdot)$ *local* if $a_\emptyset = 0$ and for any open $\omega \subset \omega^*$, it holds
$$a_{\omega^*}(v,w) = a_\omega(v,w) + a_{(\omega^* \setminus \omega)}(v,w)$$
for all sections $v, w \in H(\omega^*)$.

Locality of a bilinear form essentially ensures that it behaves like an integral on a function space, at least in the ways that we care about in this thesis. In particular, we will make use of the following properties.

Proposition 3.15. *Let $a(\cdot,\cdot)$ be a local, positive semi-definite bilinear form on a presheaf H and let $\omega \subset \omega^*$ be open. Then for any section $v \in H(\omega^*)$, we have*

$$a_\omega(v,v) \leq a_{\omega^*}(v,v).$$

Proof. By locality, we have

$$a_{\omega^*}(v,v) = a_\omega(v,v) + a_{\omega^* \setminus \omega}(v,v)$$

and both summands on the right are non-negative by positive semi-definiteness of $a(\cdot,\cdot)$. □

Definition 3.16. Let Ω be a topological space and $\{\omega_i\}_{i\in I}$ a family of open subsets. Let $\xi \in \mathbb{N} \cup \{\infty\}$ be the maximum number of them that overlap at any given point. More precisely, let

$$\xi = \sup_{x\in\Omega} |\{i \in I \mid x \in \omega_i\}|.$$

If $\xi < \infty$ we say that the family admits a finite colouring with ξ colours, or, for short, that it allows ξ-colouring.

Proposition 3.17. *Let $a(\cdot,\cdot)$ be a local, positive semi-definite form on a local presheaf H. Then for all families of open subsets $\{\omega_i\}_{i\in I}$ allowing ξ-colouring, we have*

$$\sum_{i\in I} a_{\omega_i}(v,v) \leq \xi a_\omega(v,v),$$

where $\omega = \bigcup_i \omega_i$.

Proof. First note that since $a(\cdot,\cdot)$ is positive semi-definite, all occurring terms in this discussion will be non-negative. The idea is as follows. By locality of $a(\cdot,\cdot)$, we can decompose $a_\omega(v,v)$ into contributions over regions where exactly a given subset of the $\{\omega_i\}_i$ overlap. In the worst case, such a region appears ξ times on the left-hand side of the claimed inequality, while on the right, each region appears only once. However, as discussed in Example 3.12, we intend to use non-standard topologies to define what can be plugged into H, and in these cases, things may not be so intuitive. Thus, we give a formal proof here, using induction over ξ.

For $\xi = 0$, all ω_i must be empty and since $a_\emptyset = 0$ by definition, the result is trivial. So let $\xi \in \mathbb{N}$. Assume we have shown the claim for all finite families that allow $(\xi-1)$-colouring and consider a finite family $\{\omega_i\}_{i\in I}$ that allows ξ-colouring. The idea now is to split off the part where exactly ξ of the subsets overlap. For any $J \subset I$, define

$$U_J := \bigcap_{j\in J} \omega_j, \quad \text{and let} \quad U := \bigcup_{\substack{J\subset I \\ |J|=\xi}} U_J.$$

As a union of finite intersections, U is open. We set $\widetilde{\omega_i} := (\omega_i \mathring{\setminus} U)$. Note that

$$\widetilde{\omega} := \bigcup_{i\in I} \widetilde{\omega_i} = (\omega \mathring{\setminus} U),$$

which uses the set-topological fact that for any subsets $A_i, B \subset \Omega$ with each A_i open, we have

$$(\bigcup_{i\in I} \mathring{A}_i \setminus B) = \bigcup_{i\in I} (A_i \mathring{\setminus} B).$$

Now by locality of the $a(\cdot,\cdot)$, we get

$$\begin{aligned} a_\omega(v,v) &= a_{\widetilde{\omega}}(v,v) + a_U(v,v), \\ a_{\omega_i}(v,v) &= a_{\widetilde{\omega_i}}(v,v) + a_{U\cap\omega_i}(v,v), \end{aligned}$$

and since at most $\xi - 1$ of the $\widetilde{\omega_i}$ overlap, it holds that

$$\sum_{i\in I} a_{\widetilde{\omega_i}}(v,v) \leq (\xi-1) a_{\widetilde{\omega}}(v,v).$$

Next, note that the U_J in the definition of U are pairwise disjoint, since otherwise we would have an $x \in \Omega$ where at least $\xi+1$ of the ω_i overlap. Thus, the U_J split U into disconnected parts, and we have by locality of $a(\cdot,\cdot)$ that

$$a_U(v,v) = \sum_{J\in\mathcal{I}} a_{U_J}(v,v) \quad \text{and} \quad a_{U\cap\omega_i}(v,v) = \sum_{J\in\mathcal{I}} a_{U_J\cap\omega_i}(v,v),$$

where $\mathcal{I} := \{J \subset I \mid |J| = \xi\}$. Now note that

$$U_J \cap \omega_i = \begin{cases} U_J, & \text{if } i \in J, \\ \emptyset & \text{else,} \end{cases}$$

and that the first case is true for exactly ξ indices $i \in I$. Thus, we have

$$\sum_{i \in I} a_{U \cap \omega_i}(v, v) = \sum_{i \in I} \sum_{J \in \mathcal{I}} a_{U_J \cap \omega_i}(v, v) = \xi \sum_{J \in \mathcal{I}} a_{U_J}(v, v) = \xi a_U(v, v).$$

Summing up we have found that

$$\begin{aligned} \sum_{i \in I} a_{\omega_i}(v, v) &= \sum_{i \in I} a_{\widetilde{\omega_i}}(v, v) + \sum_{i \in I} a_{U \cap \omega_i}(v, v) \\ &\leq \xi a_{\widetilde{\omega}}(v, v) + \xi a_U(v, v) \\ &= \xi a_\omega(v, v). \end{aligned}$$

□

Remark 3.18. Note that if $F : H \hookrightarrow L$ is an embedding of presheaves and $a(\cdot, \cdot)$ is a bilinear form on L, then the composition $a(\cdot, \cdot) \circ (F \times F)$ defines a bilinear form on H. We usually continue to denote this bilinear form by $a(\cdot, \cdot)$, or refer to it as the restriction of $a(\cdot, \cdot)$ to H. This restriction inherits many properties from its parent. For instance, if $a(\cdot, \cdot)$ is local, positive semi-definite or symmetric, then so is its restriction to any subsheaf.

3.1.3 Extendables

We now need a generalized version of $H_0^1(\omega)$. Notably, this space is a subspace of $H^1(\omega)$, but can also be considered as a subspace of $H^1(\Omega)$ by extending with zero. In the abstract framework we define such extensions as the a-adjoint of the restrictions with additional properties, as detailed below.

Definition 3.19 (Extendables)**.** Let H be a presheaf on a topological space $(\Omega, \mathcal{O})$, with a symmetric, positive semi-definite bilinear form $a(\cdot, \cdot)$ on H. Then we call a family $\{H_0(\omega)\}_{\omega \in \mathcal{O}}$ of closed subspaces $H_0(\omega) \subset H(\omega)$ *extendables* for H, if the following hold for any open $\omega \subset \omega^*$.

(i) The bilinear form a_ω is coercive on $H_0(\omega)$.

(ii) There are (continuous and linear) maps

$$\iota_{\omega, \omega^*} : H_0(\omega) \to H_0(\omega^*)$$

that are right-inverses to the restrictions, and at the same time their a-adjoints. More precisely, we have

$$R_{\omega^*, \omega} \circ \iota_{\omega, \omega^*} = \mathrm{Id}_{H_0(\omega)},$$

and for all $v \in H_0(\omega)$ it holds that

$$a_{\omega^*}(\iota_{\omega, \omega^*} v, w) = a_\omega(v, R_{\omega^*, \omega} w) \text{ for all } w \in H(\omega^*). \tag{3.2}$$

We call the triple (H, H_0, a) a *presheaf with extendables*. The ι_{ω,ω^*} are called *extensions*, and we often write just ι for short. If multiple presheaves are in play, we sometimes add an upper index ι^H to clearly mark the extensions in H.

Note that H_0 will usually not be a subsheaf of H since restrictions of functions in H_0 do typically not lie in H_0 again. To see this, consider the ongoing example of the diffusion equation with continuous finite elements, where the ι are given by extending with zero.

Proposition 3.20. *Let (H, H_0, a) be a presheaf with extendables. Then the following hold for any open $\omega \subset \omega^* \subset \omega^{**}$.*

(i) The extension ι_{ω,ω^*} is unique.

(ii) The extension ι_{ω,ω^*} is an a-isometry.

(iii) The extensions are compatible with the sheaf structure, i.e. $\iota_{\omega,\omega^{**}} = \iota_{\omega^*,\omega^{**}} \circ \iota_{\omega,\omega^*}$.

Proof. For (i) set $X = \left(R_{\omega^*,\omega}\right)^{-1}\left(H_0(\omega)\right)$ and note that X is closed as the preimage of a closed set under a continuous map. In particular, X is a Hilbert space and thus ι_{ω,ω^*} must be the unique adjoint of the restriction

$$R_{\omega^*,\omega}|_X : X \to H_0(\omega).$$

For (ii), note that since ι is a right-inverse of R as well as its adjoint, we have

$$a_{\omega^*}(\iota v, \iota w) = a_\omega(R\iota v, w) = a_\omega(v, w)$$

for any $v, w \in H_0(\omega)$. Finally, (iii) follows from the ι being adjoints of the restrictions, since we have

$$R_{\omega^{**},\omega} = R_{\omega^*,\omega} \circ R_{\omega^{**},\omega^*}$$

and thus the statement follows from the usual interplay of concatenations and adjoints. □

Notation 3.21. By the above proposition, the ι are injective isometries that are compatible with the overall structure. Thus, we can think of $H_0(\omega)$ as a subspace of $H_0(\omega^*)$ for any $\omega^* \supset \omega$. We will often do so, suppressing the ι for ease of readability, unless we want to emphasize their existence and properties.

Proposition 3.22. *In the setting of Definition 3.19, let additionally $a(\cdot,\cdot)$ be local. Then ι fulfills (3.2) if and only if for all $v \in H_0(\omega)$ we have*

$$a_{(\omega^* \setminus \overline{\omega})}(\iota v, w) = 0 \text{ for all } w \in H(\omega^*).$$

Proof. By locality of $a(\cdot,\cdot)$, we have for all $v \in H_0(\omega)$ and $w \in H(\omega^*)$ that

$$a_{(\omega^* \setminus \overline{\omega})}(\iota v, w) = a_{\omega^*}(\iota v, w) - a_\omega(v, w).$$

So the left-hand side is zero if and only if the two terms on the right-hand side are identical. The latter is equivalent (3.2). □

In particular, if $a(\cdot,\cdot)$ is local, and we are working with a sheaf of functions, then extending with zero will always satisfy the demands on the ι. To make zero extensions well-defined, we will have to choose $H_0(\omega)$ appropriately of course, i.e. with homogeneous Dirichlet boundary conditions on the interior boundary $\partial\omega \cap \Omega$.

Taking this further, extensions are well-defined for PDE applications as long as we equip the spaces $H_0(\omega)$ with any boundary conditions for which the equation in Proposition 3.22 is well-defined, which corresponds to solving the given PDE on $(\omega^* \mathring{\setminus} \omega)$ with zero source term.

A crucial ingredient in domain decomposition methods are the local solves, given algebraically by orthogonal projections. A generalization of these to the presheaf setting can be defined as follows.

Definition 3.23 (Orthogonal projections)**.** Let (H, H_0, a) be a presheaf with extendables. Then for any open $\omega \subset \omega^*$, we define the composite map

$$\pi_{\omega^*,\omega} : H_0(\omega^*) \xrightarrow{\pi} \iota(H_0(\omega)) \xrightarrow{\iota^{-1}} H_0(\omega),$$

where π is the a-orthogonal projection from $H_0(\omega^*)$ onto $\iota(H_0(\omega))$. In particular, $\pi_{\omega^*,\omega}(v)$ is uniquely defined by

$$a_\omega(\pi_{\omega^*,\omega} v, w) = a_{\omega^*}(v, \iota w) \text{ for all } w \in H_0(\omega).$$

We refer to $\pi_{\omega^*,\omega}$ as the orthogonal projection from $H_0(\omega^*)$ to $H_0(\omega)$ and will often omit the subscripts for better readability.

Note that by definition of the orthogonal projections, we have $\pi \circ \iota = \mathrm{Id}$. So π is another left-inverse to the extension ι, besides the restriction R. Moreover, we have $R \circ \pi = R \circ \iota \circ \pi = \pi$ since R is a left-inverse of ι. The following proposition states that also $\pi \circ R = \pi$ whenever this concatenation is defined.

Proposition 3.24. *Let (H, H_0, a) be a presheaf with extendables over Ω. Then for any open $\omega \subset \omega^* \subset \omega^{**}$ we have*

$$\pi_{\omega^{**},\omega} = \pi_{\omega^*,\omega} \circ R_{\omega^{**},\omega^*}.$$

In particular, it holds that

$$\pi_{\Omega,\omega} = \pi_{\omega,\omega} \circ R_{\Omega,\omega}.$$

Proof. Let $v \in H_0(\omega^{**})$ and $w \in H_0(\omega)$ Then

$$\begin{aligned} a_\omega(\pi_{\omega^*,\omega}\left(R_{\omega^{**},\,\omega^*} v\right),\, w) &= a_{\omega^*}\left(R_{\omega^{**},\omega^*} v,\, \iota_{\omega,\omega^*} w\right) \\ &= a_{\omega^{**}}\left(v,\, \iota_{\omega^*,\omega^{**}}\left(\iota_{\omega,\omega^*} w\right)\right) \\ &= a_{\omega^{**}}\left(v,\, \iota_{\omega,\omega^{**}} w\right), \end{aligned}$$

and thus $\pi_{\omega^*,\omega} \circ R_{\omega^{**},\omega^*}$ satisfies the unique property of $\pi_{\omega^{**},\omega}$. □

The above proposition essentially states that the projections are compatible with the general sheaf structure. The special case for $\pi_{\Omega,\omega}$ is important, since these projections will later correspond to local PDE solves. Above, we see that under relatively mild assumptions these can indeed be solved locally, i.e. after restricting to $H(\omega)$.

Proposition Proposition 3.24 in particular states that orthogonality is kept intact under restrictions and extensions. The following corollary gives a special case of this general statement which we will need in later chapters.

Corollary 3.25. *Let (H, H_0, a) be a presheaf with extendables. Then for any $v \in H(\omega^*)$ we have*

$$a_\omega(v - \pi v,\, w) = 0 \text{ for all } w \in H_0(\omega)$$

where $\pi = \pi_{\omega^*,\omega}$.

Proof. For $v \in H(\omega^*)$, $w \in H_0(\omega)$ and being precise with the restriction $R = R_{\omega^*,\omega}$ it holds that

$$a_\omega(v - \pi v,\ w) = a_\omega(Rv - R\iota\pi v,\ w) = a_{\omega^*}(v - \iota\pi v,\ \iota w) = 0,$$

where for the last equality we used that $\iota\pi$ is the orthogonal projection onto $\iota(H_0(\omega))$. □

For standard subspace correction theory, we will need a strengthened Cauchy-Schwarz inequality. This is satisfied under mild assumptions as we show below.

Proposition 3.26 (Strengthened Cauchy-Schwarz inequality)**.** *Let (H, H_0, a) be a presheaf with extendables and let $a(\cdot,\cdot)$ be local. Let further $\omega_1, \omega_2 \subset \omega^*$ be open. Then for any $v_1 \in H_0(\omega_1), v_2 \in H_0(\omega_2)$ we have*

$$||a_{\omega^*}(\iota v_1, \iota v_2)| \leq \epsilon \sqrt{a_{\omega^*}(\iota v_1, \iota v_1)} \sqrt{a_{\omega^*}(\iota v_2, \iota v_2)}$$

with

$$\epsilon = \begin{cases} 0 & \text{if } \omega_1 \cap \omega_2 = \emptyset \\ 1 & \text{otherwise.} \end{cases}$$

Proof. With $\epsilon = 1$ this is simply the Cauchy-Schwarz inequality. The case $\epsilon = 0$ follows from locality of $a(\cdot,\cdot)$ since

$$a_{\omega^*}(\iota v_1, \iota v_2) = a_{(\omega^* \setminus \omega_1)}(\iota v_1, \iota v_2) + a_{(\omega^* \setminus \omega_2)}(\iota v_1, \iota v_2)$$

and both terms on the right are zero by Proposition 3.22. □

Finally, we need to extend the notion of maps of presheaves to maps of presheaves with extendables.

Definition 3.27 (Maps of presheaves with extendables)**.** Let (H, H_0, a) and (L, L_0, b) be presheaves with extendables. Then a map of presheaves with extendables

$$\mathcal{F} : (H, H_0, a) \to (L, L_0, b)$$

is a map of presheaves $\mathcal{F} : H \to L$ that is compatible with extensions and the bilinear forms, more precisely we demand that for all open ω that

(i) It holds that $\mathcal{F}_\omega(H_0(\omega)) \subset L_0(\omega)$.

(ii) The following diagram commutes

$$\begin{array}{ccc} H(\omega^*) & \xrightarrow{\mathcal{F}_{\omega^*}} & L(\omega^*) \\ {\scriptstyle \iota^H_{\omega,\omega^*}}\uparrow & & {\scriptstyle \iota^L_{\omega,\omega^*}}\uparrow \\ H_0(\omega) & \xrightarrow{\mathcal{F}_\omega} & L_0(\omega). \end{array}$$

(iii) We have $a_\omega(\cdot,\cdot) = b(\mathcal{F}_\omega \cdot,\ \mathcal{F}_\omega \cdot)$.

If $\mathcal{F}$ is additionally an embedding, we write $(H, H_0, a) \subset (L, L_0, b)$.

Notation 3.28. Given a presheaf with extendables (H, H_0, a) over Ω, open subsets $\omega \subset \omega^* \subset \Omega$, and $v \in H(\omega^*)$, recall that we may abbreviate

$$v|_\omega = R_{\omega^*,\omega}(v).$$

We will further write

$$\|v\|_{H(\omega)} := \|v|_\omega\|_H := \|v|_\omega\|_{H(\omega)}$$

for the underlying norm of $v|_\omega$ in the Banach space $H(\omega)$. This norm will rarely be used in this thesis. Instead, we usually consider

$$\|v\|_{a_\omega} := \|v|_\omega\|_a := (a_\omega(v|_\omega, v|_\omega))^{1/2},$$

the semi-norm induced by $a(\cdot,\cdot)$, also called the *energy norm*. As previously mentioned, we often abbreviate

$$a_\omega(v,w) := a_\omega(v|_\omega, w|_\omega)$$

whenever v, w are sections of H on supersets of ω. Finally, we will often identify elements $v \in H_0(\omega)$ with their extension in $\iota v \in H(\omega^*)$, suppressing the ι in the notation.

3.1.4 Partition of Unity

In the following sections, we will frequently encounter situations where we would like to glue together locally defined functions that do not necessarily agree on the overlaps of their domains. Since there is no canonical way to perform this gluing, we must make a specific choice.

More precisely, given an open cover $\Omega = \bigcup_i \omega_i$ and a local presheaf H on Ω, we seek a map Glue that acts as a left-inverse to the family of restriction maps, as illustrated in the diagram below:

$$H(\Omega) \underset{\text{Glue}}{\overset{\text{Restrict}}{\rightleftarrows}} \prod_{i\in I} H(\omega_i)$$

Here, Restrict $:= \prod_i R_{\Omega,\omega_i}$ denotes the product of the appropriate restrictions in H. Requiring that Glue is a left-inverse means that

$$\text{Glue} \circ \text{Restrict} = \text{id}_{H(\Omega)}.$$

If the sheaf comes equipped with extendables, the construction of such a gluing map can be formalized using the concept of a partition of unity.

Definition 3.29 (Partition of unity for presheaves)**.** Let (H, H_0, a) be a local presheaf with extendables over Ω, and let $\{\omega_i\}_{i\in I}$ be an open cover, i.e. $\Omega = \bigcup_{i\in I} \omega_i$. A *partition of unity* subordinate to this cover is a collection of maps

$$\chi_i : H(\omega_i) \to H_0(\omega_i)$$

such that the following conditions hold:

(i) The composite map

$$\text{Glue} : \prod_{i\in I} H(\omega_i) \xrightarrow{\prod_{i\in I}\chi_i} \prod_{i\in I} H_0(\omega_i) \xrightarrow{\sum_{i\in I}\iota_{\omega_i}} H(\Omega)$$

is well-defined, where $\iota_{\omega_i} : H_0(\omega_i) \to H(\Omega)$ denotes the extensions in H.

(ii) The maps χ_i act as a partition of unity in the sense that

$$\text{Glue} \circ \text{Restrict} = \text{id}_{H(\Omega)},$$

where $\text{Restrict} := \prod_{i \in I} R_{\Omega,\omega_i}$.

For any $\omega^* \supset \omega_i$, we often abbreviate $\chi_{\omega^*,\omega_i} := \chi_i \circ R_{\omega^*,\omega_i}$ and may simply write χ_i for χ_{ω^*,ω_i} when there is no risk of confusion.

Condition (i) in Definition 3.29 addresses the possibility that the sum may be infinite, which could lead to it being undefined. For finite covers, this issue does not arise, and condition (i) is therefore trivially satisfied. We include the more general infinite case solely to align with standard definitions of partitions of unity found in the literature, where they are typically formulated in terms of functions, as described below.

Definition 3.30 (Partition of unity of functions)**.** Let Ω be a topological space with an open cover $\{\omega_i\}_{i \in I}$. A *partition of unity of functions* $\{\phi_i\}_{i \in I}$ subordinate to this cover is a collection of functions $\phi_i : \Omega \to [0, 1]$ such that for every $x \in \Omega$:

(i) There exists a neighbourhood of x in which all but finitely many of the ϕ_i vanish.

(ii) We have $\sum_{i \in I} \phi_i = 1$, where 1 denotes the constant function with value 1.

(iii) The support of each ϕ_i is contained in ω_i, i.e. $\text{supp}(\phi_i) \subset \omega_i$.

We say that $\{\phi_i\}_{i \in I}$ is continuous, differentiable, or measurable if each ϕ_i individually has the respective property.

There are general theorems that guarantee the existence of partitions of unity of functions with desirable properties in a wide range of settings. However, these results are not directly relevant for our purposes, as we will only consider finitely many subdomains ω_i covering either pieces of $\mathbb{R}^d$ or discrete index sets. Nevertheless, for completeness, we briefly mention a classic existence theorem. This result is commonly found in textbooks on differential topology and geometry, where partitions of unity are used to extend concepts from $\mathbb{R}^n$ to manifolds, such as integration or (Riemannian) metrics.

Theorem 3.31. *Given a paracompact smooth manifold Ω with any open cover $\{\omega_i\}_{i \in I}$, there exists a partition of unity of smooth functions, subordinate to this cover.*

Proof. See for instance [40]. □

For typical presheaves, it is often straightforward to define an abstract partition of unity based on a given partition of unity of functions as illustrated by the following examples.

Example 3.32. Consider the presheaf with extendables $(L^2, L^2, (\cdot,\cdot)_{L^2})$ over open $\Omega \subset \mathbb{R}^d$ with cover $\{\omega_i\}_{i \in I}$ and subordinate partition of unity of measurable functions $\{\phi_i\}_{i \in I}$. Then the maps

$$\begin{aligned} \chi_i : L^2(\omega_i) &\to L^2(\omega_i) \\ f &\mapsto (x \mapsto \phi_i(x) f(x)) \end{aligned}$$

give a partition of unity subordinate to the given cover.

Proof. By definition, the functions ϕ_i are bounded, so multiplying by them preserves square-integrability. Furthermore, property (i) ensures that, for any $x \in \Omega$, only finitely many terms in the sum of Definition 3.29 are non-zero. Property (ii) guarantees that the maps χ_i sum to the identity, while (iii) ensures that each χ_i is well-defined — that is, their images lie in $\iota(L^2(\omega_i))$, which we naturally identify with $L^2(\omega_i)$. □

Intuitively, such a partition of unity corresponds to taking a weighted average. If the functions ϕ_i possess properties that are preserved under addition and multiplication (such as continuity), then the χ_i and the Glue map will preserve these properties as well. Consequently, the same strategy can be used to define partitions of unity for sheaves of continuous functions, and similar settings. However, certain cases require more care. For instance, consider continuous piecewise linear functions on a simplicial mesh. The product of two such functions results in a piecewise polynomial of degree 2. In this case, we must introduce an interpolation operator and verify that the resulting partition of unity still sums to 1.

Example 3.33 (Partition of Unity for Finite Elements)**.** Consider a conforming simplicial mesh resolving a domain $\Omega \subset \mathbb{R}^d$, and let (H, H_0, a) be the presheaf of continuous piecewise linear functions, with $a(\cdot,\cdot)$ the energy form of the diffusion equation and H_0 the subspaces with homogeneous Dirichlet boundary conditions on interior boundaries. Let $\{\omega_i\}_{i\in I}$ be a finite cover of Ω resolved by the given mesh, and let $\{\phi_i\}_{i\in I}$ be a subordinate partition of unity of functions. The maps

$$\begin{aligned} \chi_i : H(\omega) &\to H_0(\omega) \\ f &\mapsto (x \mapsto I_h(\phi_i \cdot f)(x)) \end{aligned}$$

define a partition of unity subordinate to the given cover, where I_h is the Lagrange interpolation operator and $\phi_i \cdot f$ denotes the pointwise multiplication of functions.

Proof. The primary task is to verify that the χ_i still sum to the identity after interpolation, i.e. that $\sum_i \chi_i v = v$ for all $v \in H(\Omega)$. To see this, observe that at the vertices, the interpolation is exact, and hence the finite element function $\sum_i \iota\chi_i v$ equals v at these points. Since the values at the vertices uniquely determine v, it follows that $\sum_i \chi_i v = v$ everywhere. Note that no regularity of the ϕ_i is required for this argument. □

Remark 3.34. In both Example 3.32 and Example 3.33, the partition of unity maps $\chi_i : H(\omega_i) \to H(\omega_i)$ can be extended to maps $H \to H$ of presheaves. That is, for each ω, we can define maps $\chi_{i,\omega} : H(\omega) \to H_0(\omega)$ that are compatible with the restriction structure in H, and satisfy $\chi_{i,\omega_i} = \chi_i$. This is generally expected when the partition of unity is based on pointwise multiplication with scalar-valued functions, as is the case in these examples.

Viewing a partition of unity as a collection of maps of presheaves in this manner has many advantages. However, this viewpoint is not applicable to the canonical partition of unity for generalized elements that we establish in Section 3.3.3.

Before proceeding to the next section, we introduce the concept of ξ-colouring for a partition of unity. The key observation is that the colouring constant associated with the supports of a partition of unity can be smaller than that of the original cover $\omega_{ii\in I}$. This is in particular the case for spectral methods with oversampling.

Definition 3.35 (Support)**.** Let (H, H_0, a) be a presheaf with extendables over Ω and let $\omega \subset \Omega$ be open. Consider a map $f : H(\omega) \to H(\omega)$. Then we say that f has

support in an open set $\operatorname{supp}(f) \subset \omega$ if there is a map

$$\hat{f} : H(\operatorname{supp}(f)) \to H_0(\operatorname{supp}(f))$$

such that

$$f = \iota_{\operatorname{supp}(f),\omega} \circ \hat{f} \circ R_{\omega,\operatorname{supp}(f)}.$$

We call $\operatorname{supp}(f)$ a supporting set of f.

Example 3.36. If $f : H(\omega) \to H(\omega)$ is given by pointwise multiplication with a continuous function $\phi : \omega \to \mathbb{R}$, a supporting set for f can be chosen straightforwardly as

$$\operatorname{supp}(f) = \operatorname{supp}(\phi).$$

This set is open, as it is the preimage of the open set $\mathbb{R} \setminus \{0\}$ under the continuous function ϕ.

Definition 3.37 (ξ-colouring of a partition of unity)**.** Let (H, H_0, a) be a presheaf with extendables over Ω, and let $\{\omega_i\}_{i\in I}$ be an open cover with a subordinate partition of unity $\{\chi_i\}_{i\in I}$. We say that the partition of unity admits *ξ-colouring* if there is a family $\{\operatorname{supp}(\chi_i)\}_{i\in I}$ that admits ξ-colouring, where for each $i \in I$, the set $\operatorname{supp}(\chi_i)$ is a supporting set of χ_i.

3.1.5 Suitable Presheaves

Building on the previously defined properties of presheaves, we now define the central concept of our abstract framework.

Definition 3.38 (Suitable presheaf)**.** A local presheaf with extendables (H, H_0, a) over a topological space Ω is called a *suitable presheaf* if $a(\cdot,\cdot)$ is local, all restrictions are surjective, and $H(\Omega) = H_0(\Omega)$. As with all presheaves in this thesis, suitable presheaves will always be assumed to be presheaves of (at least) separable Banach spaces unless otherwise specified.

As a presheaf with extendables, a suitable presheaf offers restrictions and extensions that are compatible with each other. Beyond that, a key property is the locality of $a(\cdot,\cdot)$, which we so far kept separate to clarify where exactly it comes in. This is the case in particular in Proposition 3.17 and Proposition 3.26, which will be crucial in later chapters.

The remaining requirements, i.e. $H(\Omega) = H_0(\Omega)$ and surjective restrictions, are less critical. If they are initially not satisfied, a suitable presheaf can often still be obtained by considering the canonical subsheaf with respect to $H_0(\Omega) \subset H(\Omega)$. In PDE settings, this usually corresponds to imposing global boundary conditions on all involved functions spaces as examples will illustrate in the following section.

3.2 Examples

To demonstrate that typical PDE applications fit into the abstract framework, this section discusses how H^1 functions, continuous finite elements and a Discontinuous Galerkin (DG) discretization give a suitable presheaf with bilinear forms arising from elliptic PDE problems.

As a model problem, we consider the heterogeneous diffusion equation introduced in Section 1.3.1, with the bilinear form

$$a(v, w) = \int (A\nabla v) \cdot \nabla w \, \mathrm{d}x, \tag{3.3}$$

defined on a Lipschitz domain $\Omega \subset \mathbb{R}^d$ that is resolved by a given mesh. We call this the energy form in the following.

To ensure the proper definition of boundary conditions and local problems, we restrict the class of open sets admissible in the sheaf H. In particular, all such sets should be regular enough to support well-defined boundary conditions. To this end, we introduce a custom topology $\mathcal{O}$ on Ω for each setting. In the continuous cases we can simply take the topology made up of all standard-open sets resolved by a given mesh. For the DG discretization we work over the discrete topology on the set of mesh faces instead. Note that these topologies serve only to define the domains ω for which the presheaf structure must be specified — for all other matters the standard topologies are used.

3.2.1 Continuous Settings

In the continuous setting of H^1 functions or subsheaves thereof, let $\mathcal{O}$ denote the topology consisting of all standard-open sets that are resolved by a given mesh. Note that $\mathcal{O}$ is indeed a topology, as it is closed under finite intersections and unions. This topology is introduced primarily to ensure that all ω used in the subsequent sheaves have Lipschitz boundaries. The topology $\mathcal{O}$ will only be employed for this purpose. In particular, by boundaries we continue to refer to the boundaries with respect to the standard topology.

Example 3.39 (H^1 setting)**.** Consider the sheaf of H^1 functions over $(\Omega, \mathcal{O})$, with a mixture of homogeneous Dirichlet and Neumann conditions on $\partial\Omega$. For extendables $H^1_0(\omega) \subset H^1(\omega)$, consider functions that additionally have zero trace on the interior boundary $\partial\omega \cap \Omega$. Then (H^1, H^1_0, a) is a local presheaf with extendables, where the extensions ι_{ω,ω^*} are given by extending functions with zero outside ω. In particular, the energy form $a(\cdot,\cdot)$ is symmetric, positive semi-definite on H^1, and coercive on H^1_0. Moreover, $a(\cdot,\cdot)$ is local. Indeed, for open $\omega \subset \omega^*$, we have

$$\begin{aligned} a_{\omega^*}(v, w) &= \int_\Omega (A\nabla v) \cdot \nabla w \, dx \\ &= \int_\omega (A\nabla v) \cdot \nabla w \, dx + \int_{\omega^* \setminus \mathring{\omega}} (A\nabla v) \cdot \nabla w \, dx \\ &= a_\omega(v, w) + a_{\omega^* \setminus \mathring{\omega}}(v, w) \end{aligned}$$

for all $v, w \in H^1(\omega^*)$.

Note that here, the presheaf H^1 includes global boundary conditions without explicit mention in the notation. This illustrates the effect of assuming surjective

restrictions and $H(\Omega) = H_0(\Omega)$ in the definition of a suitable presheaf: in PDE applications these assumptions typically correspong to always enforcing global boundary conditions on all involved functions spaces.

Example 3.40 (Finite Element setting)**.** Let $V \subset H^1(\Omega)$ be the subspace of continuous finite elements on the given mesh and let H be the canonical subsheaf of H^1 with respect to V. Then with $H_0(\omega) = H_0^1(\omega) \cap H(\omega)$, the triple (H, H_0, a) is a local presheaf with extendables. Indeed, extending with zero continues to work in H, and the restriction of $a(\cdot, \cdot)$ to H inherits symmetry, positive semi-definiteness and locality from H^1, as well as coercivity on H_0.

Remark 3.41 (Linear Elasticity)**.** For linear elasticity in the setting of $H = \left(H^1\right)^d$, the arguments are exactly the same as for the diffusion equation. Equipping H with a mixture of homogeneous Dirichlet and Neumann conditions on global boundaries, and H_0 with homogeneous Dirichlet boundary conditions on interior boundaries, turns (H, H_0, a) into a suitable presheaf. Here, $a(\cdot, \cdot)$ is the energy form of linear elasticity, as described in Section 1.3.2. Continuous finite element discretizations are again a canonical subsheaf of the infinite-dimensional case, and thus inherit its properties.

3.2.2 Discontinuous Galerkin

Discontinuous Galerkin (DG) methods approximate the solution to a given PDE with piecewise polynomials that are not required to be continuous between mesh cells. This leads to challenges in defining the gradient of such functions, which are resolved by introducing penalty terms on the mesh faces.

Here the construction of a suitable presheaf becomes more intricate. In particular, locality of $a(\cdot,\cdot)$ is a challenge, since sets of measure zero contribute to the bilinear form in this setting. We show how this can be resolved in the case of the Weighted Symmetric Interior Penalty (WSIP) Discontinuous Galerkin (DG) method from [29], applied to the heterogeneous diffusion equation.

The goal of this section is to construct a suitable presheaf (H, H_0, a^h) where for any open ω the space $H(\omega)$ consists of piecewise polynomials (which are allowed to be discontinuous), and a^h is the WSIP discretization of the bilinear form

$$a(v,w) = \int (A\nabla v)\cdot\nabla w \,\mathrm{d}x.$$

Assume that the domain $\Omega \subset \mathbb{R}^d$ is resolved by an affine conforming mesh $\mathcal{T}$. To be precise, we discuss DG functions over the disjoint union of mesh cells, rather than viewing them as "two-valued functions" defined directly on Ω.

Definition 3.42 (Disjoint union)**.** For any index set I and a family of sets $\{X_i\}_{i\in I}$, we define the disjoint union as

$$\bigsqcup_{i\in I} X_i := \bigcup_{i\in I} \{i\}\times X_i.$$

Further, the disjoint union of maps is defined in the obvious manner. More precisely, given $f_i : X_i \to Z$ for any set Z, let

$$\bigsqcup_{i\in I} f_i : \bigsqcup_{i\in I} X_i \to Z, \quad (i,x)\mapsto f_i(x).$$

With this, consider the projection P from the disjoint union of mesh cells onto Ω:

$$P : \dot\Omega := \bigsqcup_{\tau\in\mathcal{T}} |\tau| \to \bigcup_{\tau\in\mathcal{T}} |\tau| = \Omega,$$

which is given by the disjoint union of the identities on the cells $\tau \in \mathcal{T}$. In the following, when writing $|T|$ for the geometric realization of a union of mesh cells $T \subset \mathcal{T}$, this always refers to the realization in the disjoint union $\dot\Omega$, unless otherwise specified. Moreover, set brackets will often be omitted if they contain a single object. For instance, we may abbreviate the geometric realization of a singleton set $\{\tau\}$ as $|\tau|$ for readability.

Further, for any given cell $\tau \in \mathcal{T}$, we denote by $\mathbb{P}_k(|\tau|)$ an appropriate polynomial space of degree k for DG methods. For instance, polynomials of degree k for simplicial τ, or polynomials with maximum degree k with respect to a reference element for quadrilateral τ. Using this, the DG space is defined as

$$V_{DG} = \bigsqcup_{\tau\in\mathcal{T}} \mathbb{P}_k(|\tau|) \subset \{\text{functions on } \dot\Omega\}.$$

One might initially expect that restricting the elements of V_{DG} to subsets of mesh cells would suffice to define a presheaf. However, the bilinear form a^h has significant contributions from face integrals. This obstructs locality of a^h unless it is properly

accounted for. Our approach is to consider H as a presheaf over the set of faces, rather than over Ω or $\mathcal{T}$.

By $\mathcal{F}$ we denote the set of faces in the mesh $\mathcal{T}$, and by $\mathcal{F}(\tau)$ the set of faces of an individual mesh cell $\tau \in \mathcal{T}$. Moreover, for a set of faces $\omega \subset \mathcal{F}$ let $\mathcal{T}(\omega)$ denote the set of mesh cells that are adjacent to any of the faces in ω, i.e.

$$\mathcal{T}(\omega) := \{\tau \in \mathcal{T} \mid \mathcal{F}(\tau) \cap \omega \neq \emptyset\}.$$

Let $\mathcal{O} = \mathcal{P}(\mathcal{F})$ be the discrete topology on $\mathcal{F}$ consisting of all subsets of $\mathcal{P}(\mathcal{F})$.

Definition 3.43. In the notation from above, let the DG presheaf H on $(\mathcal{F}, \mathcal{O})$ be defined by setting

$$H(\omega) := \bigsqcup_{\tau \in \mathcal{T}(\omega)} \mathbb{P}_k(|\tau|) \subset \{\text{functions on } \bigsqcup_{\tau \in \mathcal{T}(\omega)} |\tau|\}$$

for any $\omega \subset \mathcal{F}$. Restrictions in H are given by the usual restrictions of functions. For a single $F \in \mathcal{F}$ we often write $H(F) := H(\{F\})$ for short.

Notation 3.44. Throughout this subsection, H will always denote the DG presheaf.

Remark 3.45. The DG presheaf is a local presheaf. However, gluing is not possible, so it is not a sheaf. To see this, consider two single faces $F_1, F_2 \in \mathcal{F}$ that bound a common cell $\tau \in \mathcal{T}$. Let $0 = v_1 \in H(F_1)$ and let $v_2 \in H(F_2)$ be constant 1. Then $v_1 = v_2$ on $\{F_1\} \cap \{F_2\} = \emptyset$. However, they cannot be glued as required by Definition 3.5 since they assign incompatible values on the shared cell τ.

Next, we need to define the bilinear forms a^h_ω for any $\omega \subset \mathcal{F}$. Each of them must be positive semi-definite, not just the global form $a^h_{\mathcal{F}}$. Thus, we will be careful while developing this.

Corresponding to the disjoint union $\dot{\Omega}$, consider the disjoint union of faces and the projection

$$P : \dot{\mathcal{F}} = \bigsqcup_{\tau \in \mathcal{T}} \mathcal{F}(\tau) \to \bigcup_{\tau \in \mathcal{T}} \mathcal{F}(\tau)$$

where in $\dot{\mathcal{F}}$ all faces appear exactly once (exterior faces) or twice (interior faces). For an $F \in \dot{\mathcal{F}}$ we write $\tau(F)$ to denote the unique mesh cell that F belongs to. Further, by $\dot{\mathcal{F}}(T)$ we denote all faces in $\dot{\mathcal{F}}$ that belong to at least one cell in $T \subset \mathcal{T}$.

If $P(F_1) = P(F_2)$ for two faces $F_1, F_2 \in \dot{\mathcal{F}}$, we say they are neighbours. In the following we denote the neighbour of any interior face F by $\hat{F}$. Moreover, we assign a sign $\sigma(F) \in \{-1, +1\}$ to all faces $F \in \dot{\mathcal{F}}$ such that $\sigma(F) = -\sigma(\hat{F})$.

Given a set of faces $U \subset \dot{\mathcal{F}}$, let $|U|$ denote the set of all points in $\dot{\Omega}$ that lie on a face in U, but are not a vertex. Note that all $x \in |\dot{\mathcal{F}}|$ lying on an interior face F have a unique neighbour, namely the only element in $\left(P^{-1}(P(x))\right) \setminus \{x\}$. We denote this neighbour of x by $\hat{x}$.

For any $x \in |\dot{\mathcal{F}}|$, let $n(x)$ be the unit normal pointing outwards from the (unique) cell containing x, and let $\sigma(x) = \sigma(F)$ where $F \in \dot{\mathcal{F}}$ is the unique face containing x.

With this, we define the functions

$$\begin{aligned}\delta, \mu &: |\dot{\mathcal{F}}| \to \mathbb{R},\\ \delta(x) &:= n^\top(x) A(x) n(x),\\ \mu(x) &:= \begin{cases} \dfrac{\delta(x)}{\delta(x) + \delta(\hat{x})} & \text{on interior faces,} \\ 1 & \text{on exterior faces.} \end{cases}\end{aligned}$$

For any $\omega \subset \mathcal{F}$ and $v \in H(\omega)$, we further set

$$\begin{aligned}[\![v]\!] &: |P^{-1}(\omega)| \to \mathbb{R},\\ [\![v]\!](x) &:= \begin{cases} \sigma(x)\big(v(x) - v(\hat{x})\big) & \text{on interior faces,} \\ v(x) & \text{on exterior faces.} \end{cases}\end{aligned}$$

Finally, for any face $F \in \dot{\mathcal{F}}$, let h_F denote the diameter of F, and let $n_\partial(F) := n_\partial(\tau(F))$ be the number of faces of $\tau(F)$.

With this in place, we now define for any $F \in \dot{\mathcal{F}}$ the bilinear forms a_F, m_F, s_F, and a_F^h on $H(P(F))$ as follows:

$$\begin{aligned} a_F(v,w) &:= \frac{1}{n_\partial(F)} \int_{\tau(F)} (A\nabla v) \cdot \nabla w \, dx \\ &= \frac{1}{n_\partial(F)} (\sqrt{A}\nabla v,\ \sqrt{A}\nabla w)_{L^2(\tau(F))}, \end{aligned}$$

$$\begin{aligned} m_F(v,w) &:= \int_{|F|} \mu n^\top A \nabla v [\![w]\!] \, ds \\ &= (n^\top \sqrt{A}\nabla v,\ \mu\sqrt{A} n [\![w]\!])_{L^2(F)}, \end{aligned}$$

$$\begin{aligned} s_F(v,w) &:= \int_{|F|} \frac{\mu^2 \delta}{h_F} [\![v]\!]\, [\![w]\!] \, ds \\ &= \frac{1}{h_F} (\mu\sqrt{A} n [\![v]\!],\ \mu\sqrt{A} n [\![w]\!])_{L^2(F)}, \end{aligned}$$

and, with a parameter $\alpha > 0$ whose role will be discussed shortly,

$$a_F^h(v,w) := a_F(v,w) - m_F(v,w) - m_F(w,v) + \alpha s_F(v,w).$$

Note that we have attached a fraction of a volume term to each face here. This is to ensure that all individual a_F^h are positive semi-definite, and is visualized in Figure 3.1.

The following approach to showing non-negativity of the a_F^h is standard for DG methods. Here we merely decompose the analysis into smaller pieces than usual. First, consider the following bound on m_F.

Proposition 3.46. *With the notation from above, and for any $F \in \dot{\mathcal{F}}$ it holds for all $v \in H(P(F))$ that*

$$\sup_{w \in H(P(F))} \frac{m_F(v,w)^2}{s_F(w,w)} \leq h_F \|\sqrt{A}\nabla v\|^2_{L^2(F)}.$$

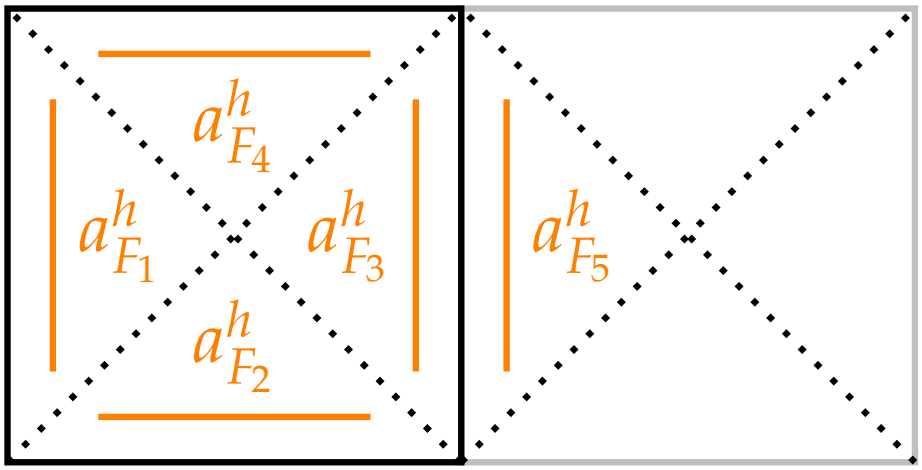


FIGURE 3.1: Given a quadrilateral mesh with two cells, consider $\omega \subset \mathcal{F}$ given in black. Then $P^{-1}(\omega) \subset \dot{\mathcal{F}}$ are the faces indicated in orange. Each of them has its own bilinear form a_F^h that is positive semi-definite. The dotted diagonals hold no explicit meaning, but the resulting triangles can be used to visualize the term a_F, which is a quarter of the volume integral on the respective cell.

Proof. Let $v, w \in H(P(F))$. Then, using the Cauchy-Schwarz inequality, we have

$$\begin{aligned} |m_F(v,w)| &= |(\sqrt{A}\nabla v,\ \mu\sqrt{A}n[\![w]\!])_{L^2(F)}| \\ &\leq \left(h_F^{1/2}\|\sqrt{A}\nabla v\|_{L^2(F)}\right) \underbrace{\left(h_F^{-1/2}\|\mu\sqrt{A}n[\![w]\!]\|_{L^2(F)}\right)}_{=(s_F(w,w))^{1/2}}. \end{aligned}$$

If $s_F(w,w) \neq 0$, this yields the result. If $s_F(w,w) = 0$, we also have $m_F(v,w) = 0$ and so the result can be continuously extended to this case. □

Next, we need to bring the volume integral a_F into the equation. The discrete trace inequality is the tool of choice for this.

Lemma 3.47 (Discrete trace inequality)**.** *Let $\mathcal{T}_h$ be a shape regular family of meshes. Then there is a constant $c > 0$ such that for all h, for all $\tau \in \mathcal{T}_h$, and for all $v \in \mathbb{P}_k(|\tau|)$, we have*

$$\|\sqrt{A}\nabla v\|_{L^2(F)} \leq c h_F^{-1/2}\|\sqrt{A}\nabla v\|_{L^2(\tau)},$$

where $F \in \dot{\mathcal{F}}(\tau)$ is any face of τ.

Proof. This is a special case of the discrete trace inequality, see [27, Lemma 12.8]. □

Together, the two previous results allow us to bound a_F^h from below in terms of $a_F + \alpha s_F$, provided α is chosen large enough. This is made precise in the following proposition.

Proposition 3.48. *With the notation from above, and in particular with c the constant from Lemma 3.47, let $F \in \dot{\mathcal{F}}$ and assume*

$$\alpha > c^2 n_\partial(F).$$

Then for all $v \in H(P(F))$ it holds that

$$a_F^h(v,v) \geq \frac{\alpha - n_\partial(F)c^2}{1+\alpha}\left(a_F(v,v) + s_F(v,v)\right).$$

Proof. As F is fixed, we write $n_\partial = n_\partial(F)$ for short. First recall that we have

$$a_F^h(v,v) = a_F(v,v) - 2m_F(v,v) + \alpha s_F(v,v).$$

With Proposition 3.46 and using the discrete trace inequality, we can bound the middle terms via

$$\begin{aligned} |m_F(v,v)|^2 &\leq \sup_{w \in H(P(F))} \frac{m_F(v,w)^2}{s_F(w,w)} s_F(v,v) \\ &\leq s_F(v,v) h_F \|\sqrt{A}\nabla v\|^2_{L^2(F)} \\ &= s_F(v,v)\; n_\partial c^2\; a_F(v,v). \end{aligned}$$

So we have

$$a_F^h(v,v) \geq a_F(v,v) - 2c\left(n_\partial a_F(v,v) s_F(v,v)\right)^{1/2} + \alpha s_F(v,v)$$

and the result follows from the fact that for any $x, y \in \mathbb{R}$ and positive $p, q > 0$, it holds that

$$\frac{p-q^2}{1+p}(x^2+y^2) \leq x^2 - 2qxy + py^2,$$

see [28, Exercise 37.2]. □

The constant α has to be chosen larger than $n_\partial(F)c^2$. This is the case for DG in general, and not particular to our presheaf setting. We thus leave more explicit bounds on c to the literature, see for instance [27, §12.2].

With the individual a_F^h bounded from below, we can safely use sums of them as needed, without worrying about positive semi-definiteness. For the following, note that a_F, m_F, s_F, and a_F^h give well-defined bilinear forms on any $H(\omega)$ as long as $P(F) \in \omega$.

Definition 3.49. With the notation from above let

$$a_\omega^h := \sum_{F \in P^{-1}(\omega)} a_F^h.$$

We call the resulting bilinear form a^h on the DG presheaf H the WSIP discretization of a. Further, we denote by $(a+s)$ the bilinear form

$$(a+s)_\omega := \sum_{F \in P^{-1}(\omega)} (a_F + s_F).$$

Remark 3.50. In this formulation, we conceptually view a^h as being assembled over faces, where each interior face appears twice and carries a fraction of the associated volume integral. As a result, contributions from volume integrals may appear up to n_∂ times, depending on how many faces a given cell has. This interpretation is primarily for theoretical convenience. In practical implementations, each cell and face is typically visited only once during matrix assembly, and redundant computations are avoided.

Corollary 3.51. *The WSIP discretization a^h is positive semi-definite if $\alpha > n_\partial c^2$, where n_∂ is the maximum number of faces of any cell in the given mesh, and c is the constant from the discrete trace inequality. Even more, a^h is coercive with respect to $(a+s)$, i.e. for all $v \in H(\omega)$ we have*

$$a^h_\omega(v,v) \geq \frac{\alpha - n_\partial c^2}{1+\alpha}(a+s)_\omega(v,v).$$

Proof. This follows by bounding the individual contributions a^h_F in the sum a^h_ω with Proposition 3.48. □

What remains is the construction of extendables H_0. For the extensions ι we intend to use extensions with zero, and in principle this can be applied to all functions in the discontinuous setting. However, a^h_ω is not coercive on $H(\omega)$ unless ω touches a global boundary. So $H_0(\omega) \subset H(\omega)$ has to be a proper subspace. To this end, for any $\omega \subset \mathcal{F}$, define the set of outer cells as

$$\partial\omega := \{\tau \in \mathcal{T}(\omega) \mid \mathcal{F}(\tau) \cap \omega^C \neq \emptyset\},$$

i.e. as all mesh cells that have at least one face in ω, and at least one face not in ω. Note that this is a slight abuse of notation. The set $\partial\omega$ is not the topological boundary of ω, which is always empty in the discrete topology.

Definition 3.52. With the notation from above, and in particular with H the DG presheaf, let for any $\omega \subset \mathcal{F}$ the space $H_0(\omega)$ be defined by

$$H_0(\omega) := \{v \in H(\omega) \mid v|_{|\partial\omega|} = 0\}.$$

We call (H, H_0, a^h) the DG presheaf (with extendables).

The following observation shows that extending with zero is compatible with the DG presheaf. This makes it indeed a presheaf with extendables — even a suitable presheaf — as the subsequent proposition lays out.

Lemma 3.53. *Let (H, H_0, a^h) be the DG presheaf and $\omega \subset \omega^* \subset \mathcal{F}$. Then for any $v \in H(\omega^*)$ and $w \in H_0(\omega)$ we have*

$$a^h_{\omega^*}(v, \iota w) = a^h_\omega(Rv, w),$$

where ιw is the extension of w with zero and $R = R_{\omega^*,\omega}$ the appropriate restriction. Further, for such v, w, the volume integrals $a_F(v,w)$ are zero whenever $\tau(F)$ is an outer cell.

Proof. Simply note that a^h_ω already includes all contributions to $a^h_{\omega^*}$ where $w \in H_0(\omega)$ can be non-zero. Moreover, the volume integrals on the outer cells vanish, as there we have $w = 0$. □

Remark 3.54. The second statement in Lemma 3.53 clarifies that volume integrals on outer cells can be disregarded in local solves, i.e. in the computation of the a^h-orthogonal projections $\pi_{\omega^*,\omega}$.

Proposition 3.55. *The DG presheaf with extendables (H, H_0, a^h) is a suitable presheaf, where extensions are given by extending DG functions with zero.*

Proof. H is a local presheaf, as previously noted, since the restrictions are well-defined and if two functions agree locally on ω_i, they also agree on $\cup_i \omega_i$. To see that a^h is

local, note that for any $\omega \subset \omega^* \subset \mathcal{F}$ we have $(\omega^{*\mathring{}}\backslash\omega) = \omega^*\backslash\omega$ in the discrete topolgy. Thus

$$a^h_{\omega^*} = a^h_{(\omega^*\backslash\omega)} + a^h_\omega,$$

since a^h is by definition a sum over individual face contributions.

As a^h is bounded from below with $(a+s)$, we can see that a^h is coercive on $H_0(\omega)$. Indeed, note in $(a+s)$ all individual face and volume integrals are positive semi-definite. Thus, any strictly positive contribution can never be made up for with a negative contribution elsewhere. In particular, any functions in $\ker(a+s)$ can have no jumps on the mesh faces and must be constant within mesh cells — so they are constant everywhere. If $\mathcal{T}(\omega)$ touches the global Dirichlet boundary, the constant has to be 0. If not, then $\mathcal{T}(\omega)$ contains outer cells of ω, likewise fixing the constant.

By Lemma 3.53, extending with zero provides well-defined extensions. To show suitability, note that by definition we have $H(\Omega) = H_0(\Omega)$ and the restrictions are surjective. Finally, all involved vector spaces are finite dimensional and thus separable Banach spaces. □

To summarize, we have shown that the WSIP method naturally fits into the abstract framework developed in this chapter. We constructed the DG presheaf over the discrete topology on the set $\mathcal{F}$ of mesh faces to ensure both locality and coercivity of the associated bilinear form. This setup enables the application of the general theory with only a minor deviation from standard practice: the inclusion of fractional volume integrals over outer cells of a given subdomain. These terms are necessary to ensure locality of a^h. As discussed in Remark 3.54, they do not affect the evaluation of orthogonal projections.

Overall, the primary technical challenge lay in expressing the bilinear form a^h in a manner compatible with the sheaf structure and in verifying its coercivity. The latter is not particular to the presheaf setting, but a standard technicality in Discontinuous Galerkin methods.

3.3 Generalized Elements

In this section, we introduce the concept of generalized elements, which form the foundational building blocks for the multilevel spectral elements developed later. Generalized elements extend the idea of Generalized Finite Elements (GFEM, [59]) to the abstract setting of suitable presheaves. Defined over any suitable presheaf, generalized elements employ local approximation spaces that are assembled into a global coarse space using a partition of unity. This resulting global space can itself be interpreted as a suitable presheaf by means of the canonical subsheaf construction (Proposition 3.13).

In this chapter, we focus on the hierarchical structure of generalized elements, and defer the detailed construction of spectral local approximation spaces to the next chapter. Specifically, we describe how a canonical partition of unity can be defined for the presheaf of generalized elements, assuming a mild unambiguity condition. Under this assumption, a natural multilevel hierarchy emerges, forming the basis for the preconditioners considered later.

3.3.1 Two-Level Construction

First, we consider the two-level construction of generalized elements, noting that they automatically inherit the structural properties required to form a suitable presheaf in their own right.

Definition 3.56 (Generalized elements)**.** Let (H, H_0, a) be a suitable presheaf over Ω, and let $\{\omega_i\}_{i\in I}$ be a finite open cover with a subordinate partition of unity $\{\chi_i\}_{i\in I}$. For each $i \in I$, choose finite-dimensional subspaces

$$S_i \subset H(\omega_i),$$

and define the global approximation space by

$$V := \sum_{i\in I} \left(\iota^H_{\omega_i,\Omega}\right) \chi_i S_i.$$

Let $\mathcal{H}$ be the canonical subsheaf associated with the inclusion $V \subset H(\Omega)$, as described in Proposition 3.13. Define the spaces

$$\mathcal{H}_0(\omega) := \sum_{i\in J_\omega} \left(\iota^L_{\omega_i,\omega}\right) \chi_i S_i,$$

where $J_\omega := \{i \in I \mid \omega_i \subset \omega\}$. We call $(\mathcal{H}, \mathcal{H}_0, a)$ the presheaf of *generalized elements* for (H, H_0, a) with respect to the cover $\{\omega_i\}_{i\in I}$, the partition of unity $\{\chi_i\}_{i\in I}$, and the local approximation spaces $\{S_i\}_{i\in I}$.

Remark 3.57. In the above definition, the assumption of finite dimensionality on the local approximation spaces S_i assures that the canonical subsheaf construction is a presheaf of separable Banach spaces. It can be replaced by assuming that all resulting $\mathcal{H}(\omega) \subset H(\omega)$ are closed subspaces. This may be tedious to verify, however, and in practice the usefulness of infinite-dimensional approximation spaces is questionable.

Proposition 3.58. *With the notation from Definition 3.56, in particular with $(\mathcal{H}, \mathcal{H}_0, a)$ the presheaf of generalized elements for (H, H_0, a), the inclusion*

$$(\mathcal{H}, \mathcal{H}_0, a) \subset (H, H_0, a)$$

is an embedding of suitable presheaves.

Proof. By Proposition 3.13, $\mathcal{H}$ is a local presheaf, and the bilinear form $a(\cdot,\cdot)$ is local, symmetric, and positive semi-definite on $\mathcal{H}$.

To verify that $\mathcal{H}_0$ gives well-defined extendables, we check the two conditions in Definition 3.19. Let $\omega \subset \omega^*$ be open. For condition (i), observe that $\mathcal{H}_0(\omega) \subset L_0(\omega)$, so $a(\cdot,\cdot)$ is coercive on $\mathcal{H}_0(\omega)$. For condition (ii), we define the extensions $\iota^{\mathcal{H}}$ by restricting the extensions from H, i.e. we set $\iota^{\mathcal{H}} := \iota^H|_{H_0}$. Then,

$$\iota^H_{\omega,\omega^*}(\mathcal{H}_0(\omega)) = \iota^H_{\omega,\omega^*}\left(\sum_{i\in J_\omega}\left(\iota^H_{\omega_i,\omega}\right)\chi_i S_i\right) = \sum_{i\in J_\omega}\left(\iota^H_{\omega_i,\omega^*}\right)\chi_i S_i \subset \mathcal{H}_0(\omega^*),$$

as required.

By construction, the embedding $\mathcal{H} \hookrightarrow H$ is compatible with restrictions and extensions, and therefore

$$(\mathcal{H}, \mathcal{H}_0, a) \subset (H, H_0, a)$$

is an embedding of presheaves with extendables. To show that $(\mathcal{H}, \mathcal{H}_0, a)$ is a suitable presheaf, first note that $\mathcal{H}(\Omega) = \mathcal{H}_0(\Omega)$ by definition. Further, the restriction maps in $\mathcal{H}$ are surjective. Indeed, for the canonical subsheaf we have $R^{\mathcal{H}}_{\Omega,\omega} = R^L_{\Omega,\omega}$, and by construction $\mathcal{H}(\omega) = R^L_{\Omega,\omega}(H(\Omega))$. Finally, each space $\mathcal{H}(\omega)$ is finite-dimensional, and hence a separable Banach space. □

3.3.2 Multilevel Domain Decomposition

As shown above, if (H, H_0, a) is a suitable presheaf, then the corresponding presheaf of generalized elements is also suitable. This allows us to recursively construct generalized elements over generalized elements, leading to a multilevel hierarchy of suitable presheaves. In this subsection, we develop preliminary terminology for this procedure.

Definition 3.59 (Hierarchy of suitable presheaves)**.** Let Ω be a topological space and L a natural number. By a hierarchy of suitable presheaves of level L over Ω, we refer to a chain of embeddings of suitable presheaves

$$(\mathcal{H}_0, \mathcal{H}_{0,0}, a) \subset (\mathcal{H}_1, \mathcal{H}_{1,0}, a) \subset \cdots \subset (\mathcal{H}_L, \mathcal{H}_{L,0}, a),$$

over Ω.

Definition 3.60 (Multilevel domain decomposition)**.** Let Ω be a topological space and L a natural number. By a *(multilevel) domain decomposition* of Ω, we refer to a collection of open subsets $\{\omega_{l,i}\}_{(l,i)\in I}$ that is constructed as follows. On level L we start with a natural number $M_L \in \mathbb{N}$ and an open cover

$$\Omega = \bigcup_{i=1}^{M_L} \omega_{L,i}.$$

Now let $1 \leq l < L$, assume we have chosen $M_{l+1} \in \mathbb{N}$, and defined an open cover

$$\Omega = \bigcup_{i=1}^{M_{l+1}} \omega_{l+1,i}.$$

Then choose a cover of the index set $I_{l+1} := \{1, \dots, M_{l+1}\}$. More precisely, let $M_l < M_{l+1}$, and let $\{I_{l,i}\}_{i \le M_l}$ be subsets of I_{l+1} such that

$$I_{l+1} = \bigcup_{i=1}^{M_l} I_{l,i}.$$

The open cover $\{\omega_{l,i}\}_{i=1}^{M_l}$ is then defined by setting

$$\omega_{l,i} := \bigcup_{j \in I_{l,i}} \omega_{l+1,j}.$$

Finally, on level 0 we always choose $M_0 = 1$, which results in $\omega_{0,1} = \Omega$. We call the $\{I_l\}_{l \le L}$ the index hierarchy, and $\{I_{l,i}\}_{i=1}^{M_l}$ the index cover of level l. We often work with the complete index set

$$I := \{(l,i) \in \mathbb{N}_0 \times \mathbb{N} \mid l \le L, i \le M_l\},$$

and use the individual I_l only when needed for clarity in multilevel constructions.

Definition 3.61 (Multilevel index partition)**.** Let $\{\omega_{l,i}\}_{(l,i)\in I}$ be a multilevel domain decomposition of a topological space Ω with index cover $\{I_{l,i}\}_{i=1}^{M_l}$. An *index partition* (or *discrete partition of unity*) for this domain decomposition is a collection of maps $\{\phi_{l,i}\}_{(l,i)\in I}$, such that for each level $l < L$ the family $\{\phi_{l,i}\}_{i=1}^{M_l}$ forms a partition of unity of functions subordinate to $\{I_{l,i}\}_{i=1}^{M_l}$. More precisely, for each $l < L$ and $i \le M_l$ we have

$$\phi_{l,i} : I_{l+1} \to [0,1], \quad \operatorname{supp}(\phi_{l,i}) \subset I_{l,i}, \quad \text{and} \quad \sum_{i=1}^{M_l} \phi_{l,i} \equiv 1.$$

For notational convenience, we set $I_{L+1} = \Omega$ and $\phi_{L,i} = 0$. The $\phi_{L,i}$ will never be used.

Definition 3.62 (Multilevel colouring)**.** Let $\{\omega_{l,i}\}_{l,i}$ be a multilevel domain decomposition, such that for each l the cover $\{\omega_{l,i}\}_{i=1}^{M_l}$ allows ξ-colouring. Then we say that the multilevel domain decomposition allows ξ-colouring. Similarly, let $\{\chi_{l,i}\}_{i=1}^{M_l}$ be a partition of unity subordinate to $\{\omega_{l,i}\}_{i=1}^{M_l}$ for each individual level $l \le L$. If on each level these allow ξ-colouring, we say that $\{\chi_{l,i}\}_{(l,i)\in I}$ allows ξ-colouring.

In practice, the number of levels and subdomains in domain decomposition methods is always finite, so the setups we consider necessarily allow for ξ-colouring with, at worst, $\xi = \max_l M_l$. However, since ξ appears as a constant in many theoretical results, obtaining a sharper bound is both desirable and often possible. For instance, in the case of a structured decomposition into rectangular subdomains $\omega_{l,i}$, one typically finds $\xi = 4$ in two dimensions or $\xi = 8$ in three dimensions — provided that subdomains do not overlap with the neighbours of their neighbours. Importantly, ξ is usually independent of the number of subdomains M_l, which is a key property when considering scalability.

3.3.3 Canonical Partition of Unity

Given a suitable presheaf (H, H_0, a) and a multilevel domain decomposition, we aim to construct a hierarchy of increasingly coarser suitable presheaves by repeatedly forming generalized elements. While the conceptual idea is straightforward, several

technical details require careful treatment. In particular, in practice we must construct a partition of unity appropriate for each coarser level. This subsection provides a canonical way to do so, based on a multilevel index partition.

Notation 3.63. Throughout this subsection, let (H, H_0, a) be a suitable presheaf over Ω, associated with a finite open cover $\{\omega_i\}_{i\in I}$, a subordinate partition of unity $\{\chi_i\}_{i\in I}$, and local approximation spaces S_i.

An important practical concern when working with generalized elements $(\mathcal{H}, \mathcal{H}_0, a)$ for (H, H_0, a) is how to construct suitable bases. In a two-level method, it suffices to provide a basis for $\mathcal{H}(\Omega)$ since the local spaces $\mathcal{H}(\omega)$ are never used. This is often straightforward in applications. The standard approach is to take bases X_i for the local approximation spaces S_i, and hope that the global union $X := \bigcup_i \chi_i X_i$ is linearly independent over Ω.

This assumption implicitly requires two conditions. First, the partition functions χ_i must act injectively on X_i. Second, the global collection X must not lose linear independence when assembled. This is reasonable to expect in practice if each S_i has relatively low dimension compared to its ambient space $H(\omega_i)$.

However, when constructing the next level of generalized elements, it becomes necessary to consider local spaces $\mathcal{H}(\omega)$ for subsets $\omega \subset \Omega$. A natural idea is to restrict the global basis X to ω, but this fails for several reasons. Most obviously, any basis X_i whose support does not intersect ω will vanish on ω and must be discarded. While this can be handled easily, the remaining functions still do not necessarily form a basis for $\mathcal{H}(\omega)$.

To illustrate this, consider the case where (H, H_0, a) is the sheaf of continuous piecewise linear finite elements over a domain $\Omega \subset \mathbb{R}^2$, and $\{\omega_i\}_{i\in I}$ is a structured cover of overlapping square subdomains. Suppose that neighbouring patches overlap by two layers of mesh cells. Then a patch ω_i will share exactly four cells — corresponding to nine degrees of freedom — with a diagonal neighbour ω_j. If X_j contains more than nine basis vectors, the restriction of $\chi_j X_j$ to ω_i will necessarily become linearly dependent. Yet we cannot simply discard these functions, since some of them may still be required to span $\mathcal{H}(\omega_i)$. We discuss in Section 5.2.2 how this issue can be addressed in implementations.

Next, we formulate a slightly stronger assumption than X being a basis for $\mathcal{H}(\Omega)$, which is still weak enough to be met in typical applications.

Definition 3.64 (Unambiguity, components)**.** Let $(\mathcal{H}, \mathcal{H}_0, a)$ be the presheaf of generalized elements for (H, H_0, a) with respect to the cover $\{\omega_i\}_{i\in I}$. We call $(\mathcal{H}, \mathcal{H}_0, a)$ *unambiguous* if for any open ω and with

$$J := \{i \in I \mid \omega_i \subset \omega\}$$

the following holds. Whenever we have

$$v_i, w_i \in \left(\iota^{\mathcal{H}}_{\omega_i,\omega}\right) \chi_i S_i, \quad i \in J$$

with

$$\sum_{i\in I} v_i|_\omega = \sum_{i\in I} w_i|_\omega =: v,$$

then $v_i = w_i$ for all $i \in J$. In this case we call v_i the i-th *component* of v over ω.

Unambiguity now enables the definition of a partition of unity on generalized elements by projecting a global section $v \in \mathcal{H}(\Omega)$ onto its local components. In the

construction, we use index notation of the form $\omega_{*,i}$. Here, the asterisk is purely symbolic and serves as a placeholder that will later be replaced by the level index in the multilevel setting.

Definition 3.65 (Canonical partition of unity)**.** Let $(\mathcal{H}, \mathcal{H}_0, a)$ be the presheaf of generalized elements for (H, H_0, a). Assume that $(\mathcal{H}, \mathcal{H}_0, a)$ is unambiguous, let $M \in \mathbb{N}$, and let $\{I_i\}_{i=1}^M$ be a cover for the index set, i.e. we have

$$I = \bigcup_{i=1}^{M} I_i.$$

Let $\{\phi_i\}_{i=1}^M$ be a partition of unity of functions on I subordinate to this cover. More precisely, let $\phi_i : I \to \mathbb{R}$ with

$$\operatorname{supp}(\phi_i) \subset I_i, \quad \text{and} \quad \sum_{i=1}^{M} \phi_i \equiv 1.$$

For any $i \leq M$, set $\omega_{*,i} := \bigcup_{j \in I_i} \omega_j$ and consider the maps

$$\begin{aligned} \chi_{*,i} : \mathcal{H}(\omega_{*,i}) &\to \mathcal{H}_0(\omega_{*,i}) \\ w &\mapsto \sum_{j \in I_i} \phi_i(j) w_j, \end{aligned}$$

where w_j is the j-th component of w. We call these $\{\chi_{*,i}\}_{i=1}^M$ the *canonical partition of unity* for the generalized elements $(\mathcal{H}, \mathcal{H}_0, a)$ with respect to the index partition $\{\phi_i\}_{i=1}^M$.

Proposition 3.66. *The canonical partition of unity for generalized finite elements is well-defined and indeed gives a partition of unity.*

Proof. With the notation from Definition 3.65, note that by unambiguity, the j-th component over $\omega_{*,i}$ is well-defined for all $j \in I_i$. To see that the $\chi_{*,j}$ sum up to the identity, let $v \in H(\Omega)$. By unambiguity we can write

$$v = \sum_{j \in I} \iota_{\omega_j,\Omega}\, v_j$$

with the $v_j \in \chi_j S_i$ the unique components of v. With this we have

$$\chi_{*,i}(v) = \chi_{*,i} R_{\Omega,\omega_{*,i}} \sum_{j \in I} \left(\iota_{\omega_j,\Omega}\right) v_j = \chi_{*,i} \sum_{j \in I} \left(\iota_{\omega_j,\omega_{*,i}}\right) v_j = \sum_{j \in I} \phi_i(j) \ \left(\iota_{\omega_j,\omega_{*,i}}\right) v_j.$$

Whenever $\phi_i(j) \neq 0$, it holds $\omega_j \subset \omega_{*,i}$, and thus $\iota_{\omega_j,\omega_{*,i}}$ is well-defined where required. Since the ϕ_i sum up to 1, it follows that

$$\sum_{i=1}^{M} \left(\iota_{\omega_{*,i},\Omega}\right) \chi_{*,i}(v) = \sum_{i=1}^{M} \sum_{j \in I} \phi_i(j) \left(\iota_{\omega_{*,i},\Omega}\right) \left(\iota_{\omega_j,\omega_{*,i}}\right) v_j = v.$$

□

Thus, we have constructed a partition of unity for the suitable presheaf of generalized elements using only an index partition as input. Since such a partition is always available in practical settings, there is no obstruction to defining yet another level of generalized elements on top. We make this precise below.

3.3.4 Multilevel Generalized Elements

After the necessary preliminaries, we are now in a position to define multilevel generalized elements — a hierarchy of generalized elements constructed at successively coarser levels.

Definition 3.67 (Multilevel generalized elements)**.** Let (H, H_0, a) be a suitable presheaf and let $\{\omega_{l,i}\}_{(l,i)\in I}$ be a multilevel domain decomposition of level L with an index partition $\{\phi_{l,i}\}_{(l,i)\in I}$. Further, let $\{\chi_{L,i}\}_{i\leq M_L}$ be a partition of unity subordinate to the cover $\{\omega_{L,i}\}_{i\leq M_L}$. By *multilevel generalized elements* we refer to a hierarchy of suitable presheaves $\{(\mathcal{H}_l, \mathcal{H}_{l,0}, a)_l\}_{l\leq L}$, which are constructed by iteratively taking generalized elements of the previous level, starting with

$$(\mathcal{H}_L, \mathcal{H}_{L,0}, a) := (H, H_0, a).$$

For $0 \leq l < L$, assume that $(\mathcal{H}_l, \mathcal{H}_{l,0}, a)$ and a partition of unity $\{\chi_{l,i}\}_{i\leq M_l}$ subordinate to the cover $\{\omega_{l,i}\}_{i\leq M_l}$ are already defined. Then we choose local approximation spaces

$$S_{l,i} \subset \mathcal{H}_l(\omega_{l,i})$$

for all $i \leq M_l$, and set $(\mathcal{H}_{l-1}, \mathcal{H}_{l-1,0}, a)$ to be the presheaf of generalized elements for $(\mathcal{H}_l, \mathcal{H}_{l,0}, a)$ with respect to the cover, the partition of unity and the local approximation spaces of the given level. As the next partition of unity $\{\chi_{l-1,i}\}_{i\leq M_{l-1}}$, we choose the canonical partition of unity of $(\mathcal{H}_{l-1}, \mathcal{H}_{l-1,0}, a)$ with respect to the index partition $\{\phi_{l-1,i}\}_{i\leq M_{l-1}}$, where we assume that $(\mathcal{H}_{l-1}, \mathcal{H}_{l-1,0}, a)$ is unambiguous.

This procedure is repeated until we reach level $l = 0$ with only a single subdomain $\omega_{0,1} = \Omega$, where we set $S_{0,1} = 0$ for notational convenience. The hierarchy $\{(\mathcal{H}_l, \mathcal{H}_{l,0}, a)_l\}_{l\leq L}$ is called the hierarchy of *multilevel generalized elements* for (H, H_0, a) with respect to the domain decomposition $\{\omega_{l,i}\}_{(l,i)\in I}$, the initial partition of unity $\{\chi_{L,i}\}_{i\leq M_L}$, the index partitions $\{\phi_{l,i}\}_{(l,i)\in I}$ and the local approximation spaces $\{S_{l,i}\}_{(l,i)\in I}$.

Remark 3.68. The above definition assumes unambiguity of the involved generalized elements. This is a requirement for the canonical partition of unity to be well-defined. It can be relaxed by providing an alternative construction of partitions of unity $\{\chi_{l,i}\}_{l\leq M_l}$ for all levels $l < L$.

Up to this point, the construction remains a recipe — albeit a powerful one. Once a given PDE setting can be shown to give rise to a suitable presheaf, we immediately obtain a corresponding hierarchy of multilevel generalized elements. The practical effectiveness of this framework, however, depends crucially on the choice of local approximation spaces $S_{l,i}$. In this thesis, we focus on the spectral approach described in Chapter 2, which is generalized to the setting of multilevel generalized elements in Chapter 4.

Chapter 4

Spectral Elements

The previous chapter introduced generalized elements, which rely on local approximation spaces whose construction has not yet been addressed. In this chapter, we fill this gap by providing a general recipe for incorporating spectral methods into our abstract framework. Within the setting of generalized elements, we will consider symmetric positive semidefinite bilinear forms a_i, b_i over each subdomain ω_i. We then define local spectral problems of the form

$$b_i(v, \cdot) = \lambda a_i(v, \cdot) \tag{4.1}$$

which determine the functions to be included in the local approximation spaces. For the theory we will mostly use the equivalent formulation

$$b_i(v, \cdot) = \mu(a_i + b_i)(v, \cdot) \tag{4.2}$$

with $\lambda = \frac{\mu}{1-\mu}$. This formulation is advantageous because $a_i + b_i$ is often coercive, which simplifies much of the analysis. Moreover, it clarifies how to handle the case of $\lambda = \infty$.

Notation 4.1. In the following, when denoting eigenvalues of local eigenproblems in the context of spectral methods by λ we are referring to the formulation (4.1). Likewise, we are referring to (4.2) when denoting them by μ.

To allow for greater flexibility, we also accommodate the singular case where $\ker(a_i + b_i) \neq 0$. Since in this case the eigenproblem is not well-defined on its entire domain, it must be considered on the quotient space by $\ker(a_i + b_i)$ instead. This introduces some additional theoretical complexity but renders the method more flexible. For instance, it enables the definition and analysis of GenEO variants for a DG discretization in Section 4.2.5.

To improve readability, we first present the concept of abstract spectral elements for the non-singular case in Section 4.1.1. Subsequently, we develop the theoretical tools required for handling the general case considered in Section 4.1.3. The section concludes with a generalization of the local approximation property known from spectral methods to the abstract framework.

Section 4.2 gives abstract formulations of Full GenEO, GenEO, and MS-GFEM, along with the corresponding local approximation results. We also briefly review theoretical insights behind MS-GFEM, particularly the near-exponential decay of its eigenvalues. These methods are then applied to both continuous finite elements and the previously introduced discontinuous WSIP discretization. For the latter, we provide two GenEO variants, one of which is used in the numerical experiments of [10], which we discuss later in this thesis.

Finally, we introduce multilevel spectral elements in Section 4.3. This is straightforward, as the multilevel structure of generalized elements from Section 3.3.4 already provides a suitable framework. It remains only to discuss the specific assumptions for spectral elements, which are typically simply passed on from finer to coarser levels.

We will analyse the use of these abstract multilevel spectral elements in preconditioners for a non-restricted additive Schwarz scheme in Chapter 5 and for a restricted hybrid Schwarz scheme in Chapter 6.

4.1 Spectral Elements

In this section, we develop spectral elements as a special case of the generalized elements from Section 3.3. The idea is to obtain the approximation spaces S_i by solving local eigenproblems, following the methods described in Chapter 2.

First, we consider the simpler case of non-singular eigenproblems to clarify the general idea. We then develop the theory needed to handle the singular case and to state the approximation result Theorem 4.15, which we will use in subsequent chapters.

4.1.1 Spectral Elements, Non-Singular Case

We begin by describing spectral elements in the simplified setting where the local eigenproblems are non-singular. This case illustrates the core ideas cleanly, without the technical complications of the general case.

Definition 4.2 (Spectral elements, non-singular case)**.** Let (H, H_0, a) be a suitable presheaf over Ω, and let $(\mathcal{H}, \mathcal{H}_0, a)$ be the presheaf of generalized elements for (H, H_0, a), with respect to a cover $\{\omega_i\}_{i\in I}$, a partition of unity $\{\chi_i\}_{i\in I}$, and local approximation spaces $\{S_i\}_{i\in I}$.

For each $i \in I$, let $\widetilde{V}_i \subset H(\omega_i)$ be a closed subspace, and let a_i, b_i be symmetric, positive semi-definite bilinear forms on $\widetilde{V}_i$, such that $a_i + b_i$ is coercive. Consider the local eigenproblems of finding $(\mu, v) \in \mathbb{R} \times \widetilde{V}_i$ such that

$$b_i(v, w) = \mu(a_i + b_i)(v, w) \quad \text{for all } w \in \widetilde{V}_i. \tag{4.3}$$

Suppose that each S_i includes all eigenspaces corresponding to eigenvalues $\mu > \mu_{\max}$, for some fixed $0 < \mu_{\max} < 1$, that is:

$$S_i \supset \ker(b_i - \mu(a_i + b_i)) \quad \text{for all } \mu > \mu_{\max}.$$

Then $(\mathcal{H}, \mathcal{H}_0, a)$ is called the presheaf of *spectral elements* for (H, H_0, a), with respect to the bilinear forms $\{a_i\}_{i\in I}$, $\{b_i\}_{i\in I}$, the given cover, the partition of unity, and with eigenvalue threshold

$$\lambda_{\max} := \frac{\mu_{\max}}{1 - \mu_{\max}}.$$

Spectral elements satisfy a local approximation property, which we investigate in the following. First, the proposition below establishes a few basic facts.

Proposition 4.3. *Let V be a separable Banach space, and let a, b be two symmetric, positive definite bilinear forms on V, such that $a + b$ is coercive. Consider the eigenproblem of finding $(\mu, v) \in \mathbb{R} \times V$ such that*

$$b(v, w) = \mu(a + b)(v, w) \quad \textit{for all } w \in V. \tag{4.4}$$

Then the following statements hold:

(i) Eigenspaces corresponding to distinct eigenvalues are pairwise orthogonal with respect to a, b, and $a + b$.

(ii) There are at most countably many eigenvalues.

(iii) (μ, v) is an eigenpair of (4.4) if and only if

$$b(v, w) = \lambda a(v, w) \quad \textit{for all } w \in V, \tag{4.5}$$

where $\lambda = \frac{\mu}{1-\mu}$, *and we write* $\lambda = \infty$ *for the case* $\mu = 1$.

(iv) The eigenvalues satisfy $0 \leq \mu \leq 1$.

Proof. For (i), let (μ_1, w_1), (μ_2, w_2) be eigenpairs. Then,

$$\mu_1(a+b)(w_1, w_2) = b(w_1, w_2) = b(w_2, w_1) = \mu_2(a+b)(w_1, w_2).$$

In particular, $(a+b)(w_1, w_2) = 0$ if $\mu_1 \neq \mu_2$. In this case we also have

$$b(w_1, w_2) = \mu_1(a+b)(w_1, w_2) = 0,$$

and consequently

$$a(w_1, w_2) = (a+b)(w_1, w_2) - b(w_1, w_2) = 0.$$

Therefore, the eigenspaces are pairwise orthogonal with respect to each of the bilinear forms a, b and $a+b$.

For (ii), the coercivity of $a+b$ implies that $(V, a+b)$ is a separable Hilbert space, since we assumed V to be separable. Thus, there can be at most countably many mutually orthogonal nontrivial eigenspaces.

Statement (iii) follows from an easy calculation: subtract $\mu b(v, w)$ from the right-hand side of (4.4), and then divide by $(1-\mu)$. The case $\mu = 1$ implies that the corresponding eigenvectors lie in $\ker(a)$, which makes $\lambda = \infty$ a natural extension of $\lambda = \frac{\mu}{1-\mu}$.

Statement (iv) follows from (iii): if $\mu < 0$ or $\mu > 1$, then $\lambda < 0$, which contradicts the positive semi-definiteness of b. □

By the preceding proposition, there are two equivalent formulations of the local eigenproblems underlying spectral elements: (4.4) and (4.5). In this thesis, we adopt (4.4) as the basis for the definition, as it is more concise. The alternative formulation, (4.5), is more commonly found in the literature. This is due to its direct relationship with the threshold parameter $\lambda_{\max}$, which enters the practically relevant approximation results. To avoid ambiguity, we adopt the following notational convention.

Notation 4.4. In the context of spectral elements, we will consistently use μ to denote eigenvalues arising from the formulation in (4.4), and λ to refer to the corresponding eigenvalues in the equivalent form (4.5).

For example, if we state that the spectrum of a local eigenproblem accumulates near $\lambda = 1$, this corresponds to an accumulation point near $\mu = 1/2$ in the formulation of Definition 4.2. Next, consider the following basic approximation result that provides the main motivation for our definition of spectral elements.

Lemma 4.5. *In the setting of Proposition 4.3, assume additionally that the span of the eigenvectors is dense in* V *and define*

$$W := \sum_{\mu > \mu_{\max}} \ker(b - \mu(a+b))$$

to be the span of the dominant eigenspaces. Let $p : V \to W$ *denote the* $a+b$*-orthogonal projection onto* W. *Then the following statements hold:*

(i) For all $v \in V$, *we have*

$$b(v - p(v), v - p(v)) \leq \lambda_{\max} a(v, v), \quad \text{with } \lambda_{\max} = \frac{\mu_{\max}}{1 - \mu_{\max}}.$$

(ii) The projection is orthogonal with respect to both a and b. More precisely, for all $v \in V$ it holds that

$$a(v - p(v), p(v)) = 0, \quad \textit{and} \quad b(v - p(v), p(v)) = 0.$$

Proof. By Proposition 4.3, and using the assumption that the eigenvectors form a dense subset, we obtain a countable $(a+b)$-orthonormal basis $\{v_n\}_n$ of eigenvectors for V, with corresponding eigenvalues $\{\mu_n\}_n$. For $n \neq m$ we have

$$b(v_n, v_m) = \mu_n (a+b)(v_n, v_m) = 0,$$

and thus this basis is also b-orthogonal. Let

$$v = \sum_n \alpha_n v_n, \quad \text{and } v - p(v) = \sum_n \widetilde{\alpha}_n v_n$$

for coefficients $\alpha_n, \widetilde{\alpha}_n \in \mathbb{R}$. Since $v - p(v)$ lies in the $(a+b)$-orthogonal complement of the dominant eigenspaces W, we must have

$$\widetilde{\alpha}_n = \begin{cases} 0 & \text{if } \mu_n > \mu_{\max}, \\ \alpha_n & \text{otherwise.} \end{cases}$$

By b- and $(a+b)$-orthogonality of the v_n it holds that

$$\begin{aligned} b(v - p(v), v - p(v)) &= \sum_n \widetilde{\alpha}_n^2 \, b(v_n, v_n) \\ &= \sum_n \widetilde{\alpha}_n^2 \, \mu_n (a+b)(v_n, v_n) \\ &\leq \mu_{\max} \sum_n \widetilde{\alpha}_n^2 (a+b)(v_n, v_n) \\ &\leq \mu_{\max} \sum_n \alpha_n^2 a(v_n, v_n) + \mu_{\max} \sum_n \widetilde{\alpha}_n^2 b(v_n, v_n) \\ &= \mu_{\max} \, a(v, v) + \mu_{\max} \, b(v - p(v), v - p(v)). \end{aligned}$$

To conclude the proof of (i), we subtract $\mu_{\max} b(v - p(v), v - p(v))$ from both sides and divide by $1 - \mu_{\max}$ to obtain

$$b(v - p(v), v - p(v)) \leq \frac{\mu_{\max}}{1 - \mu_{\max}} a(v, v).$$

For (ii), note that since the v_n are $(a+b)$- and b- orthogonal, they are also a-orthogonal. In particular

$$a(v - p(v), p(v)) = \sum_n (\alpha_n - \widetilde{\alpha}_n) \widetilde{\alpha}_n \, a(v_n, v_n) = 0,$$

and similarly $b(v - p(v), p(v)) = 0$, completing the proof. □

The effectiveness of the above result depends critically on the choice of the bilinear forms a_i and b_i. Before addressing this in Section 4.2, we first work on developing spectral elements for the case where $\ker(a_i + b_i) \neq \{0\}$.

4.1.2 Splitting Bilinear Forms

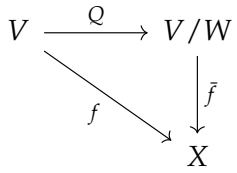
In order to extend the definition of spectral elements to settings where $\ker(a_i + b_i)$ may be nontrivial, we require a framework that allows us to work with quotient spaces. In this subsection, we thus develop a generalized approximation result for local eigenproblems modulo $\ker(a_i + b_i)$, as well as an alternative modulo $\ker(a_i)$. First, we establish some general terminology.

Definition 4.6 (Splitting subspace)**.** Let V be a topological vector space, $W \subset V$ a subspace, and let $Q : V \to V/W$ be the corresponding quotient map. We say that W (or the projection Q) is *splitting* if Q admits a right-inverse; that is, there exists a (continuous and linear) map $g : V/W \to V$ such that $Q \circ g = \mathrm{Id}$. In this case, we say that the map g *splits* W.

Remark 4.7. Let V be a topological vector space, $W \subset V$ a splitting subspace and g a map that splits W. Then $V = W \oplus g(V/W)$. Moreover, with Q the quotient map, we have $Q(g \circ Q - \mathrm{id}_V) = 0$ and thus $\mathrm{Image}(g \circ Q - \mathrm{id}_V) \subset W$.

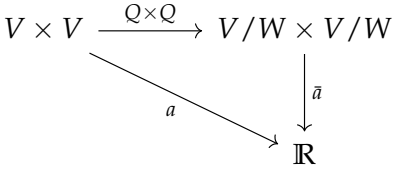
Proposition 4.8 (Quotients of maps)**.** *Let V and X be topological vector spaces, and let $Q : V \to V/W$ be a quotient map for a splitting subspace $W \subset V$. Then the following hold true.*

(i) Let $f : V \to X$ be a (continuous linear) map such that $W \subset \ker(f)$. Then there exists a unique (continuous and linear) map $\bar{f} : V/W \to X$ such that



commutes. We call $\bar{f}$ the quotient *of the map f by W and denote it by f/W, or more often simply by $\bar{f}$.*

(ii) Let $a(\cdot,\cdot)$ be a (continuous) bilinear form on V such that $W \subset \ker(a)$. Then there is exactly one (continuous) bilinear form $\bar{a}$ on V/W such that



commutes. We call this the quotient *of the bilinear form a by W and denote it by a/W, or more often simply by $\bar{a}$.*

Proof. For (i), note that uniqueness follows from the fact that Q is surjective. To show existence, let $g : V/W \to V$ split Q and define $\bar{f} := f \circ g$. Then

$$\bar{f} \circ Q = f \circ g \circ Q = f \circ (\mathrm{id}_V + (g \circ Q - \mathrm{id}_V)) = f,$$

since $\mathrm{Image}(g \circ Q - \mathrm{id}_V) \subset W \subset \ker(f)$. Statement (ii) follows by similar arguments for $\bar{a}(\cdot,\cdot) := a(g(\cdot), g(\cdot))$. □

Remark 4.9. Note that in a Hilbert space V, all closed subspaces W are splitting since we can decompose $V = W \oplus W^{\perp}$.

We will only consider applications to Hilbert spaces V in this thesis. Thus, we could use the decomposition $V = W \oplus W^\perp$ to simplify some of the definitions in this section. However, there are already three bilinear forms in play: a_{ω_i}, a_i, and b_i. We thus refrain from additionally introducing an underlying inner product. Below, we see that the approximation result Lemma 4.5 continues to hold true, even if the eigenspaces therein are considered only over the quotient $V/\ker(a+b)$.

Lemma 4.10. *Let V be a separable Banach space and let a, b be two symmetric, positive definite bilinear forms on V. Suppose that $\ker(a+b)$ splits and that $(\bar{a}+\bar{b})$ is coercive on the quotient space $\bar{V} := V/\ker(a+b)$. Assume further that the eigenvectors of*

$$\bar{b}(v,\cdot) = \mu(\bar{a}+\bar{b})(v,\cdot) \tag{4.6}$$

form a dense subset of $\bar{V}$, where $\bar{a}, \bar{b}$ are the quotients of a, b by $\ker(a+b)$. Let $\mu_{\max} \in (0,1)$ and define

$$W := \sum_{\mu > \mu_{\max}} \ker(\bar{b} - \mu(\bar{a}+\bar{b})) \subset \bar{V}.$$

Finally, let g be any map that splits $\ker(a+b)$. Then there is a map $p : V \to g(W)$ such that:

(i) For all $v \in V$ we have with $\lambda_{\max} = \frac{\mu_{\max}}{1-\mu_{\max}}$ that

$$b(v-p(v), v-p(v)) \le \lambda_{\max} a(v,v).$$

(ii) The map p is a-orthogonal and b-orthogonal, i.e. for all v it holds that

$$a(v-p(v), p(v)) = 0, \quad \text{and} \quad b(v-p(v), p(v)) = 0.$$

Proof. By Lemma 4.5 there is an $\bar{a}$- and $\bar{b}$-orthogonal map $\bar{p} : \bar{V} \to W$ such that

$$\bar{b}(v-\bar{p}(v), v-\bar{p}(v)) \le \lambda_{\max} \bar{a}(v,v)$$

for all $v \in \bar{V}$. With this, define p as the concatenation

$$p : V \xrightarrow{Q} \bar{V} \xrightarrow{\bar{p}} W \xrightarrow{g} g(W)$$

where $Q : V \to \bar{V} = V/\ker(a+b)$ is the quotient map. Then we have for all $v \in V$ that

$$\begin{aligned} b(v-p(v), v-p(v)) &= \bar{b}(Qv - Qg\bar{p}Q(v), Qv - Qg\bar{p}Q(v)) \\ &= \bar{b}_i(Qv - \bar{p}Q(v), Qv - \bar{p}Q(v)) \\ &\le \lambda_{\max} \bar{a}_i(Qv, Qv) \\ &= \lambda_{\max} a_i(v,v). \end{aligned}$$

To see (ii), note that for any $v \in V$ we have

$$a(v-p(v), p(v)) = \bar{a}(Qv - \bar{p}Q(v), \bar{p}Q(v)) = 0$$

since by Lemma 4.5, $\bar{p}$ is $\bar{a}$ orthogonal. Similarly,

$$b(v-p(v), p(v)) = \bar{b}(Qv - \bar{p}Q(v), \bar{p}Q(v)) = 0.$$

□

For theoretical considerations, it can be more convenient to work in the quotient by $\ker(a)$, rather than by $\ker(a+b)$. This is because, in typical applications, it is already known that $a/\ker(a)$ is coercive, even in infinite-dimensional settings. Thus, we investigate the same approximation result for this case. The only property that is lost is the b-orthogonality of the projection p.

Lemma 4.11. *Let V be a separable Banach space and let a, b be two symmetric, positive definite bilinear forms on V. Assume that $\ker(a)$ splits and that $\bar{a}$ is coercive on $\bar{V} := V/\ker(a)$. Let g be a map that splits $\ker(a)$ and define the bilinear form*

$$\bar{b}(\cdot,\cdot) := b(g\cdot, g\cdot)$$

on $\bar{V}$. Suppose that the eigenvectors of the problem

$$\bar{b}(v,\cdot) = \mu(\bar{a}+\bar{b})(v,\cdot)$$

form a dense subset of $\bar{V}$. Let $\mu_{\max} \in (0,1)$ and set

$$W := \sum_{\mu > \mu_{\max}} \ker(\bar{b} - \mu(\bar{a}+\bar{b})) \subset \bar{V}.$$

Then there is a map $p : V \to \ker(a) + g(W)$ such that:

(i) For all $v \in V$ we have with $\lambda_{\max} = \frac{\mu_{\max}}{1-\mu_{\max}}$ that

$$b(v-p(v), v-p(v)) \leq \lambda_{\max} a(v,v).$$

(ii) The map p is a-orthogonal, i.e. for all v it holds that

$$a(v-p(v), p(v)) = 0.$$

Proof. Since $\bar{a}$ is coercive on $\bar{V}$, so is $\bar{a}+\bar{b}$ and thus by Lemma 4.5 there is an $\bar{a}$- and $\bar{b}$-orthogonal map $\bar{p} : \bar{V} \to W$ such that

$$\bar{b}(v-p(v), v-p(v)) \leq \lambda_{\max} \bar{a}(v,v)$$

for all $v \in \bar{V}$. Let $Q : V \to \bar{V}$ be the quotient map and set

$$p : V \to \ker(a) + g(W), \quad v \mapsto (v - gQ(v)) + g\bar{p}Q(v).$$

Note that this is well-defined since $v - gQ(v) \in \ker(a)$. Using the definition of $\bar{b}$ we see that for all $v, w \in V$ it holds

$$\begin{aligned} b(v-p(v),\ v-p(v)) &= b(g(Q(v)-\bar{p}Q(v)),\ g(Q(v)-\bar{p}Q(v))) \\ &= \bar{b}(Qv-\bar{p}(Qv),\ Qv-\bar{p}(Qv)) \\ &\leq \lambda_{\max}\bar{a}(Qv,Qv) \\ &= \lambda_{\max} a(v,v). \end{aligned}$$

Finally, note that for any $v \in V$ we have

$$\begin{aligned} a(v - p(v),\ p(v)) &= a(gQ(v) - g\bar{p}Q(v),\ (v - gQ(v)) + g\bar{p}Q(v)) \\ &= a(gQ(v) - g\bar{p}Q(v),\ g\bar{p}Q(v)) \\ &= \bar{a}(Qv - \bar{p}(Qv),\ \bar{p}(Qv)) \\ &= 0, \end{aligned}$$

where we used the $\bar{a}$-orthogonality of $\bar{p}$. □

Remark 4.12. Note that in the modulo $\ker(a)$-case, the bilinear form $\bar{b}$ is not the quotient $b/\ker(a)$, since we do not necessarily have $\ker(a) \subset \ker(b)$. Rather, in this case $\bar{b}$ depends on the choice of the splitting map g.

To summarize, we have shown that it is sufficient to consider eigenspaces in the quotients $V/ker(a+b)$, or even $V/ker(a)$, to obtain analogies to Lemma 4.5. This will allow us to define spectral elements with singular local eigenproblems.

4.1.3 Spectral Elements, General Case

With the preliminary work above, we can now define spectral elements for cases where the local eigenproblems may be singular. We will encounter these for instance in Example 4.49.

Definition 4.13 (Spectral elements, general case)**.** Let (H, H_0, a) be a suitable presheaf over Ω and let $(\mathcal{H}, \mathcal{H}_0, a)$ be the presheaf of generalized elements for (H, H_0, a) with respect to a cover $\{\omega_i\}_{i\in I}$, a partition of unity $\{\chi_i\}_{i\in I}$ and local approximation spaces $\{S_i\}_{i\in I}$.

For all $i \in I$, let $\widetilde{V}_i \subset H(\omega_i)$ be a closed subspace, and let a_i, b_i be symmetric, positive semi-definite bilinear forms on $\widetilde{V}_i$. Let $\mu_{\max} \in (0,1)$ and assume one of the following two scenarios for all $i \in I$.

(i) Assume $\ker(a_i + b_i)$ is split by a map g_i, and that $(\bar{a}_i + \bar{b}_i)$ is coercive on $\bar{V}_i := \widetilde{V}_i/\ker(a_i + b_i)$. In this scenario, let $\bar{a}_i, \bar{b}_i$ denote the quotients of a_i, b_i by $\ker(a_i + b_i)$. Further, suppose that for all $\mu > \mu_{\max}$ we have

$$S_i \supset g_i(\ker(\bar{b}_i - \mu(\bar{a}_i + \bar{b}_i))).$$

(ii) Assume $\ker(a_i)$ is split by a map g_i and that $\bar{a}_i$ is coercive on $\bar{V}_i := \widetilde{V}_i/\ker(a_i)$. In this case we denote by $\bar{a}_i$ the quotient of a_i on $\bar{V}_i$, and define the bilinear form $\bar{b}_i$ on $\bar{V}_i$ by

$$\bar{b}_i(\bar{v}, \bar{w}) := b_i(g_i\bar{v}, g_i\bar{w}).$$

Suppose that for all $\mu > \mu_{\max}$ we have

$$S_i \supset g_i(\ker(\bar{b}_i - \mu(\bar{a}_i + \bar{b}_i))), \quad \text{and further } S_i \supset \ker(a_i).$$

Then $(\mathcal{H}, \mathcal{H}_0, a)$ is called (the suitable presheaf of) *spectral elements* for (H, H_0, a) with respect to the bilinear forms $\{a_i\}_{i\in I}, \{b_i\}_{i\in I}$, the cover $\{\omega_i\}_{i\in I}$, the partition of unity $\{\chi_i\}_{i\in I}$, the splitting maps g_i, and with eigenvalue threshold

$$\lambda_{\max} := \frac{\mu_{\max}}{1 - \mu_{\max}}.$$

In this setting, the *local eigenproblems* refer to the problems of finding $(\mu, v) \in \mathbb{R} \times \bar{V}_i$ such that

$$\bar{b}_i(v, w) = \mu(\bar{a}_i + \bar{b}_i)(v, w) \quad \text{for all } w \in \bar{V}_i, \tag{4.7}$$

or equivalently

$$\bar{b}_i(v, w) = \lambda \bar{a}_i(v, w) \quad \text{for all } w \in \bar{V}_i, \tag{4.8}$$

where $\lambda = \frac{\mu}{1-\mu}$, see also Notation 4.4.

Remark 4.14. In the case of spectral elements under assumption (i), the choice of the splitting map g_i has no effect on the generalized elements $\mathcal{H}$, provided that $\ker(a_i + b_i) \subset \ker(\chi_i)$. This is because $\mathcal{H}$ is constructed from the components $\chi_i S_i$, not the S_i themselves.

A similar observation holds under assumption (ii), if $\ker(a_i) \subset \ker(\chi_i)$. However, in this case, the local eigenproblems themselves still depend on the choice of the splitting map g_i, as the bilinear forms $\bar{b}_i$ are not canonically defined in this case.

The two separate sets of assumptions in the definition of spectral elements feel a bit cumbersome. The reason we consider both cases here is that (i) is the more appealing case in practice. For one, it is always satisfied in finite dimension. Further, it leaves the bilinear forms $\bar{b}_i$ independent of the splitting maps g_i. So, by the above remark, their choice does not matter at all, as long as $\ker(a_i + b_i) \subset \ker(\chi_i)$. Moreover, in many typical applications we have $\ker(a_i + b_i) = 0$, which means one often does not have to deal about quotients of bilinear forms at all.

By "always satisfied in finite dimension" we refer in particular to the coercivity. A concern is, of course, whether for a shape-regular family of meshes $\mathcal{T}_h$ this coercivity is uniform as $h \to 0$. This is where the set of assumptions (ii) is more attractive. Typically, a_i will be simply a_{ω_i} — the bilinear form associated to the given PDE — and thus showing coercivity of $a_i / \ker(a_i)$ is essentially equivalent to showing that the local solves are well-defined.

In that sense, (ii) seems to be the superior set of assumptions. However, in practice we do not want to deal with local eigenproblems that may depend on splitting maps g_i. We will reconsider this when discussing applications in Section 4.2.4 and Section 4.2.5.

4.1.4 Approximation Property

After the preliminary work of previous sections, it is now easy to state a local approximation result for spectral elements in broad generality.

Theorem 4.15 (Local approximation with spectral elements)**.** *For spectral elements, with the notation from Definition 4.13, assume additionally that for $i \in I$ the span of the eigenvectors of the local eigenproblem is dense in $\bar{V}_i$. Then there is a map $p_i : \widetilde{V}_i \to S_i$ such that*

(i) For all $v \in \widetilde{V}_i$, we have

$$b_i(v - p_i(v), v - p_i(v)) \leq \lambda_{\max} a_i(v, v).$$

(ii) The map p_i is a_i-orthogonal, i.e. for all v it holds that

$$a_i(v - p_i(v), p_i(v)) = 0.$$

Proof. For the first set of assumptions in Definition 4.13 where we work modulo $\ker(a + b)$, this is Lemma 4.10. For the second set assumptions which works modulo $\ker(a)$, this is Lemma 4.11. □

This local approximation result forms the foundation for all robustness and convergence theorems presented in the following. It generalizes the corresponding estimates established for GenEO [56], MS-GFEM [46], and related spectral methods.

4.1.5 Eigenvalue Decay

The local approximation property of spectral elements depends critically on the chosen eigenvalue threshold. A key question is whether a prescribed threshold

$$0 < \lambda_{\max} < \infty$$

can actually be achieved. There are two obstacles to this. First, if the eigenvalues of the local eigenproblem on ω_i accumulate at some value $\lambda_* > 0$, then no finite-dimensional local approximation space S_i can yield a threshold $\lambda_{\max} \leq \lambda_*$. Second, even if 0 is the only accumulation point, the eigenvalues may decay so slowly that the required local spaces S_i become prohibitively large, rendering the construction impractical. These concerns can be addressed using a compactness argument, as outlined in the following.

Proposition 4.16. *Consider spectral elements with the notation from Definition 4.13. For each $i \in I$, let (X_i, c_i) be a separable Hilbert space. Assume that the local bilinear forms b_i are given by*

$$b_i(\cdot, \cdot) := c_i(T_i \cdot, T_i \cdot)$$

for a map $T_i : \widetilde{V}_i \to X_i$. Then, if T_i is compact, the corresponding local eigenproblem has a dense set of eigenvectors, and zero is the only accumulation point of its eigenvalues.

Proof. Let $i \in I$. Recall that the local eigenproblems are defined on the quotient space $\bar{V}_i$. First, assume that we work modulo $\ker(a_i + b_i)$, i.e. with assumption (i) in Definition 4.13. In this case, we consider

$$\bar{V}_i = \widetilde{V}_i / \ker(a_i + b_i).$$

Let $\bar{T}_i := T_i \circ g_i$, where $g_i : \bar{V}_i \to \widetilde{V}_i$ is the splitting map from Definition 4.13. We have

$$\bar{b}_i(\cdot, \cdot) = b_i(g_i \cdot, g_i \cdot) = c_i(\bar{T}_i \cdot, \bar{T}_i \cdot),$$

and the map $\bar{T}_i : (\bar{V}_i, \bar{a}_i + \bar{b}_i) \to (X_i, c_i)$ has a well-defined adjoint $\bar{T}_i^*$, which satisfies

$$c_i(\bar{T}_i \cdot, \bar{T}_i \cdot) = (\bar{a}_i + \bar{b}_i)(\bar{T}_i^* \bar{T}_i \cdot, \cdot).$$

By Proposition 1.18, $\bar{T}_i$ is compact, and thus $\bar{T}_i^* \bar{T}_i$ is a self-adjoint compact operator. Moreover, the local eigenproblems read

$$(\bar{a}_i + \bar{b}_i)(\bar{T}_i^* \bar{T}_i v, \cdot) = \mu(\bar{a}_i + \bar{b}_i)(v, \cdot)$$

which is equivalent to $\bar{T}_i^* \bar{T}_i v = \mu v$ by the representation theorem. Thus, the result follows with Theorem 1.22.

Now assume we are working modulo $\ker(a_i)$, i.e. with assumption (ii) in Definition 4.13. Then we have $\bar{V}_i = \widetilde{V}_i / \ker(a_i)$ and the result follows exactly as in the first case, by simply replacing any appearance of $(\bar{a}_i + \bar{b}_i)$ with $\bar{a}_i$. □

The above proposition allows the use of any Hilbert space (X_i, c_i) in the definition of b_i, as it is only concerned with the spectrum of the arising local eigenproblem.

Whether a particular choice of X_i leads to a meaningful local approximation property in Theorem 4.15 is a separate question. Notably, the setting in Proposition 4.16 applies to all three spectral methods discussed in Chapter 2, except that the T_i are not necessarily compact. This is precisely where the concept of oversampling becomes important, as explained below.

Definition 4.17 (Factorization of maps)**.** Let f and g be maps. We say that f *factors through* g if there is a map ψ such that we have

$$f = \psi \circ g.$$

If f and g are morphisms in a particular category — say continuous, linear maps between separable Banach spaces — then ψ is required to also be a morphism in that category.

Remark 4.18. Let g be a compact map between normed vectors spaces. Then, by Proposition 1.18, any map that factors through g is compact.

Definition 4.19 (Factorization through subsets)**.** Let (H, H_0, a) be a presheaf with extendables over Ω. Let X be a normed vector space, let $\omega \subset \omega^* \subset \Omega$ be open and consider a subspace $\widetilde{V} \subset H(\omega^*)$. We say that a map $f : \widetilde{V} \to X$ *factors through* ω if it factors through the restriction $(R_{\omega^*,\omega})\,|_{\widetilde{V}}$.

Remark 4.20. Let (H, H_0, a) be a presheaf with extendables over Ω, let

$$f : H(\omega) \to H(\omega)$$

be a map, and let $\operatorname{supp}(f)$ be a supporting set of f (cf. Definition 3.35). Then f factors through $\operatorname{supp}(f)$. In particular, partition of unity operators χ_i factor through their supporting sets.

Definition 4.21 (Oversampling)**.** Consider spectral elements with the notation from Definition 4.13. For each $i \in I$, let (X_i, c_i) be a separable Hilbert space. Assume the local bilinear forms b_i are given by

$$b_i(\cdot, \cdot) := c_i(T_i \cdot, T_i \cdot)$$

for a (continuous and linear) map $T_i : \widetilde{V}_i \to X_i$. Assume further that each T_i factors through an open set $\hat{\omega}_i \subset \omega_i$. Then, if Ω is equipped with a notion of distance, the *amount of oversampling* is defined as

$$\min_{i \in I} \operatorname{dist}(\partial\omega_i, \hat{\omega}_i).$$

We speak of spectral elements *with oversampling* if the amount of oversampling is strictly positive.

Remark 4.22. It is more common to define oversampling by introducing larger sets $\omega_i^* \supset \omega_i$, such as in [46, 42] — rather than by introducing smaller $\hat{\omega}_i \subset \omega_i$. This is a less appealing viewpoint, however, at least for the purpose of this thesis: with the traditional notation, all ω_i in this document would have to be renamed to ω_i^*. In contrast, the $\hat{\omega}_i$ only appear when discussing compactness of the restrictions.

Example 4.23. For Full GenEO and MS-GFEM — to be discussed for the presheaf setting in Section 4.2 — we will have

$$X_i = H_0(\omega_i) \quad \text{and} \quad T_i = (\chi_i)\,|_{\widetilde{V}_i},$$

where χ_i is the partition of unity operator. Thus, choosing a partition of unity with supporting sets $\operatorname{supp}(\chi_i)$ that are smaller than needed, i.e. such that

$$\operatorname{dist}(\partial\omega_i, \operatorname{supp}(\chi_i)) > 0,$$

equates to oversampling for these methods.

Example 4.24. Another common choice for the T_i is to directly take restrictions to smaller subsets, i.e.

$$X_i = H(\hat{\omega}_i) \quad \text{and} \quad T_i = R_{\omega_i,\hat{\omega}_i}$$

for $\hat{\omega}_i \subset \omega_i$. Consider, for instance, the GMsFEM eigenproblem (2.7), but with the left-hand side integrated only over $\hat{\omega}_i$.

Proposition 4.25. *Consider spectral elements with oversampling in the notation from Definition 4.21, and assume that the restrictions*

$$(R_{\omega_i,\hat{\omega}_i})\,|_{\widetilde{V}_i}$$

are compact. Then the local eigenproblem have a dense set of eigenvectors, and zero is the only accumulation point of their eigenvalues.

Proof. By assumption, the T_i factor through compact maps and are thus compact themselves. The result then follows with Proposition 4.16. □

In typical applications, the restriction operators are not compact on the full spaces $\widetilde{V}_i = H(\omega_i)$, but they often are compact when restricted to the locally harmonic spaces

$$\widetilde{V}_i = H_0(\omega_i)^{\perp_a} \subset H(\omega_i).$$

In such cases, oversampling ensures that the eigenvalues have no accumulation point greater than zero. The plots in Figure 4.1 of the spectra of GenEO ($\widetilde{V}_i = H(\omega_i)$) and MS-GFEM ($\widetilde{V}_i = H_0(\omega_i)^{\perp_a}$) demonstrate this behaviour.

It may not be immediately obvious how the theoretical result applies in practice, since we generally work with finite-dimensional discretizations over a fine mesh. In this case, all linear maps are compact and there are only finitely many eigenvalues, so accumulation points do not occur. Thus, the above results are automatically satisfied. However, we usually consider discretizations over a family of shape-regular meshes of size h. As $h \to 0$, the finite-dimensional spaces approximate the infinite-dimensional setting. In many applications, this holds not only for the original PDE problem, but also for the given local eigenproblems. Therefore, decay of the eigenvalues in the infinite-dimensional setting often carries over to their finite-dimensional discretizations in the limit $h \to 0$. Conversely, if the infinite-dimensional eigenproblem has an accumulation point at $\lambda = x$, then the discretization will typically exhibit a large cluster of eigenvalues near x.

While the outlined compactness argument guarantees eigenvalue decay to 0, it does not offer a guaranteed asymptotic rate. This is because compactness is a qualitative property that is either given or not. The procedure can be refined by considering the n-width of the $R_{\omega_i,\hat{\omega}_i}$, which essentially quantifies their compactness. For MS-GFEM, this approach leads to a guaranteed near-exponential decay rate [46, 42] under assumptions that we briefly review in Section 4.2.3.

4.2 Examples

In this thesis, we consider three examples of spectral elements that can be formulated within the general framework of suitable presheaves: Full GenEO [57], GenEO [56], and MS-GFEM [46]. We outlined the intuition and historical background of these methods in Chapter 2. For all the discussed methods, we show that

$$\ker(a_i + b_i) \subset \ker(a_i) \cap \ker(\chi_i), \tag{4.9}$$

which typically leads to $\ker(a_i + b_i) = 0$ in applications. We also discuss the various arising local approximation results. Notably, of the three methods, only MS-GFEM exhibits decay of its local eigenproblems to zero. This comes at the cost of significantly more expensive local eigenproblems, however, as we will investigate in later chapters.

We then consider the application of these spectral methods to previously discussed examples of suitable presheaves: H^1 functions, continuous finite elements, and Discontinuous Galerkin (DG) elements. In principle, there is not much to say, since these were already shown to be suitable presheaves and spectral elements were constructed to work in generality in this context. The main concern in practice is how to work in the quotient by $\ker(a_i + b_i)$. Ideally, it simply holds that $\ker(a_i + b_i) = 0$, which is often the case due to (4.9).

For DG, there is an additional question: whether one can ignore volume integrals over outer cells. This can indeed be done, and we will adopt this approach in variations of the abstract GenEO eigenproblems. This leads to more technical considerations, and we will encounter the case where $\ker(a_i + b_i) \neq 0$.

4.2.1 Full GenEO

The Full GenEO eigenproblems are a variant of the standard GenEO approach (Generalized Eigenproblems in the Overlap). They exhibit slightly improved approximation properties compared to GenEO, and are described in [57]. We present them here within the abstract framework of a suitable presheaf, as they serve as a common foundation for both GenEO and MS-GFEM.

Example 4.26 (Full GenEO)**.** Consider spectral elements with the notation from Definition 4.13 where we set $\widetilde{V}_i = H(\omega_i)$ and

$$\begin{aligned} a_i(v, w) &:= a_{\omega_i}(v, w), \\ b_i(v, w) &:= a_{\omega_i}(\chi_i v, \chi_i w). \end{aligned}$$

We call this setup spectral elements with Full GenEO eigenproblems, or simply Full GenEO elements.

Remark 4.27. In the setting of Full GenEO elements, we have $\ker(b_i) = \ker(\chi_i)$, and thus $\ker(a_i + b_i) \subset \ker(a_i) \cap \ker(\chi_i)$ for all i.

Proof. Simply note that the range of χ_i lies in $H_0(\omega_i)$, where a_{ω_i} is coercive. □

Lemma 4.28 (Local approximation with Full GenEO)**.** *Consider Full GenEO elements with the notation from Definition 4.13, in particular with threshold $\lambda_{\max}$. Then, for all $i \in I$, there is a map $p_i : H(\omega_i) \to S_i$ such that for any $\omega^* \supset \omega_i$ and with $R = R_{\omega^*,\omega_i}$, the following hold.*

(i) For all $v \in H(\omega^*)$, we have

$$a_{\omega_i}(\chi_i(Rv - p_i(Rv)), \chi_i(Rv - p_i(Rv))) \leq \lambda_{\max} a_{\omega_i}(Rv, Rv).$$

(ii) The map p_i is a_{ω_i}-orthogonal, i.e. for all $v \in H(\omega^*)$, it holds that

$$a_{\omega_i}(Rv - p_i(Rv), p_i(Rv)) = 0.$$

Proof. These are exactly the properties of p_i from Theorem 4.15. Note that here we use $\widetilde{V}_i = H(\omega_i)$. □

Among the examples considered here, Full GenEO provides the strongest local approximation property. However, this comes with a significant drawback: the eigenvalues of the Full GenEO eigenproblems accumulate near $\lambda = 1$. As a result, the threshold $\lambda_{\max}$ can be at best equal to 1.

4.2.2 GenEO

As discussed in Chapter 2, GenEO is a popular spectral method for PDE problems with highly heterogeneous coefficients. Here, we define it within our abstract framework. The main difference from Full GenEO is that the b_i in this case disregard what happens in the interior of their respective domains.

Example 4.29 (GenEO)**.** In the notation from Definition 4.13, consider spectral elements with $\widetilde{V}_i = H(\omega_i)$ and

$$\begin{aligned} a_i(v, w) &:= a_{\omega_i}(v, w), \\ b_i(v, w) &:= a_{\omega_i^{\text{ovlp}}}(\chi_i v, \chi_i w), \end{aligned}$$

where ω_i^{ovlp} is defined as the complement of the largest open subset of ω_i where χ_i restricts to the identity. More precisely, let

$$U_i := \bigcup_{U \in \mathcal{U}_i} U,$$

with

$$\mathcal{U}_i := \{U \subset \omega_i \text{ open} \mid \text{ for all } v \in \widetilde{V}_i : \ (\chi_i v)|_U = v|_U\},$$

and we define

$$\omega_i^{\text{ovlp}} := \omega_i \setminus^{\circ} U_i.$$

We call this setup spectral elements with GenEO eigenproblems, or simply GenEO elements.

The intuition behind restricting the Full GenEO eigenproblem in this way is that, for the coarse space, one only cares about what happens along the boundaries of the subdomains ω_i. This is because we aim to use this within a Schwarz method, which handles everything in $H_0(\omega_i)$ via local solves already. In particular, for any $v \in H_0(U_i) \subset H(\omega_i)$, it holds by locality of $a(\cdot, \cdot)$ that

$$a_{\omega_i}(\chi_i v, \chi_i w) = \underbrace{a_{\omega_i^{\text{ovlp}}}(\chi_i v, \chi_i w)}_{=0} + a_{U_i}(v, w) = \underbrace{a_{\omega_i^{\text{ovlp}}}(v, w)}_{=0} + a_{U_i}(v, w) = a_{\omega_i}(v, w).$$

Thus, any such v is an eigenvector of the Full GenEO eigenproblem with eigenvalue $\lambda = 1$. In contrast, for the GenEO eigenproblem, the same functions are eigenvectors with $\lambda = 0$. None of these elements of $H_0(U_i)$ are likely to contribute meaningfully to the coarse space S_i in two-level Schwarz schemes, since their contribution to the

global solution can already be resolved by the local solves. The GenEO eigenproblems are designed to completely ignore them — not just by assigning them eigenvalue zero, but also by excluding them from the computation of b_i altogether. This yields two advantages. First, it reduces the accumulation of eigenvalues around $\lambda = 1$. Second, it theoretically lowers the cost of the eigensolves, as many entries in the matrix representation of b_i are zero.

However, in practice, the cluster of eigenvalues remains too large to achieve a threshold $\lambda_{\max} < 1$. Moreover, the speed-up is negligible in typical implementations. Overall, the differences between GenEO and Full GenEO seem to be minor in practical settings.

The theoretical results established for Full GenEO largely still apply, although the approximation bound deteriorates slightly: the constant in Lemma 4.31 becomes $(1 + \lambda_{\max})$.

Proposition 4.30. *In the setting of GenEO elements, we have* $\ker(a_i + b_i) \subset \ker(a_i) \cap \ker(\chi_i)$ *for all* $i \in I$.

Proof. By locality of $a(\cdot,\cdot)$, and with U_i as in Example 4.29, we have

$$a_{U_i}(v, w) \leq a_{\omega_i}(v, w)$$

for all $v, w \in H(\omega_i)$. Now let $v \in \ker(a_i + b_i)$. Then, again by locality of $a(\cdot,\cdot)$, it holds that

$$\begin{aligned} 0 = b_i(v, v) &= a_{\omega_i^{\text{ovlp}}}(\chi_i v, \chi_i v) + \underbrace{a_{U_i}(v, v)}_{=0} \\ &= a_{\omega_i^{\text{ovlp}}}(\chi_i v, \chi_i v) + a_{U_i}(\chi_i v, \chi_i v) \\ &= a_{\omega_i}(\chi_i v, \chi_i v). \end{aligned}$$

Moreover, the range of χ_i lies in $H_0(\omega_i)$, where a_{ω_i} is coercive. Thus, $\chi_i v = 0$. □

Lemma 4.31 (Local approximation with GenEO)**.** *Consider Full GenEO elements with the notation from Definition 4.13, in particular with threshold* $\lambda_{\max}$. *Then for all* $i \in I$ *there is a map* $p_i : H(\omega_i) \to S_i$ *such that for any* $\omega^* \supset \omega_i$ *and with* $R = R_{\omega^*,\omega_i}$ *the following hold.*

(i) For all $v \in H(\omega^*)$, *we have*

$$a_{\omega_i}(\chi_i(Rv - p_i(Rv)), \chi_i(Rv - p_i(Rv))) \leq (1 + \lambda_{\max}) a_{\omega_i}(Rv, Rv).$$

(ii) The map p_i *is* a_{ω_i}*-orthogonal, i.e. for all* $v \in H(\omega^*)$ *it holds that*

$$a_{\omega_i}(Rv - p_i(Rv), p_i(Rv)) = 0.$$

Proof. As for Full GenEO, we have $\widetilde{V}_i = H(\omega_i)$, and we again use the p_i from Theorem 4.15. Statement (ii) follows directly from Theorem 4.15 since $a_{\omega_i} = a_i$.

Now recall that the partition of unity is the identity on $U_i = (\omega_i \backslash \mathring{\omega}^{\text{ovlp}})$, i.e. $R_{\omega_i,U_i} \circ \chi_i = R_{\omega_i,U_i}$. With this and locality of $a(\cdot,\cdot)$, we have

$$\begin{aligned} a_{\omega_i}(\chi_i(v - p_i v), \chi_i(v - p_i v)) &= \left(a_{U_i} + a_{\omega_i^{\text{ovlp}}}\right)(\chi_i(v - p_i v), \chi_i(v - p_i v)) \\ &= a_{U_i}(v - p_i v,\ v - p_i v) + b_i(v - p_i v,\ v - p_i v) \\ &\leq a_{\omega_i}(v - p_i v,\ v - p_i v) + \lambda_{\max} a_{\omega_i}(v\, v), \end{aligned}$$

where in the last step we used Proposition 3.15 for the first term and Theorem 4.15 for the second. By a_{ω_i}-orthogonality of p_i we further get

$$a_{\omega_i}(v - p_i v,\ v - p_i v) \leq a_{\omega_i}(v,\ v),$$

yielding (i). □

4.2.3 MS-GFEM

So far, we have not provided bounds on the number of eigenvectors that must be included in the local approximation spaces S_i to achieve a prescribed threshold $\lambda_{\max}$. MS-GFEM [46] addresses this by guaranteeing near-exponential decay of the local eigenvalues under reasonable assumptions. It does so by restricting the Full GenEO eigenproblems to the a-harmonic subspaces $\widetilde{V}_i = H_0(\omega_i)^{\perp_a} \subset H(\omega_i)$. The guiding idea, similar to GenEO, is that components already resolved by the local solves need not be represented in the approximation spaces. Compared to GenEO, MS-GFEM offers a significant improvement, in particular by eliminating the eigenvalue cluster near 1, making it possible in practice to achieve arbitrarily small values of $\lambda_{\max}$.

This improvement, however, comes at the cost of significantly more complex eigenproblems. While the problems are defined over the smaller space $H_0(\omega_i)^{\perp_a}$, an explicit basis for this space is typically unavailable. Consequently, the orthogonality constraint must be incorporated directly into the eigenproblem. Although there are proposals for efficient implementations (see [45]), in the author's experience the eigensolves still tend to be substantially more expensive than those of GenEO or Full GenEO.

Example 4.32 (MS-GFEM)**.** In the notation from Definition 4.13, spectral elements with respect to $\widetilde{V}_i = H_0(\omega_i)^{\perp_a} \subset H(\omega_i)$ and

$$\begin{aligned} a_i(v, w) &:= a_{\omega_i}(v, w) \\ b_i(v, w) &:= a_{\omega_i}(\chi_i v, \chi_i w) \end{aligned}$$

are called spectral elements with MS-GFEM eigenproblems, or just MS-GFEM elements. Note that MS-GFEM is typically employed with oversampling; see Definition 4.21.

Remark 4.33. Since the MS-GFEM eigenproblems are a restriction of the Full GenEO eigenproblems, here we likewise have $\ker(a_i + b_i) \subset \ker(a_i) \cap \ker(\chi_i)$ for all i.

Lemma 4.34 (Local approximation with MS-GFEM)**.** *Consider MS-GFEM elements with the notation from Definition 4.13. Then, for all $i \in I$, there is a map*

$$p_i : H(\omega_i)^{\perp_a} \to S_i$$

such that for any $\omega^* \supset \omega_i$, and with $R = R_{\omega^*,\omega_i}$, $\pi = \pi_{\omega_i,\omega^*}$, the following holds: for all $v \in H_0(\omega^*)$ we have for $\widetilde{v} = Rv - \pi v$ that

$$a_{\omega_i}(\chi_i(\widetilde{v} - p_i(\widetilde{v})), \chi_i(\widetilde{v} - p_i(\widetilde{v}))) \leq (1 + \lambda_{\max}) a_{\omega_i}(Rv, Rv).$$

Proof. This is again the p_i from Theorem 4.15. Here, we have to account for the fact that $\widetilde{V}_i = H_0(\omega_i)^{\perp_a} \neq H(\omega_i)$. This is why the result is stated only for $Rv - \pi v$, rather than v itself. Simply note that $Rv - \pi v \in H_0(\omega_i)^{\perp_a}$ since for any $w \in H_0(\omega_i)$ we have that

$$a_{\omega_i}(Rv - \pi v, w) = a_{\omega_i}(Rv - R\iota\pi v, w) = a_{\omega^*}(v - \iota\pi v, \iota w) = 0$$

by definition of π as the a-orthogonal projection $H_0(\omega^*) \to H_0(\omega_i)$. □

Notably, the above approximation result appears strictly weaker than that of Full GenEO, since it only holds in the a-harmonic space $H_0(\omega_i)^{\perp_a}$. In return however, the achievable eigenvalue threshold $\lambda_{\max}$ for MS-GFEM is arbitrarily small in typical applications. We cite the corresponding result here, adjusted to our framework for elliptic PDEs. In principle, the strategy is to make use of oversampling and the n-width of compact restrictions, as outlined in Section 4.1.5. Details can be found in [42].

Notation 4.35. For the remainder of this subsection, let (H, H_0, a) be a suitable presheaf. Further, let L be a sheaf of separable Hilbert spaces over $(\Omega, \mathcal{O})$, whose inner product is local. Moreover, we assume there is an embedding of sheaves

$$H \hookrightarrow L.$$

Definition 4.36 (Caccioppoli-type inequality)**.** If there exists a constant $C_{\text{cac}} > 0$ such that for any $\omega \subset \omega^* \in \mathcal{O}$ with $\delta := \operatorname{dist}(\omega, \partial\omega^* \setminus \partial\Omega) > 0$ we have

$$\|v\|_{a_\omega} \leq \frac{C_{\text{cac}}}{\delta} \|v\|_{L(\omega^* \setminus \omega)} \quad \text{for all } v \in H_0(\omega^*)^{\perp_a},$$

we say that the embedding $H \hookrightarrow L$ satisfies a Caccioppoli-type inequality.

Definition 4.37 (Weak approximation property)**.** We say that the embedding $H \hookrightarrow L$ satisfies the weak approximation property if the following holds. Let $\omega \subset \omega^* \subset \omega^{**} \in \mathcal{O}$ such that

$$\delta := \operatorname{dist}(\omega^*, \partial\omega^{**} \setminus \partial\Omega) > 0,$$

and assume that $\operatorname{dist}(\boldsymbol{x}, \partial\omega^* \setminus \partial\Omega) \leq \delta$ for all $\boldsymbol{x} \in \partial\omega \setminus \partial\Omega$. Then for each $m \in \mathbb{N}$, there exists an m-dimensional space $Q_m(\omega^{**}) \subset L(\omega^{**})$ such that for all $v \in H_0(\omega^{**})^{\perp_a}$,

$$\inf_{w \in Q_m(\omega^{**})} \|v - w\|_{L(\omega^* \setminus \omega)} \leq C_{\text{wa}} |\mathsf{V}_\delta(\omega^* \setminus \omega)|^\alpha m^{-\alpha} \|v\|_{a_{\omega^*}}, \tag{4.10}$$

where

$$\mathsf{V}_\delta(\omega^* \setminus \omega) := \{\boldsymbol{x} \in \omega^{**} : \operatorname{dist}(\boldsymbol{x}, \omega^* \setminus \omega) \leq \delta\}$$

is the δ-vicinity of the set $\omega^* \setminus \omega$, and C_{wa} and α are positive constants independent of ω^{**}, ω^*, ω, and m.

Theorem 4.38 (Exponential convergence for local approximations)**.** *Consider the MS-GFEM eigenproblems for H and let $\lambda_{i,k}$ be the k-th largest eigenvalue on the subdomain ω_i. Let further the Caccioppoli-type inequality and the weak approximation property be satisfied. Then, there exist positive constants k_0, b, and C, such that for any $k > k_0$,*

$$\lambda_{i,k}^{1/2} \leq C\, e^{-bk^\alpha}, \tag{4.11}$$

where C only depends on the partition of unity operator χ_i, $\alpha > 0$ is the same constant as in the weak approximation property, and b is related to the amount of oversampling δ by

$$b \propto \left(\frac{\delta}{diam(\omega_i)}\right)^{\alpha - 1}$$

Proof. [42, Theorem 3.8] □

For a discussion of the assumptions of the above theorem, and many applications where they are satisfied, see [42]. They in particular hold for the diffusion equation as well as for linear elasticity, the examples considered in this thesis.

4.2.4 Application to Continuous Settings

In applications, the first step is to verify whether a given problem fits into the framework of a suitable presheaf. This was established in Section 3.2 for the diffusion equation using H^1 functions, continuous finite elements, and the discontinuous WSIP discretization. Consequently, generalized elements can be applied in these cases. For spectral elements to be well-defined, several properties must be checked (cf. Definition 4.13). The discussion that follows applies equally to Full GenEO, GenEO, and MS-GFEM.

First, symmetry and positive semi-definiteness of the bilinear forms a_i and b_i follow directly from those of a_{ω_i}. Second, in the continuous setting, we work with separable Hilbert spaces where all closed subspaces split, and kernels are always closed. Density of eigenvectors is ensured if the bilinear forms b_i can be expressed in terms of self-adjoint compact operators. This is automatically satisfied in finite-dimensional settings, and, as discussed earlier, also in many infinite-dimensional cases when $\widetilde{V}_i = H(\omega_i)^{\perp_a}$ and oversampling is used.

Finally, coercivity of $a_i + b_i$ is trivial in finite dimensions when working modulo $\ker(a_i + b_i)$. For theoretical considerations in infinite dimensions on the other hand, the setting modulo $\ker(a_i)$ may be preferable: for the discussed methods we have $a_i = a_{\omega_i}$ and thus coercivity of $a_i / \ker(a_i)$ is well-established for the problems considered in this thesis. Figure 4.1 offers a demonstration that the spectra of the local eigenproblems decay rapidly enough to be practically useful. For MS-GFEM, this observation is further supported by the results on exponential eigenvalue decay from [42], which we briefly reviewed in Section 4.2.3.

The remaining practical issue is how to work in the quotient space

$$\bar{V}_i = \widetilde{V}_i / \ker(a_i + b_i).$$

Fortunately, in many typical applications one simply has $\ker(a_i + b_i) = 0$. The standard argument for this is as follows. We have shown that for the eigenproblems considered here, it holds

$$\ker(a_i + b_i) \subset \ker(a_i) \cap \ker(chi_i).$$

Now, $\ker(a_i)$ is typically known to be finite-dimensional. For instance, it consists of constant functions in the case of the diffusion equation, or of rigid body modes in linear elasticity. It then suffices to check that these do not lie in the kernel of χ_i.

Although the abstract definition of a partition of unity leaves considerable flexibility, in practice the χ_i are typically implemented as diagonal matrices with entries in $[0, 1]$. These arise from pointwise multiplication with a partition of unity function, followed by interpolation into the finite element basis. With this construction, it is easy to see that applying χ_i to a function with full support in ω_i does not yield zero unless $\chi_i = 0$ everywhere. But if $\chi_i = 0$ on all degrees of freedom, then the corresponding subdomain is ignored during the construction of generalized elements — so the question whether its eigenproblem is well-posed becomes irrelevant.

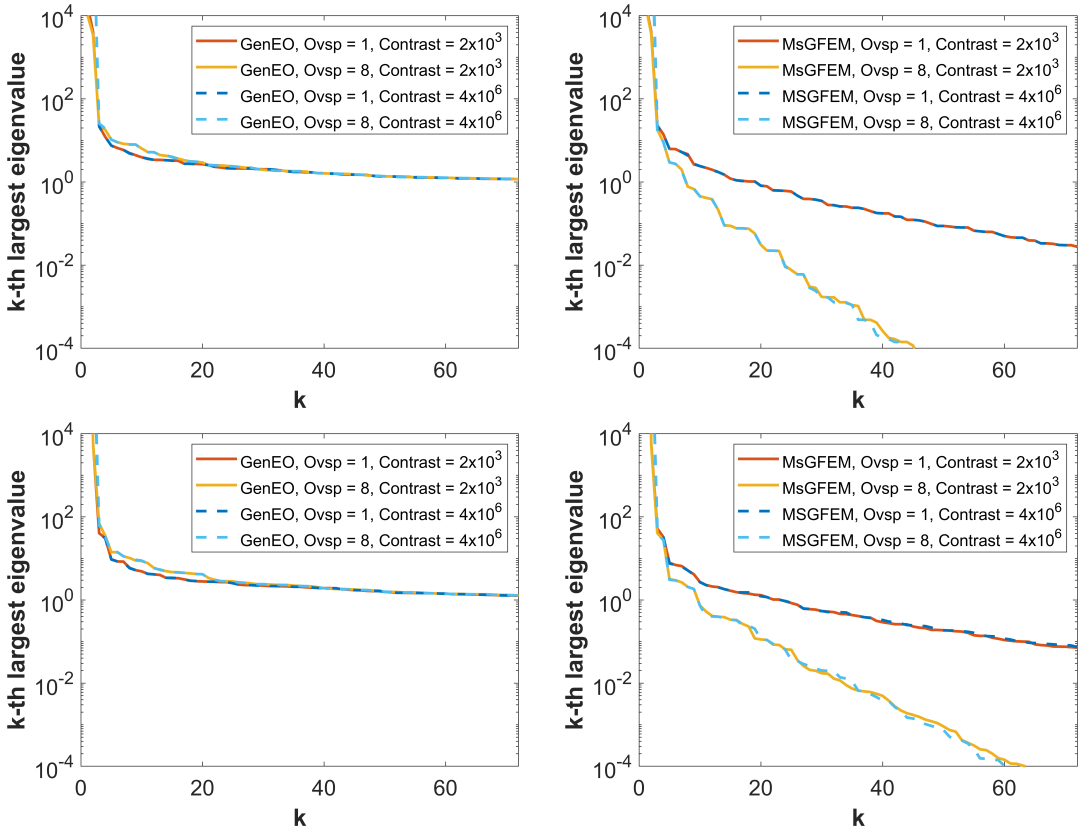


FIGURE 4.1: Spectra of local eigenproblems for MS-GFEM and GenEO for the subdomains shown in Figure 4.2, for varying coefficient contrast and amount of oversampling. Top: boundary subdomain. Bottom: interior subdomain. Details on the experiment are given in Section 6.3.2.

4.2.5 Variants for Discontinuous Galerkin

Recall from Section 3.2.2 that we defined the DG presheaf (H, H_0, a^h) for the heterogeneous diffusion equation, based on the WSIP method. We showed this presheaf to be suitable, provided that the penalty parameter α is chosen large enough. Throughout this section, we assume that this condition is satisfied and continue to use the notation from Section 3.2.2.

Example 4.39 (Spectral elements for DG)**.** For the Full GenEO, GenEO, and MS-GFEM eigenproblems, essentially the same considerations apply to the DG presheaf (H, H_0, a^h) as in the continuous setting. Suppose that $p(\mathcal{T}((\omega_i))$ is connected; that is, the union of the inner cells of ω_i forms a connected domain. Then, by Corollary 3.51, the kernel of a_i consists precisely of the constant functions. Consequently, $\ker(a_i + b_i)$ consists of constant functions that are also in $\ker(\chi_i)$. In practice, this typically means that $\ker(a_i + b_i) = 0$. Since we are working in a finite-dimensional setting, density of the eigenvectors is not an issue.

Thus, the mentioned spectral methods can be applied to the DG sheaf without any modifications. However, the bilinear form a^h includes contributions from volume integrals over outer cells, essentially forcing one cell layer of oversampling. For computing the local projections π, we argued that these outer contributions are always zero (cf. Remark 3.54). However, this is not necessarily true for the local

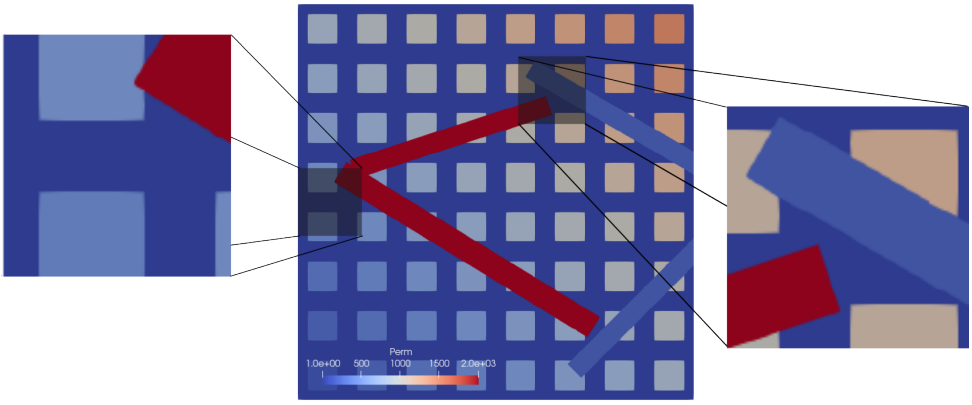


FIGURE 4.2: Diffusion coefficient and subdomain selection for Figure 4.1. Left: boundary subdomain corresponding to the top row in Figure 4.1. Right: interior subdomain corresponding to the bottom row in Figure 4.1.

eigenproblems — in those, the outer terms do appear. In the following, we examine whether these terms can still be neglected and, if so, what the implications are.

First, observe that for the bilinear forms b_i, there is no issue. These are defined as variants of $a^h_{\omega_i}(\chi_i \cdot, \chi_i \cdot)$, and since $\mathrm{Image}(\chi_i) \subset H_0(\omega_i)$, their contributions from the outer cells vanish by Lemma 3.53. However, the same does not hold for the forms $a_i = a^h_{\omega_i}$. To address this, we now define a modified bilinear form on H in which the outer contributions are excluded. Recall that we are working with subdomains $\omega \subset \mathcal{F}$, but often also reason in terms of $\tilde{\mathcal{F}}$, where each face is attached to a unique mesh cell — meaning that interior faces appear twice.

Given any $\omega \subset \mathcal{F}$, define $\widetilde{\omega}$ to be the set of all faces in ω that do not belong to, or touch, an outer cell:

$$\widetilde{\omega} := \{F \in \omega \mid \mathcal{T}(\{F\}) \cap \partial\omega = \emptyset\}.$$

Now consider $a^h_{\widetilde{\omega}}$ as a bilinear form on $H(\omega)$. This in particular excludes all volume and face contributions from $\partial\omega$. Moreover, by the locality of a^h, it satisfies the inequality

$$a^h_{\widetilde{\omega}}(v, v) \leq a^h_{\omega}(v, v) \text{ for all } v \in H(\omega).$$

We can use $a^h_{\widetilde{\omega}}$ to define GenEO variants that yield similar approximation results as in Example 4.26 and Example 4.29.

Example 4.40 (Full GenEO, DG Variant 1)**.** Let (H, H_0, a^h) be the DG presheaf, and otherwise use the notation from Definition 4.13. Consider spectral elements with respect to $\widetilde{V}_i = H(\omega_i)$ and

$$\begin{aligned} a_i(v, w) &:= a^h_{\widetilde{\omega}_i}(v, w), \\ b_i(v, w) &:= a^h_{\omega_i}(\chi_i v, \chi_i w). \end{aligned}$$

We refer to this as DG Variant 1 for Full GenEO.

Lemma 4.41 (Local approximation with Full GenEO, DG Variant 1)**.** *Consider the Full GenEO variant from Example 4.40, using the notation from Definition 4.13. Then, for all $i \in I$, there exists a map $p_i : H(\omega_i) \to S_i$ such that for any $\omega^* \supset \omega_i$, and with $R = R_{\omega^*, \omega_i}$, the following holds.*

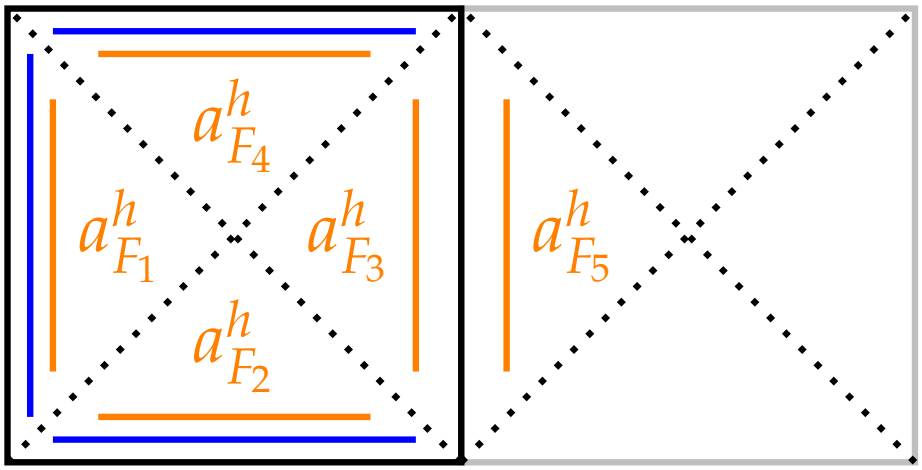


FIGURE 4.3: Continuation of Figure 3.1. $\omega \subset \mathcal{F}$ shown in black, $P^{-1}(\omega) \subset \dot{\mathcal{F}}$ in orange, and $\widetilde{\omega} \subset \dot{\mathcal{F}}$ in blue.

(i) For all $v \in H(\omega^*)$, we have

$$a^h_{\omega_i}(\chi_i(Rv - p_i(Rv)),\ \chi_i(Rv - p_i(Rv))) \leq \lambda_{\max}\, a^h_{\omega_i}(Rv, Rv).$$

(ii) The projection is $a^h_{\widetilde{\omega}_i}$-orthogonal, i.e. for all $v \in H(\omega^*)$ it holds that

$$a^h_{\widetilde{\omega}_i}(Rv - p_i(Rv),\ p_i(Rv)) = 0.$$

Proof. We use the p_i from Theorem 4.15. Statement (ii) follows directly, since $a^h_{\widetilde{\omega}_i} = a_i$. Moreover,

$$\begin{aligned} a^h_{\omega_i}(\chi_i(v - p_i v),\ \chi_i(v - p_i v)) &= b_i(v - p_i v,\ v - p_i v) \\ &\leq \lambda_{\max}\, a^h_{\widetilde{\omega}_i}(v - p_i v,\ v - p_i v). \end{aligned}$$

By $a^h_{\widetilde{\omega}_i}$-orthogonality of p_i, we further have

$$a^h_{\widetilde{\omega}_i}(v - p_i v,\ v - p_i v) \leq a^h_{\widetilde{\omega}_i}(v,\ v),$$

and by locality of a^h,

$$a^h_{\widetilde{\omega}_i}(v,\ v) \leq a^h_{\omega_i}(v,\ v).$$

□

The same can be done for (non-full) GenEO of course.

Example 4.42 (GenEO, DG Variant 1)**.** Let (H, H_0, a^h) be the DG presheaf, and otherwise use the notation from Definition 4.13. Consider spectral elements with respect to $\widetilde{V}_i = H(\omega_i)$ and

$$\begin{aligned} a_i(v, w) &:= a^h_{\widetilde{\omega}_i}(v, w), \\ b_i(v, w) &:= a^h_{\omega_i^{\text{ovlp}}}(\chi_i v, \chi_i w), \end{aligned}$$

where ω_i^{ovlp} is defined as in Example 4.29. We refer to this as DG Variant 1 for GenEO.

Lemma 4.43 (Local approximation with GenEO, DG Variant 1)**.** *Consider the GenEO variant from Example 4.42, using the notation from Definition 4.13. Then, for all $i \in I$, there exists a map $p_i : H(\omega_i) \to S_i$ such that for any $\omega^* \supset \omega_i$, and with $R = R_{\omega^*,\omega_i}$, the following holds.*

(i) For all $v \in H(\omega^*)$, we have

$$a^h_{\omega_i}\left(\chi_i(Rv - p_i(Rv)),\ \chi_i(Rv - p_i(Rv))\right) \leq (1 + \lambda_{\max})\, a^h_{\omega_i}(Rv, Rv).$$

(ii) The projection is $a^h_{\widetilde{\omega}_i}$-orthogonal, i.e. for all $v \in H(\omega^*)$ it holds that

$$a^h_{\widetilde{\omega}_i}(Rv - p_i(Rv),\ p_i(Rv)) = 0.$$

Proof. We use the projection p_i from Theorem 4.15. Statement (ii) follows directly from that lemma, since $a^h_{\widetilde{\omega}_i} = a_i$. Note that the partition of unity χ_i vanishes on all outer cells of ω_i, so it can be the identity on at most $H(\widetilde{\omega}_i)$. In particular, we have $U_i = \omega_i \setminus \omega_i^{\text{ovlp}} \subset \widetilde{\omega}_i$. Thus, we get:

$$\begin{aligned} a^h_{\omega_i}\left(\chi_i(v - p_i v),\ \chi_i(v - p_i v)\right) &= \left(a^h_{U_i} + a^h_{\omega_i^{\text{ovlp}}}\right)\left(\chi_i(v - p_i v),\ \chi_i(v - p_i v)\right) \\ &= a^h_{U_i}(v - p_i v,\ v - p_i v) + b_i(v - p_i v,\ v - p_i v) \\ &\leq (1 + \lambda_{\max})\, a^h_{\widetilde{\omega}_i}(v - p_i v,\ v - p_i v), \end{aligned}$$

where we used Proposition 3.15 for the first term and Theorem 4.15 for the second. By the $a^h_{\widetilde{\omega}_i}$-orthogonality of p_i, we further have

$$a^h_{\widetilde{\omega}_i}(v - p_i v,\ v - p_i v) \leq a^h_{\widetilde{\omega}_i}(v,\ v),$$

and by locality of a^h,

$$a^h_{\widetilde{\omega}_i}(v,\ v) \leq a^h_{\omega_i}(v,\ v).$$

Combining these, gives the desired result. □

At first glance, the argument might seem too simple. It mainly relies on is the inequality $a^h_{\widetilde{\omega}i}(v, v) \leq a\omega_i(v, v)$. This can be achieved for more easily, however.

Example 4.44 (Trivial Full GenEO)**.** Consider spectral elements with the notation from Definition 4.13, where $\widetilde{V}_i = H(\omega_i)$ and

$$\begin{aligned} a_i(v, w) &:= 0, \\ b_i(v, w) &:= a^h_{\omega_i}(\chi_i v, \chi_i w). \end{aligned}$$

Then the resulting spectral elements exhibit the same approximation properties as those in Full GenEO. However, observe that the local approximation spaces S_i must contain the full quotient $\ker(a_i)/\ker(b_i)$ for any threshold $\lambda_{\max} < \infty$.

This example highlights a key issue: we do not provide any bound on the size of the coarse space required to achieve a target threshold $\lambda_{\max}$. This omission is not uncommon in the literature on domain decomposition methods. Nonetheless, since $\chi_i \ker(a_i)$ has to be included in the global approximation space, on should at least be able to show that it has small, finite dimension, independent of the mesh parameter h. This property does hold for the DG variant of GenEO, as we will establish in the following.

Proposition 4.45. *For DG Variant 1 of Full GenEO and GenEO, the following holds.*

(i) The kernel of a_i consists of all functions in $H(\omega_i)$ that are constant on the interior cells

$$\mathcal{T}(\widetilde{\omega}_i) = \mathcal{T}(\omega_i) \setminus \partial\omega_i,$$

and arbitrary on the boundary $\partial\omega_i$.

(ii) If χ_i factors through $\widetilde{\omega}_i$, then $\chi_i(\ker(a_i))$ has dimension at most 1.

(iii) If χ_i factors through $\widetilde{\omega}_i$, then the kernel of b_i contains at least all functions that vanish on $\mathcal{T}(\widetilde{\omega}_i)$ and are arbitrary on $\partial\omega_i$.

Proof. Recall that $a_i(v, w) = a^h_{\widetilde{\omega}_i}(Rv, Rw)$, where $R = R_{\omega_i,\widetilde{\omega}_i}$. As discussed in the proof of Proposition 3.55, the kernel of $a^h_{\widetilde{\omega}_i}$ on $H(\widetilde{\omega}_i)$ contains at most the constant functions.

Since a_i is defined on $H(\omega_i)$ but restricted via R, functions in $\ker(a_i)$ can behave arbitrarily on the outer cells $\partial\omega_i$. Therefore, $\ker(a_i)$ consists of functions that are constant on $\mathcal{T}(\widetilde{\omega}_i)$ and arbitrary elsewhere.

For (ii), if χ_i factors through $\widetilde{\omega}_i$, then it vanishes on $\partial\omega_i$. Thus applying it to $\ker(a_i)$ removes all degrees of freedom except the constant mode on the interior, yielding a space of dimension at most one.

For (iii), under the assumption that χ_i factors through $\widetilde{\omega}_i$, we have $\ker(R) \subset \ker(\chi_i) \subset \ker(b_i)$, which implies that any function vanishing on $\mathcal{T}(\widetilde{\omega}_i)$ lies in $\ker(b_i)$. □

The assumption that χ_i factors through $\widetilde{\omega}_i$ is typically satisfied in practice. To see this, note that the image of the partition of unity operator χ_i lies in $H_0(\omega_i)$, and elements thereof are required to be zero on all outer cells. Thus, if χ_i is based on point-wise multiplication with a function on $|\mathcal{T}(\omega_i)|$, this function has support in $|\mathcal{T}(\widetilde{\omega}_i)|$.

What remains to be addressed is the fact that $\ker(a_i + b_i)$ is not trivial for the DG variant 1. Thus, for once, we are in a situation where we must explicitly construct a basis for the quotient space $\widetilde{V}_i / \ker(a_i + b_i)$, along with a corresponding representation of the bilinear forms $\bar{a}_i$ and $\bar{b}_i$. Fortunately, with Proposition 4.45 at hand, this task becomes relatively straightforward. Essentially, we can simply work in $H(\widetilde{\omega}_i)$ as the following proposition states.

Proposition 4.46. *For DG Variant 1 of GenEO from Example 4.42, let $i \in I$ and consider the natural inclusion*

$$\iota : H(\widetilde{\omega}_i) \hookrightarrow H_0(\omega_i),$$

which extends functions in $H(\widetilde{\omega}_i)$ with zero. Further, define the bilinear form

$$\widetilde{b}_i(\cdot, \cdot) := b_i(\iota\cdot, \iota\cdot)$$

on $H(\widetilde{\omega}_i)$. Then ι induces an isomorphism

$$\bar{\iota} : H(\widetilde{\omega}_i) / \ker(a^h_{\widetilde{\omega}_i} + \widetilde{b}_i) \to H(\omega_i) / \ker(a^h_{\widetilde{\omega}_i} + b_i).$$

Proof. We first observe that

$$v \in \ker(a^h_{\widetilde{\omega}_i} + \widetilde{b}_i) \quad \iff \quad \iota v \in \ker(a^h_{\widetilde{\omega}_i} + b_i)$$

since $\widetilde{b}_i(v,w) = b_i(\iota v, \iota w)$ by definition. This equivalence implies that the quotient map $\bar{\iota}$ is well-defined and injective.

To prove surjectivity, we use the description of the kernel in Proposition 4.45, which shows that $\ker(a^h_{\widetilde{\omega}_i} + b_i)$ contains all functions that are supported on the boundary $\partial\omega_i$. It follows that

$$\text{Image}(\iota) + \ker(a^h_{\widetilde{\omega}_i} + b_i) = H(\omega_i).$$

This sum remains exhaustive after passing to the quotient and thus

$$H(\omega_i)/\ker(a^h_{\widetilde{\omega}_i} + b_i) = \Big(\text{Image}(\iota) + \ker(a^h_{\widetilde{\omega}_i} + b_i)\Big)/\ker(a^h_{\widetilde{\omega}_i} + b_i) = \text{Image}(\bar{\iota}).$$

Hence, $\bar{\iota}$ is also surjective, and therefore an isomorphism. □

Remark 4.47. Thanks to Proposition 4.46, we have shown that the eigenproblems in DG Variant 1 of GenEO can be solved over the smaller space

$$H(\widetilde{\omega}_i)/\ker(a^h_{\widetilde{\omega}_i} + \widetilde{b}_i).$$

Moreover, on $H(\widetilde{\omega}_i)$ we typically have $\ker(a^h_{\widetilde{\omega}_i} + \widetilde{b}_i) = 0$, by a similar argument as in Example 4.39. In this case, we can use extension with zero for the splitting maps g_i.

Attempting to define a similar variant for MS-GFEM encounters a challenge: the space $\widetilde{V}_i = H(\omega_i)^{\perp_a}$ generally requires contributions from the outer cells for its computation. In practice, this orthogonality constraint is built into the eigenproblem itself. As a result, removing the outer cells from the formulation is less appealing, as they will appear in the calculation either way. Moreover, the main motivation for the current effort is to avoid being forced into at least one layer of oversampling — but oversampling is the preferred mode of operation for MS-GFEM.

So instead we consider another GenEO variant — namely, the one used [10], the implementation of which we will review in Chapter 5. We again modify the definition of a_i in GenEO, this time replacing it with the bilinear form $\mathring{a}^h$ defined on H by setting

$$\mathring{a}^h_\omega(v,w) := a^h_{\widetilde{\omega}}(v,w) + \sum_{F\in O_\omega} a_F(v,w), \tag{4.12}$$

where O_ω is the set of all faces $F \in \dot{\mathcal{F}}$ that do not belong to $\widetilde{\omega}$, but whose associated cell $\tau(F)$ lies on the boundary $\partial\widetilde{\omega}$ (i.e. it is an outer cell with respect to $\widetilde{\omega}$).

In other words, we "round up" the fractional volume integrals in $a^h_{\widetilde{\omega}}$ to full cell integrals. To see this more concretely, recall that for each face F,

$$a_F(v,w) = \frac{1}{n_\partial(F)}\left(\sqrt{A}\nabla v,\ \sqrt{A}\nabla w\right)_{L^2(\tau(F))},$$

and that for $F \in O_\omega$, these terms do not appear in $a^h_{\widetilde{\omega}}$. Including them in $\mathring{a}^h_\omega$ therefore completes the volume contributions for the outer cells of $\widetilde{\omega}$. For example, in terms of Section 4.2.5, computing $\mathring{a}^h_\omega$ amounts to adding the volume term associated with face F_3 to $a^h_{\widetilde{\omega}}$ — but not the other terms from $a^h_{F_3}$.

Clearly, the definition of $\mathring{a}^h$ does not fit neatly into our presheaf framework over $\mathcal{F}$. This is not a defect of the presheaf setting, but rather an unavoidable artifact of using this alternative bilinear form. As a result, this discrepancy introduces an additional constant when comparing $\mathring{a}^h$ with a^h, as quantified in the following proposition.

Proposition 4.48. *Let $\omega \subset \mathcal{F}$ and use the notation introduced above. Then for all $v \in H(\omega)$,*

$$\mathring{a}^h_\omega(v,v) \leq \max\left(1, \frac{1+\alpha}{\alpha - n_\partial c^2},\right) a^h_\omega(v,v),$$

where c is the constant from the discrete trace inequality, and n_∂ is the maximum number of faces of any cell in $\mathcal{T}$.

Proof. We decompose the full DG bilinear form as

$$a^h_\omega = a^h_{\widetilde{\omega}} + \sum_{F \in O_\omega} a^h_F + \sum_{F \in Q_\omega} a^h_F,$$

where

$$Q_\omega := \left\{ F \in P^{-1}(\omega) \;\middle|\; \tau(F) \in \partial\omega \right\}$$

is the set of all faces in $P^{-1}(\omega)$ that belong to outer cells of ω.

By Proposition 3.48 and the non-negativity of the penalty term s_F, we have

$$a_F(v,v) \leq a_F(v,v) + s_F(v,v) \leq \frac{1+\alpha}{\alpha - n_\partial c^2} \cdot a^h_F(v,v),$$

for each $F \in O_\omega$. Therefore,

$$\begin{aligned}
\mathring{a}^h_\omega(v,v) &= a^h_{\widetilde{\omega}}(v,v) + \sum_{F \in O_\omega} a_F(v,v) \\
&\leq a^h_{\widetilde{\omega}}(v,v) + \frac{1+\alpha}{\alpha - n_\partial c^2} \sum_{F \in O_\omega} a^h_F(v,v) + \sum_{F \in Q_\omega} a^h_F(v,v) \\
&\leq \max\left(1, \frac{1+\alpha}{\alpha - n_\partial c^2}\right) \cdot a^h_\omega(v,v).
\end{aligned}$$

□

Example 4.49 (GenEO, DG Neumann Variant)**.** Let (H, H_0, a^h) be the DG presheaf, and use the notation from Definition 4.13. Consider spectral elements with respect to $\widetilde{V}_i = H(\omega_i)$ and

$$\begin{aligned}
a_i(v,w) &:= \mathring{a}^h_{\omega_i}(v,w), \\
b_i(v,w) &:= a^h_{\omega_i^{\text{ovlp}}}(\chi_i v, \chi_i w),
\end{aligned}$$

where ω_i^{ovlp} is defined as in Example 4.29, and $\mathring{a}^h$ is the modified bilinear form described in (4.12). We refer to this as the DG Neumann variant for GenEO, which is also the variant used in the implementation discussed in Chapter 5.

This variant offers the same type of approximation result as for Full GenEO, albeit with a weaker constant and a similarly relaxed orthogonality statement as in Example 4.42.

Lemma 4.50 (Local approximation with GenEO, DG Neumann Variant)**.** *Consider the GenEO Neumann variant from Example 4.49, with the usual notation from Definition 4.13. Then, for all $i \in I$, there exists a map $p_i : H(\omega_i) \to S_i$ such that for any $\omega^* \supset \omega_i$, and with $R = R_{\omega^*,\omega_i}$, the following holds.*

(i) For all $v \in H(\omega^*)$, we have

$$a^h_{\omega_i}\big(\chi_i(Rv - p_i(Rv)),\, \chi_i(Rv - p_i(Rv))\big) \leq \max\left(1,\, \frac{1+\alpha}{\alpha - n_\partial c^2}\right)(1+\lambda_{\max})\, a^h_{\omega_i}(Rv, Rv).$$

(ii) The projection is $a^h_{\widetilde{\omega}_i}$-orthogonal, i.e. for all $v \in H(\omega^*)$,

$$\mathring{a}^h_{\omega_i}(Rv - p_i(Rv),\, p_i(Rv)) = 0.$$

Proof. The proof follows exactly as in Lemma 4.43, with $a^h_{\widetilde{\omega}_i}$ replaced by $\mathring{a}^h_{\omega_i}$. The only difference occurs in the final estimate, where we use the inequality

$$\mathring{a}^h_{\omega_i}(v, v) \leq \max\left(\frac{1+\alpha}{\alpha - n_\partial c^2},\, 1\right) a^h_{\omega_i}(v, v),$$

as established earlier. □

Remark 4.51. For the DG Neumann variant, the image $\chi_i(\ker a_i)$ again has dimension at most 1, by the same argument as for DG variant 1. Also, we may work over the quotient space $H(\widetilde{\omega_i})$ using the same reasoning as in Proposition 4.46 and Remark 4.47.

Remark 4.52. It is straightforward to define an analogous Neumann variants for the Full GenEO eigenproblems and the same constant of $C = \max\left(1, \frac{1+\alpha}{\alpha - n_\partial c^2}\right)$ will appear in the corresponding local approximation result. However, we omit further repetition of essentially identical steps. Moreover, in later chapters we will not explicitly restate every result for the Neumann variants. Instead, note that for these cases any bounds relying on Theorem 4.15 will still hold true, but appearances of $\lambda_{\max}$ (Full GenEO) or $(1 + \lambda_{\max})$ (GenEO) will gather C as an additional factor.

To summarize, our formulation of DG functions as a sheaf over the set of faces enables the application of previously discussed spectral methods to the WSIP discretization without modification. Moreover, the framework is flexible enough to accommodate variations in local boundary conditions, as demonstrated by the GenEO variants presented here.

4.3 Multilevel Spectral Elements

With the framework in place, we are now ready to define multilevel spectral elements – that is, multilevel generalized elements that satisfy the requirements of spectral elements at each level of the hierarchy. While the construction is conceptually straightforward, we explicitly state the definition and discuss the corresponding local approximation result below to establish consistent notation for what follows.

In the subsequent chapters, we will analyse how the established local approximation results can be assembled to obtain a bound on the condition number of the multilevel additive Schwarz operator (Chapter 5) and on the convergence rate of the multilevel restricted hybrid Schwarz method (Chapter 6).

Definition 4.53 (Multilevel spectral elements)**.** Let (H, H_0, a) be a suitable presheaf and let $\{(\mathcal{H}_l, \mathcal{H}_{l,0}, a)_l\}_l$ be multilevel generalized elements for (H, H_0, a) with respect to a domain decomposition $\{\omega_{l,i}\}_{(l,i)\in I}$, an initial partition of unity $\{\chi_{L,i}\}_{i\leq M_L}$, an index partition $\{\phi_{l,i}\}_{(l,i)\in I}$, and local approximation spaces $\{S_{l,i}\}_{(l,i)\in I}$. For all $(l,i) \in I$, let $a_{l,i}, b_{l,i}$ be symmetric, positive semi-definite bilinear forms on subspaces $\widetilde{V}_{l,i} \subset H(\omega_{l,i})$. Assume that for each $(l,i) \in I$ one of the two scenarios in Definition 4.13 is satisfied with splitting map $g_{l,i}$ and $\mu_{\max} \in (0,1)$. Then the hierarchy of suitable presheaves $\{(\mathcal{H}_l, \mathcal{H}_{l,0}, a)_l\}_l$ is called *multilevel spectral elements* with respect to the given bilinear forms, cover, index partition, and splitting maps – and with eigenvalues threshold

$$\lambda_{\max} := \frac{\mu_{\max}}{1-\mu_{\max}}.$$

Multilevel spectral elements admit a local projection map with guaranteed approximation and orthogonality properties. This result is a direct consequence of the corresponding two-level theory. We state it here in the multilevel context for clarity and completeness.

Theorem 4.54. *In the setting of Definition 4.53, assume additionally that for $(l,i) \in I$ the span of the eigenvectors of the local eigenproblem is dense in $\tilde{V}_{l,i}$. Then there is a map $p_i : \widetilde{V}_{l,i} \to S_{l,i}$ such that the following holds.*

(i) For all $v \in \widetilde{V}_{l,i}$ we have

$$b_{l,i}(v - p_{l,i}(v), v - p_{l,i}(v)) \leq \lambda_{\max} a_{l,i}(v,v).$$

(ii) The map is also $a_{l,i}$-orthogonal and $b_{l,i}$-orthogonal, i.e. for all $v \in \widetilde{V}_{l,i}$ it holds that

$$a_{l,i}(v - p_{l,i}(v), p_{l,i}(v)) = 0, \quad \text{and} \quad b_{l,i}(v - p_{l,i}(v), p_{l,i}(v)) = 0.$$

Proof. This is exactly Theorem 4.15, as this is a local statement for a single level of the multilevel hierarchy. □

Due to the preceding theorem, all previously stated local approximation results that rely on it naturally extend to the multilevel setting – for example, Lemma 4.28, Lemma 4.31, Lemma 4.34, Lemma 4.41, Lemma 4.43, and Lemma 4.50. We refrain from restating each of them individually. It suffices to observe that the corresponding two-level statements remain valid at every level of the hierarchy, with the index i replaced by (l,i).

4.3.1 Applications

We have seen in Chapter 3 how typical PDE settings fit into the framework of suitable presheaves. We then argued that the construction of generalized elements can be iterated to coarser and coarser levels, naturally resulting in a multilevel hierarchy. The previous sections then demonstrated how two-level spectral methods fit seamlessly into this generalized elements framework. What remains is to verify that the assumptions required for two-level spectral elements continue to hold on coarser levels. The following discussion applies to both Full GenEO, GenEO and MS-GFEM elements.

Recall that the preferred assumption for theoretical considerations is that the kernels $\ker(a_{l,i})$ split, and that $a_{l,i}$ is coercive on $\widetilde{V}_{l,i}/\ker(a_{l,i})$. For the methods under discussion, we have $a_{l,i} = a_{\omega_{l,i}}$, so on the finest level, the coercivity condition reduces to the well-posedness of the underlying PDE. Moreover, since each presheaf of generalized elements is embedded into its parent presheaf via an embedding that preserves the bilinear form a, the coercivity property naturally carries over to the next coarser level.

From a practical perspective on the other hand, we have previously been more concerned with whether $\ker(a_{l,i} + b_{l,i}) = 0$, as this allows us to avoid working in quotient spaces. In the multilevel setting, this becomes less relevant. As discussed in Section 3.3.3, there is typically no easily available basis for the spaces $\mathcal{H}_l(\omega_{l,i})$, meaning that we work instead with generating systems. This leads to singular eigenproblems in practice — regardless of whether $\ker(a_{l,i} + b_{l,i}) = 0$ or not.

That said, in many practical cases we still have $\ker(a_{l,i} + b_{l,i}) = 0$, for the same reasons already established in the two-level setting. Recall that we have

$$\ker(a_{l,i} + b_{l,i}) \subset \ker(\chi_{l,i}) \cap \ker(a_{l,i})$$

for the methods under discussion, and this intersection is often trivial. For instance, in the case of the diffusion equation, $\ker(a_{l,i})$ consists of at most the constant functions on any level l, and these are typically not annihilated by $\chi_{l,i}$ except in pathological cases.

In summary, multilevel spectral elements can be constructed for any suitable presheaf, provided we accept that the resulting local eigenproblems may be singular. Section 5.2.2 discusses a practical approach for handling this in implementations. So far, however, we have only presented local approximation results. In the following, we will analyse how these combine into global convergence results – first for the multilevel additive Schwarz method in Chapter 5, and then for the multilevel restricted hybrid Schwarz method in Chapter 6.

Chapter 5

Additive Schwarz for Spectral Elements

In this chapter, we develop and analyse a multilevel additive Schwarz preconditioner within the abstract framework of suitable presheaves, and demonstrate its effectiveness through numerical experiments. The motivation for this work lies in addressing the scalability challenges of classical two-level domain decomposition methods when applied to large-scale problems. In particular, we generalize and extend results from [10], GenEO is cast as a multilevel method, and applied to the heterogeneous diffusion equation as well as a carbon fibre composite problem. While the numerical results remain the same, our formulation provides a broader theoretical foundation and includes additional methods such as MS-GFEM within the same unifying framework. The GenEO multilevel preconditioner has also been studied in [4], in an algebraic setting.

The aim of a multilevel method is to decompose a Hilbert space (V, a) into a hierarchy of closed subspaces

$$V_0 \subset V_1 \subset ... \subset V_L = V,$$

that separate a given problem into different scales. The index l and the space V_l are both called the l-th level, where small l represent coarse levels, and large l fine levels. To parallelize the method, the individual levels are then themselves decomposed into closed subspaces

$$V_l = \sum_{i=1}^{M_l} V_{l,i,0}$$

with $V_{l,i,0} \subset V_l$. With this decomposition, we can define a multilevel version of the additive Schwarz method as follows.

Definition 5.1 (Multilevel additive Schwarz)**.** With the notation from above, let further $\pi_{l,i} : V \to V_{l,i,0}$ denote the a-orthogonal projection onto $V_{l,i,0}$. We call the linear map

$$\mathcal{P} := \sum_{l=0}^{L} \sum_{i=1}^{M_l} \pi_{l,i}$$

parallel subspace correction or the multilevel additive Schwarz operator.

This chapter analyses this method using standard subspace correction theory for multilevel settings, adapted to the context of suitable presheaves. The theoretical results ensure robustness and convergence guarantees for a wide class of discretizations. A central aspect of the approach is the use of the abstract spectral elements from Chapter 4, which define local approximation spaces through eigenvalue problems derived from bilinear forms on subdomains. The main theoretical result in this chapter is a condition number bound for the multilevel additive Schwarz operator, see Theorem 5.17 and Theorem 5.21. This bound depends on the local approximation property of the employed spectral method, the colouring constant ξ, and the number of levels in the hierarchy.

We report on numerical results drawn from [10], as well as on details of the implementation. The results confirm the scalability of the multilevel approach and demonstrate that significant speed-ups are possible when moving from two to three levels—e.g. in an example with approximately 30 million degrees of freedom. As in the previous chapters, we make the following abbreviations.

Notation 5.2. Throughout this section all vectors spaces are silently assumed to be at least separable Banach spaces over $\mathbb{R}$. All maps between vectors spaces are assumed to be continuous and linear. In particular, all presheaves will be assumed to be presheaves of separable Banach spaces with continuous linear maps.

5.1 Robustness

The goal of this section is to establish a robust bound on the condition number of the (non-restricted) additive multilevel Schwarz method (cf. Definition 5.1) within the framework of multilevel generalized elements. To achieve this, we adapt standard subspace correction theory to our abstract setting.

We then apply the local approximation results for Full GenEO, GenEO, and MS-GFEM from the previous chapter to derive explicit condition number bounds for each method. Since the theory is developed entirely within the framework of suitable presheaves, it applies to any discretization compatible with this structure — including, in particular, continuous finite elements and the discontinuous WSIP discretization discussed earlier.

5.1.1 Subspace Correction Theory

Standard subspace correction theory provides a general recipe to prove robustness for two-level — and more generally multilevel — preconditioners, and can be found for instance in [63, 61]. Since typically the theory is phrased for finite-dimensional vector spaces, we revisit the proofs here to adapt to the Hilbert space setting of this thesis.

The goal is to show robustness of $\mathcal{P}$ from Definition 5.1, that is to obtain a bound on the condition number independent of the mesh size h. The strategy will be to establish bounds on the Rayleigh quotient of $\mathcal{P}$, which in turn imply bounds on the condition number in any inner product with respect to which $\mathcal{P}$ is symmetric.

Definition 5.3 (Rayleigh quotient)**.** For a linear map $M : V \to V$ and an inner product c on V, the map

$$R_{M,c} : V \to \mathbb{R}$$

defined by

$$R_{M,c}(v) = \frac{c(Mv, v)}{c(v, v)}$$

is called the Rayleigh quotient of M in c.

Note that for any eigenvector ϕ with eigenvalue λ, we have $R_{M,c}(\phi) = \lambda$. In particular, an upper and lower bound on the Rayleigh quotient readily translates into a bound on the condition number for symmetric positive definite M, as the following proposition makes precise.

Proposition 5.4. *Let (V, c) be a Hilbert space and let $M : V \to V$ be an invertible, (continuous, linear) self-adjoint map with a dense set of eigenvectors and such that*

$$0 < \gamma \leq R_{M,c}(v) \leq \Gamma$$

for all $v \in V$. Then for any other coercive, symmetric bilinear form $d(\cdot, \cdot)$ in which M is also self-adjoint, we have $\kappa_d(M) \leq \Gamma/\gamma$ for the condition number in the d-norm.

Proof. Since d is coercive, the set of eigenvectors of M is also dense in (V, d). By expanding in an orthonormal basis of eigenvectors, it is easy to see that

$$\kappa_d(M) \leq \lambda_{\text{sup}}/\lambda_{\text{inf}},$$

where λ_{sup} is the supremum of eigenvalues, λ_{inf} the infimum. The result follows since the eigenvalues are bounded from above with Γ and from below with γ. □

Since $\mathcal{P}$ is a sum of a-orthogonal projections, it is symmetric with respect to the a-inner product. By the proposition above, it therefore suffices to derive appropriate bounds on its Rayleigh quotient. To that end, we begin by stating a property that will appear repeatedly in the analysis.

Definition 5.5 (Strengthened triangle inequality under the square)**.** Let (V, a) be a Hilbert space and $\{W_n\}_{n\in I}$ a finite family of closed subspaces. We say that this family satisfies the strengthened triangle inequality under the square with constant $\theta > 0$ if for any vectors$\{w_n\}_{n\in I}$ with $w_n \in W_n$ it holds that

$$a\left(\sum_{n\in I} w_n, \sum_{n\in I} w_n\right) \leq \theta \sum_{n\in I} a(w_n, w_n).$$

Lemma 5.6 (Upper bound)**.** *Let (V, a) be a Hilbert space and consider $\mathcal{P}$ from Definition 5.1. Assume that for each individual level $1 \leq l \leq L$, the family $\{V_{l,i,0}\}_i$ satisfies the strengthened triangle inequality under the square with constant θ and assume $M_0 = 1$. Then*

$$a(\mathcal{P}v, v) \leq (1 + \theta L)a(v, v) \qquad \forall v \in V.$$

Proof. On each individual level l we have by the Cauchy-Schwarz inequality

$$a\left(\sum_{i=1}^{M_l} \pi_{l,i}v, v\right) \leq \sqrt{a\left(\sum_{i=1}^{M_l} \pi_{l,i}v, \sum_{i=1}^{M_l} \pi_{l,i}v\right)}\sqrt{a(v, v)}. \tag{5.1}$$

Using the strengthened triangle inequality under the square yields

$$a\left(\sum_{i=1}^{M_l} \pi_{l,i}v, \sum_{i=1}^{M_l} \pi_{l,i}v\right) \leq \theta \sum_{i=1}^{M_l} a\left(\pi_{l,i}v, \pi_{l,i}v\right) = \theta \sum_{i=1}^{M_l} a\left(\pi_{l,i}v, v\right).$$

Plugging this into (5.1), gives

$$a\left(\sum_{i=1}^{M_l} \pi_{l,i}v, v\right) \leq \theta^{1/2}\sqrt{\sum_{i=1}^{M_l} a\left(\pi_{l,i}v, v\right)}\sqrt{a(v, v)},$$

and thus

$$a\left(\sum_{i=1}^{M_l} \pi_{l,i}v, v\right) \leq \theta a(v, v).$$

Summing over all levels completes the proof.

$$\begin{aligned} a(\mathcal{P}v, v) &= a(\pi_{0,1}v, v) + \sum_{l=1}^{L} a(\sum_{i=1}^{M_l} \pi_{l,i}v, v) \\ &\leq (1 + \theta L)a(v, v). \end{aligned}$$

□

With the upper bound established, it remains to prove a lower bound on $a(\mathcal{P}v, v)$. This is where specific properties of the coarse space come into play. Subspace correction theory provides several tools that decompose the analysis into manageable steps. We present these here in the context of an infinite-dimensional Hilbert space.

Definition 5.7 (Stable splitting)**.** Let (V, a) be a Hilbert space and $\{W_n\}_{n\in I}$ a finite family of closed subspaces. We say that this family admits a stable splitting with constant $C_0 > 0$ if for any $v \in V$ there exists a decomposition $v = \sum_{n\in I} w_n$ with $w_n \in W_n$, such that

$$\sum_{n\in I} a(w_n, w_n) \leq C_0\, a(v, v).$$

Lemma 5.8. *Let (V, a) be a Hilbert space and consider $\mathcal{P}$ from Definition 5.1. If the subspaces $\{V_{l,i,0}\}_{l,i}$ admit a stable splitting with constant C_0, then $\mathcal{P}$ satisfies the lower bound*

$$C_0^{-1} a(v, v) \leq a(\mathcal{P}v, v) \qquad \forall v \in V.$$

Proof. For short, we abbreviate

$$\sum_{l,i} := \sum_{l=1}^{L} \sum_{i=1}^{M_l}.$$

Let $v \in V$ and let $v_{l,i} \in V_{l,i,0}$ be a stable splitting of v, i.e. $v = \sum_{l,i} v_{l,i}$ with

$$\sum_{l,i} a(v_{l,i}, v_{l,i}) \leq C_0\, a(v, v).$$

Note that as a-orthogonal projections, the $\pi_{l,i}$ are a-symmetric and act as the identity on their image $V_{l,i,0}$. With this and the Cauchy-Schwarz inequality we get

$$\begin{aligned}
a(v, v) &= a(v, \sum_{l,i} v_{l,i}) = a(v, \sum_{l,i} \pi_{l,i} v_{l,i}) \\
&= \sum_{l,i} a(\pi_{l,i} v, v_{l,i}) \\
&\leq \sum_{l,i} \|\pi_{l,i} v\|_a \|v_{l,i}\|_a \\
&\leq \left(\sum_{l,i} \|\pi_{l,i} v\|_a^2 \right)^{1/2} \left(\sum_{l,i} \|v_{l,i}\|_a^2 \right)^{1/2} \\
&\leq \left(\sum_{l,i} \|\pi_{l,i} v\|_a^2 \right)^{1/2} \left(C_0 a(v, v) \right)^{1/2} \\
&= \left(a(\mathcal{P}v, v) \right)^{1/2} \left(C_0 a(v, v) \right)^{1/2},
\end{aligned}$$

where in the last step we used $\|\pi_{l,i} v\|_a^2 = a(\pi_{l,i} v, v)$. Dividing by $C_0 a(v, v)$ yields the desired result. □

Note that the result above requires a stable splitting across all levels simultaneously — a condition that is often difficult to verify directly. For this reason, we now break it down into more manageable components, following [10].

Definition 5.9 (Level-wise stable splitting)**.** Let $(V, a(\cdot, \cdot))$ be a Hilbert space and consider a multilevel hierarchy

$$V_0 \subset V_1 \subset \cdots \subset V_L = V$$

which decomposes on each level into subspaces

$$V_l = \sum_{i=1}^{M_l} V_{l,i,0},$$

where the V_l and $V_{l,i,0}$ are closed and $M_0 = 1$. We say that this hierarchy of subspaces admits a level-wise stable splitting if there are constants $C, C_1 > 0$ such that for each level l and any $v_l \in V_l$ there exists a decomposition $v_l = v_{l-1} + \sum_{i=1}^{M_l} v_{l,i}$ with $v_{l-1} \in V_{l-1}$ and $v_{l,i} \in V_{l,i,0}$, such that the following two inequalities hold.

$$\sum_{i=1}^{M_l} a(v_{l,i}, v_{l,i}) \leq C_1 a(v_l, v_l),$$
$$a\left(v_{l-1}, v_{l-1}\right) \leq C a(v_l, v_l).$$

Lemma 5.10 (Multilevel stability)**.** *A level-wise stable splitting implies an (overall) stable splitting. More precisely, with the notation from Definition 5.9, for any $v \in V$ there exists a decomposition*

$$v = \sum_{l=0}^{L} \sum_{i=1}^{M_l} v_{l,i},$$

with $v_{l,i} \in V_{l,i,0}$, such that

$$\sum_{l=0}^{L} \sum_{i=1}^{M_l} a(v_{l,i}, v_{l,i}) \leq C^L \left(1 + \frac{C_1}{C-1}\right) a(v, v).$$

That is, the family $\{V_{l,i,0}\}_{l,i}$ admits a stable splitting with constant $C^L \left(1 + \frac{C_1}{C-1}\right)$.

Proof. Let $v \in V = V_L$. Using level-wise stability we can construct a decomposition recursively, level-by-level, starting with $v_L = v$. On level l, there is a decomposition

$$v_l = v_{l-1} + \sum_{i=1}^{M_l} v_{l,i}$$

with $v_{l-1} \in V_{l-1}$ and $v_{l,i} \in V_{l,i}$ satisfying the bounds in Definition 5.9. This process ends at $v_0 \in V_0$ which is not decomposed further and yields a decomposition

$$v = v_0 + \sum_{l=1}^{L} \sum_{i=1}^{M_l} v_{l,i}.$$

By assumption, we have

$$\sum_{i=1}^{M_l} a(v_{l,i}, v_{l,i}) \leq C_1 a(v_l, v_l)$$
$$a(v_l, v_l) \leq C^{L-l} a(v, v),$$

and thus the decomposition satisfies

$$\begin{aligned} a(v_0, v_0) + \sum_{l=1}^{L} \sum_{i=1}^{M_l} a(v_{l,i}, v_{l,i}) &\leq \left(C^L + C_1 \sum_{l=1}^{L} C^{L-l} \right) a(v, v) \\ &= \left(C^L + C_1 \frac{C^L - 1}{C - 1} \right) a(v, v) \\ &\leq C^L \left(1 + \frac{C_1}{C - 1} \right) a(v, v). \end{aligned}$$

□

Summing up, we obtain the following theorem.

Theorem 5.11. *Let (V, a) be a Hilbert space and consider $\mathcal{P}$ from Definition 5.1. Assume that*

(i) For each individual level $1 \leq l \leq L$, the family $\{V_{l,i,0}\}_i$ satisfies the strengthened triangle inequality under the square with constant θ.

(ii) The hierarchy of subspaces $\{V_{l,i,0}\}_{l,i}$ admits a level-wise stable splitting with constants C, C_1.

Then the condition number of the map $\mathcal{P}$ satisfies the bound

$$\kappa_d(\mathcal{P}) \leq C^L \left(1 + \frac{C_1}{C - 1} \right) (1 + \theta L)$$

in any inner product $d(\cdot, \cdot)$ in which $\mathcal{P}$ is symmetric.

Proof. Assumption (i) implies an upper bound on the Rayleigh quotient of $\mathcal{P}$ by Lemma 5.6. Assumption (ii) implies a lower bound by Lemma 5.10 and Lemma 5.8. Together these bounds yield the result by Proposition 5.4. □

5.1.2 Robustness Result

To apply Theorem 5.11 from subspace correction theory, it remains to verify that the strengthened triangle inequality under the square holds, and that a level-wise stable splitting exists. Here, we investigate these properties in the setting of suitable presheaves.

Notation 5.12. For the remainder of this chapter, whenever we consider a hierarchy of suitable presheaves $\{(\mathcal{H}_l, \mathcal{H}_{l,0}, a)\}_{l \leq L}$ over Ω of level L, and a multilevel domain decomposition $\{\omega_{l,i}\}_{(l,i)\in I}$ of Ω, we adopt the following notation.

$$\begin{aligned} V &:= \mathcal{H}_{L,0}(\Omega), \\ V_l &:= \mathcal{H}_l(\Omega), \\ V_{l,i,0} &:= \mathcal{H}_{l,0}(\omega_{l,i}). \end{aligned}$$

Further, for any $\omega^* \supset \omega_{l,i}$, we write $R_{l,i} = R_{\omega^*,\omega_{l,i}}$ for the restriction to $\omega_{l,i}$. Note that since the embeddings $\mathcal{H}_l \subset \mathcal{H}_{l+1}$ are by assumption compatible with restrictions, we do not have to distinguish the $R^{\mathcal{H}_l}_{\omega^*,\omega_{l,i}}$ for different levels l.

In this setting, the strengthened triangle inequality under the square follows from locality of $a(\cdot,\cdot)$ and a finite ξ-colouring, as formalized in the following proposition.

Proposition 5.13. *In the setting of Notation 5.12, let the multilevel domain decomposition allow ξ-colouring. Then for each individual level $l \leq L$, the family $\{V_{l,i,0}\}_{i \leq M_l}$ satisfies the strengthened triangle inequality under the square with constant ξ. More precisely, we have for any collection of vectors $v_i \in V_{l,i,0}$, where $i = 1, \ldots, M_l$, that*

$$a\left(\sum_{i=1}^{M_l} v_i, \sum_{i=1}^{M_l} v_i\right) \leq \xi \sum_{i=1}^{M_l} a(v_i, v_i).$$

Proof. Since $a(\cdot,\cdot)$ is local, it satisfies the strengthened Cauchy-Schwarz inequality by Proposition 3.26. This yields

$$\begin{aligned} a\left(\sum_{i=1}^{M_l} v_i, \sum_{i=1}^{M_l} v_i\right) &= \sum_{i=1}^{M_l}\sum_{j=1}^{M_l} a(v_i, v_j) \\ &\leq \sum_{i=1}^{M_l}\sum_{j=1}^{M_l} \epsilon_{ij}\sqrt{a(v_i, v_i)}\sqrt{a(v_j, v_j)} \end{aligned}$$

with

$$\epsilon_{ij} = \begin{cases} 0 \text{ if } \omega_i \cap \omega_j = \emptyset \\ 1 \text{ otherwise.} \end{cases}$$

Thinking of ϵ as a matrix and considering the vector $\boldsymbol{x} \in \mathbb{R}^{M_l}$ with entries

$$\boldsymbol{x}_i := \sqrt{a(v_i, v_i)},$$

this can be written more compactly and continued as

$$\begin{aligned} a\left(\sum_{i=1}^{M_l} v_i, \sum_{i=1}^{M_l} v_i\right) &\leq \boldsymbol{x}^\top \epsilon \boldsymbol{x} \\ &\leq \|\epsilon\|_2 \boldsymbol{x}^\top \boldsymbol{x} \\ &\leq \sqrt{\|\epsilon\|_1 \|\epsilon\|_\infty} \boldsymbol{x}^\top \boldsymbol{x} \\ &= \xi \sum_{j=1}^{M_l} a(v_i, v_i). \end{aligned}$$

□

Corollary 5.14. *In the setting of Notation 5.12, let the multilevel domain decomposition allow ξ-colouring. Then the additive multilevel domain decomposition operator $\mathcal{P}$ satisfies the bound*

$$a(\mathcal{P}v, v) \leq (1 + \xi L) a(v, v) \text{ for all } v \in V.$$

Proof. Using Proposition 5.13, this follows from Lemma 5.6. □

So the additive multilevel domain decomposition operator $\mathcal{P}$ satisfies assumption (i) of Theorem 5.11. This relies on the bilinear form $a(\cdot, \cdot)$ being local, and on the domain decomposition admitting ξ-colouring. Verifying assumption (ii), on the other hand, will depend on the precise definition of the multilevel hierarchy.

We begin by discussing how local stability implies level-wise stability, and then investigate how the local approximation results of spectral methods guarantee local stability in the presheaf setting.

Definition 5.15 (Local stability)**.** The hierarchy of suitable subsheaves $\{(\mathcal{H}_l, \mathcal{H}_{l,0}, a)\}_{l \leq L}$ is said to be *locally stable* with respect to the multilevel domain decomposition $\{\omega_{l,i}\}_{(l,i) \in I}$ and with constant $C_{\text{loc}} > 0$ if the following holds.

For any level $1 \leq l \leq L$ and any $v_l \in V_l = \mathcal{H}_l(\Omega)$, there exists a decomposition

$$v_l = v_{l-1} + \sum_{i=1}^{M_l} v_{l,i},$$

where $v_{l-1} \in V_{l-1}$ and $v_{l,i} \in V_{l,i,0} = \mathcal{H}_{l,0}(\omega_{l,i})$, such that for all $i \leq M_l$,

$$a_{\omega_{l,i}}(v_{l,i}, v_{l,i}) \leq C_{\text{loc}}\, a_{\omega_{l,i}}(R_{l,i} v_l, R_{l,i} v_l),$$

where $R_{l,i} := R_{\Omega, \omega_{l,i}}$ denotes the restriction map $\mathcal{H}_l(\Omega) \to \mathcal{H}_l(\omega_{l,i})$ associated with the presheaf $\mathcal{H}_l$. If the multilevel domain decomposition is clear from the context, we simply say that the hierarchy of suitable presheaves is locally stable with constant $C_{\text{loc}} > 0$.

Lemma 5.16 (From local to level-wise stability)**.** *Let the hierarchy of suitable presheaves from Notation 5.12 be locally stable with constant $C_{\text{loc}} > 0$. Then the hierarchy of subspaces*

$$\{V_l\}_{l \leq L} = \{\mathcal{H}_l(\Omega)\}_{l \leq L}, \quad \{V_{l,i,0}\}_{(l,i) \in I} = \{\mathcal{H}_{l,0}(\omega_{l,i})\}_{(l,i) \in I}$$

admits a level-wise stable splitting (cf. Definition 5.9) with

$$C_1 = C_{\text{loc}}\, \xi \quad \text{and} \quad C = 2\left(1 + \xi^2 C_{\text{loc}}\right).$$

Proof. Let $1 \leq l \leq L$ and $v_l \in V_l$. By local stability, there exists a decomposition

$$v_l = v_{l-1} + \sum_{i=1}^{M_l} v_{l,i}, \quad \text{with } v_{l-1} \in V_{l-1},\ v_{l,i} \in V_{l,i,0},$$

such that

$$\begin{aligned}\sum_{i=1}^{M_l} a_\Omega(v_{l,i}, v_{l,i}) &= \sum_{i=1}^{M_l} a_{\omega_{l,i}}(v_{l,i}, v_{l,i}) \\ &\leq C_{\text{loc}} \sum_{i=1}^{M_l} a_{\omega_{l,i}}(R_{l,i} v_l, R_{l,i} v_l) \\ &\leq C_{\text{loc}}\, \xi\, a(v_l, v_l),\end{aligned}$$

where the last inequality uses the sub-additivity from Proposition 3.17 and the colouring assumption. To bound the coarse component, we apply the strengthened triangle inequality under the square:

$$\begin{aligned}a(v_{l-1}, v_{l-1}) &= a\left(v_l - \sum_{i=1}^{M_l} v_{l,i},\ v_l - \sum_{i=1}^{M_l} v_{l,i}\right) \\ &\leq 2a(v_l, v_l) + 2a\left(\sum_{i=1}^{M_l} v_{l,i}, \sum_{i=1}^{M_l} v_{l,i}\right) \\ &\leq 2a(v_l, v_l) + 2\xi \sum_{i=1}^{M_l} a(v_{l,i}, v_{l,i}) \\ &\leq 2a(v_l, v_l) + 2\xi^2 C_{\text{loc}}\, a(v_l, v_l) \\ &= 2(1 + \xi^2 C_{\text{loc}})\, a(v_l, v_l),\end{aligned}$$

as claimed. □

Theorem 5.17 (Robustness of the multilevel additive schwarz method)**.** *In the setting of Notation 5.12, assume that the multilevel domain decomposition admits ξ-colouring and that the hierarchy of suitable presheaves is locally stable with constant C_{loc}. Then, in any inner product d with respect to which the multilevel additive Schwarz operator $\mathcal{P}$ is symmetric, its condition number satisfies the bound*

$$\kappa_d(\mathcal{P}) \leq C^L \left(1 + \frac{\xi C_{\text{loc}}}{C - 1}\right)(1 + \xi L),$$

where $C := 2(1 + \xi^2 C_{\text{loc}})$.

Proof. The strengthened triangle inequality under the square with constant ξ follows from Proposition 5.13. In addition, Lemma 5.16 guarantees a level-wise stable splitting with constants C and $C_1 = \xi C_{\text{loc}}$. These two properties satisfy the assumptions of Theorem 5.11, from which the result follows. □

5.1.3 Examples

We now demonstrate how local stability property, holds for the three main classes of spectral elements discussed in this thesis: Full GenEO, GenEO, and MS-GFEM. In

each case, we derive explicit local stability constants that will feed into the condition number bounds for the global preconditioner.

We begin with Full GenEO. Solving the full and unrestricted eigenproblems leads to the cleanest stability constant.

Proposition 5.18 (Local stability for Full GenEO)**.** *Consider multilevel Full GenEO elements in the notation from Definition 4.2 with threshold $\lambda_{\max}$. Assume that the span of the local eigenvectors is dense in each $\widetilde{V}_{l,i}$ for all $(l,i) \in I$. Then the resulting hierarchy of suitable presheaves is locally stable with constant $\lambda_{\max}$.*

Proof. With the abbreviations from Notation 5.12, let $v_l \in V_l$. By Lemma 4.28, for each $(l,i) \in I$, there exists $s_{l,i} \in S_{l,i}$ such that

$$a_{\omega_{l,i}}\big(\chi_{l,i}(R_{l,i}v_l - s_{l,i}),\, \chi_{l,i}(R_{l,i}v_l - s_{l,i})\big) \leq \lambda_{\max}\, a_{\omega_{l,i}}(R_{l,i}v_l, R_{l,i}v_l).$$

Since $\sum_i \chi_{l,i}s_{l,i} \in V_{l-1}$, we define the decomposition

$$\begin{aligned} v_l &= \sum_i \chi_{l,i}R_{l,i}v_l \\ &= \sum_i \underbrace{\chi_{l,i}(R_{l,i}v_l - s_{l,i})}_{=:v_{l,i}} \quad + \quad \underbrace{\sum_i \chi_{l,i}s_{l,i}}_{=:v_{l-1}}. \end{aligned}$$

To verify local stability, observe that

$$a_{\omega_{l,i}}(v_{l,i}, v_{l,i}) = a_{\omega_{l,i}}\big(\chi_{l,i}(R_{l,i}v_l - s_{l,i}),\, \chi_{l,i}(R_{l,i}v_l - s_{l,i})\big)$$

which is bounded by $\lambda_{\max}\, a_{\omega_{l,i}}(R_{l,i}v_l, R_{l,i}v_l)$ by construction. Therefore, the decomposition $v_l = \sum_i v_{l,i} + v_{l-1}$ satisfies the definition of local stability with constant $\lambda_{\max}$. □

Next, we turn to the GenEO method. While its local approximation is slightly weaker than Full GenEO, the structure of the proof remains similar. The resulting stability constant increases by 1.

Proposition 5.19 (Local stability for GenEO)**.** *Consider multilevel GenEO elements using the notation from Definition 4.2, and with threshold $\lambda_{\max}$. Assume that the span of the local eigenvectors is dense in each $\widetilde{V}_{l,i}$ for all $(l,i) \in I$. Then the resulting hierarchy of suitable presheaves is locally stable with constant $1 + \lambda_{\max}$.*

Proof. The argument follows the same structure as in the Full GenEO case, except that we apply the weaker local approximation property provided by Lemma 4.31, which yields the bound with constant $1 + \lambda_{\max}$ instead of $\lambda_{\max}$. □

Finally, in the case of MS-GFEM, we benefit from stronger eigenvalue decay. However, here this comes at the cost of a more intricate stability constant, depending not only on the eigenvalue threshold but also on the norm of the partition of unity. As a result MS-GFEM is better-suited for the restricted hybrid Schwarz method — a setting in which its strengths are fully realized, as we will explore in detail in the later Chapter 6.

Proposition 5.20 (Local stability for multilevel MS-GFEM)**.** *Consider the MS-GFEM eigenproblems from Example 4.32, applied in the multilevel spectral elements setting with notation as introduced in Definition 4.2, and using an eigenvalue threshold $\lambda_{\max}$. Assume*

that the span of the local eigenvectors is dense in each $\widetilde{V}_{l,i}$ for all $(l,i) \in I$. Then the hierarchy of suitable presheaves is locally stable with constant

$$C_{\text{loc}} := \left(\sqrt{\lambda_{\max}} + \chi_{\max}\right)^2,$$

where

$$\chi_{\max} := \sup_{(l,i)\in I} \|\chi_{l,i}\|_a$$

is the supremum of the operator norms of the partition of unity functions.

Proof. We again use the abbreviations from Notation 5.12. Let $v_l \in V_l$. By Lemma 4.34, for each $(l,i) \in I$, there exists an element $s_{l,i} \in S_{l,i}$ such that

$$\|\chi_{l,i}(R_{l,i}v_l - \pi_{l,i}v_l - s_{l,i})\|^2_{a_{\omega_{l,i}}} \leq \lambda_{\max} \|R_{l,i}v_l\|^2_{a_{\omega_{l,i}}}.$$

Using this, define the decomposition:

$$\begin{aligned} v_{l,i} &:= \chi_{l,i}\left(R_{l,i}v_l - \pi_{l,i}v_l - s_{l,i} + \pi_{l,i}v_l\right), \\ v_{l-1} &:= \sum_{i=1}^{M_l} \chi_{l,i}s_{l,i}. \end{aligned}$$

Then $v_l = \sum_i v_{l,i} + v_{l-1}$ as required. The energy of $v_{l,i}$ subdomain can be bounded by

$$\begin{aligned} \|v_{l,i}\|_{a_{\omega_{l,i}}} &= \|\chi_{l,i}\left(R_{l,i}v_l - \pi_{l,i}v_l - s_{l,i} + \pi_{l,i}v_l\right)\|_{a_{\omega_{l,i}}} \\ &\leq \|\chi_{l,i}(R_{l,i}v_l - \pi_{l,i}v_l - s_{l,i})\|_{a_{\omega_{l,i}}} + \|\chi_{l,i}\pi_{l,i}v_l\|_{a_{\omega_{l,i}}} \\ &\leq \sqrt{\lambda_{\max}}\,\|R_{l,i}v_l\|_{a_{\omega_{l,i}}} + \|\chi_{l,i}\|_a\|\pi_{l,i}v_l\|_{a_{\omega_{l,i}}} \\ &\leq \left(\sqrt{\lambda_{\max}} + \|\chi_{l,i}\|_a\right)\|R_{l,i}v_l\|_{a_{\omega_{l,i}}}, \end{aligned}$$

where the last step uses $\|\pi_{l,i}v_l\|_{a_{\omega_{l,i}}} \leq \|R_{l,i}v_l\|_{a_{\omega_{l,i}}}$, since $\pi_{l,i}$ is an a-orthogonal projection. Squaring both sides yields the desired result

$$\|v_{l,i}\|^2_{a_{\omega_{l,i}}} \leq \left(\sqrt{\lambda_{\max}} + \|\chi_{l,i}\|_a\right)^2 \|R_{l,i}v_l\|^2_{a_{\omega_{l,i}}}.$$

□

With these local stability results in hand, we can now apply the general theory of Theorem 5.17 to derive explicit condition number bounds for each method. These bounds are stated in the following theorem and confirm that all three approaches yield robust multilevel preconditioners under the assumptions established above.

Theorem 5.21 (Robustness of additive multilevel spectral element preconditioners)**.** *Consider multilevel spectral elements with notation as in Definition 4.2, and let $\mathcal{P}$ denote the resulting additive Schwarz operator. Assume that the multilevel domain decomposition admits ξ-colouring.*

(i) For Full GenEO eigenproblems, assume that for all (l,i), the span of the local eigenvectors is dense in $\widetilde{V}_{l,i}$. Then

$$\kappa_2(\mathcal{P}) \leq C^L\left(1 + \frac{\xi\lambda_{\max}}{C-1}\right)(1+\xi L),$$

where $C = 2\left(1 + \xi^2 \lambda_{\max}\right)$.

(ii) For GenEO eigenproblems, assume again that the local eigenvectors are dense in each $\widetilde{V}$. *Then*

$$\kappa_2(\mathcal{P}) \leq C^L \left(1 + \frac{\xi(1+\lambda_{\max})}{C-1}\right)(1+\xi L),$$

where $C = 2\left(1 + \xi^2(1+\lambda_{\max})\right)$.

(iii) For MS-GFEM eigenproblems, assume that the local eigenvectors are dense in each $\widetilde{V}$, *and that the partition of unity satisfies*

$$\chi_{\max} := \sup_{(l,i)\in I} \|\chi_{l,i}\|_a < \infty.$$

Then

$$\kappa_2(\mathcal{P}) \leq C^L \left(1 + \frac{\xi(\sqrt{\lambda_{\max}} + \chi_{\max})^2}{C-1}\right)(1+\xi L),$$

where $C = 2\left(1 + \xi^2(\sqrt{\lambda_{\max}} + \chi_{\max})^2\right)$.

Proof. Each case follows directly from applying Theorem 5.17, which gives a general condition number bound of the form

$$\kappa_d(\mathcal{P}) \leq C^L \left(1 + \frac{\xi C_{\text{loc}}}{C-1}\right)(1+\xi L),$$

where $C := 2(1 + \xi^2 C_{\text{loc}})$, and the local stability constant C_{loc} depends on the chosen spectral method.

- **Full GenEO:** Local stability holds with constant $C_{\text{loc}} = \lambda_{\max}$, as shown in Proposition 5.18.
- **GenEO:** The local stability constant is $C_{\text{loc}} = 1 + \lambda_{\max}$, due to Proposition 5.19.
- **MS-GFEM:** Local stability holds with constant $C_{\text{loc}} = (\sqrt{\lambda_{\max}} + \chi_{\max})^2$, as established in Proposition 5.20.

□

5.2 Implementation

The multilevel spectral preconditioner described in the previous sections has been implemented and tested in a sequential setting within the DUNE software framework[1] [9, 11], as detailed in [10]. This section presents the results of those experiments. The runtime estimates reported below (e.g. in Table 5.4) reflect a parallel setting: for local computations, only the maximum time across subdomains is counted towards the total.

For the discussion in this section, we consider multilevel generalized elements $\{(\mathcal{H}_l, \mathcal{H}_{l,0}, a)\}_{l \leq L}$ with respect to the domain decomposition $\{\omega_{l,i}\}_{(l,i)\in I}$, an initial partition of unity $\{\chi_{L,i}\}_{i \leq M_L}$, index partitions $\{\phi_{l,i}\}_{(l,i)\in I}$ and the local approximation spaces $\{S_{l,i}\}_{(l,i)\in I}$. We write $\{\chi_{l,i}\}_{i\in I}$ for the arising canonical partitions of unity. Moreover, we assume that $(\mathcal{H}_L, \mathcal{H}_{L,0}, a)$ is given by the presheaf of continuous or discontinuous finite elements for the diffusion equation. As in the previous section, we adopt the abbreviations from Notation 5.12.

5.2.1 Patch-Wise Stiffness Matrices

The implementation is entirely algebraic and requires only information from the finest level. This input consists of stiffness matrices assembled over certain non-overlapping subsets of elements — referred to as patch matrices here. Their construction is crucial for enabling the recursive multilevel assembly and is described below. Modifications using other symmetric positive semidefinite splittings (e.g. those requiring only the global stiffness matrix or Neumann subdomain matrices) are also possible; see [2, 4].

Patch matrices are defined with respect to the domain decomposition on the finest level L. For each point $x \in \Omega$, consider the index set $\mathcal{J}(x)$ of subdomains containing it:

$$\begin{aligned} \mathcal{J} : \Omega &\to P(\{1, \dots, M_L\}) \\ x &\mapsto \{i \mid x \in \omega_{L,i}\}, \end{aligned}$$

where we use $P(\cdot)$ to denote the power set. For any subset $\sigma \subset \{1, \dots, M_L\}$, we call $|\sigma| := \mathcal{J}^{-1}(\sigma)$ as patch corresponding to σ. In the following, we identify an index set σ with its associated patch $|\sigma|$.

We assume that the fine mesh $\mathcal{T}_L$ resolves all patches exactly. Each mesh element $\tau \in \mathcal{T}_L$ belongs to exactly one patch. Note that patch boundaries are not included in $|\sigma|$.

For each non-empty patch σ, we assemble its volume patch matrix A_σ. In the case of continuous finite elements, this matrix consists of the volume integrals over mesh elements in $|\sigma|$. For DG methods, we also add face integrals that can be uniquely attributed to $|\sigma|$, i.e. those on global boundaries and in the interior of the patch.

What remains are faces shared between patches. To avoid duplicating these contributions, we introduce a total ordering on the power set $P(\{1, \dots, M_L\})$ and, for each ordered pair (ρ, σ) of patches sharing at least one face, we assemble the face integrals over $|\rho| \cap |\sigma|$ into the *skeleton patch matrix* $A_{(\rho,\sigma)}$.

Remark 5.22. Note that in this context, it would be more intuitive to adopt the separation of face terms developed in Section 3.2.2. Recall that the WSIP discretization of $a(\cdot,\cdot)$ was assembled from contributions over faces in the set $\dot{\mathcal{F}}$, which are defined on the disjoint union of mesh cells. Each of these faces is associated with exactly

[1] www.dune-project.org

one cell, and therefore has a natural correspondence to a single patch. At the time of implementation however, the theoretical framework discussed above had not yet been developed. Additionally, traversing each face only once is computationally more convenient and efficient.

The algorithm receives all non-empty A_σ and $A_{(\rho,\sigma)}$ as input. These patch matrices are then used to compute the required subdomain matrices on all coarser levels recursively.

5.2.2 Eigenproblems

The local approximation spaces $S_{l,i}$ are constructed from the eigenvectors of the local GenEO eigenproblems in the form

$$b_{l,i}(v,\cdot) = \lambda a_{l,i}(v,\cdot).$$

Specifically, all eigenvectors v corresponding to eigenvalues λ greater than a prescribed threshold $\lambda_{\max} < \infty$ are included in $S_{l,i}$.

To solve the generalized eigenproblems, a basis representation of the spaces $\mathcal{H}_l(\omega_{l,i})$ is needed. However, only a generating system is available in practice. To see this, recall that a subdomain $\omega_{l,i}$ on level $l < L$ is made up of subdomains from a finer level, that is,

$$\omega_{l,i} = \bigcup_{j \in I_{l,i}} \omega_{l+1,j}.$$

The space $\mathcal{H}_l(\omega_{l,i})$ is formed from eigenfunctions associated with these finer subdomains, including additional overlap contributions from the index set

$$\overline{I}_{l,i} := \{j \in \{1,\dots,P_{l+1}\} \mid \omega_{l+1,j} \cap \omega_{l,i} \neq \emptyset\}.$$

Thus, for bases $\{x_{l+1,i,n}\}_{n \leq N_{l+1,i}}$ of $S_{l+1,i}$, we obtain

$$\mathcal{H}_l(\omega_{l,i}) = \operatorname{span}(Y_{l,i})$$

with

$$Y_{l,i} := \{\chi_{l+1,j} x_{l+1,j,n} \mid j \in I_{l,i},\, n \leq N_{l+1,j}\} \cup \{R_{l,i}\chi_{l+1,j} x_{l+1,j,k} \mid j \in \overline{I}_{l,i} \setminus I_{l,i},\, n \leq N_{l+1,j}\}.$$

These functions are not guaranteed to be linearly independent — especially the contributions from the overlap $\overline{I}_{l,i} \setminus I_{l,i}$ — so we only obtain a generating system, not a basis.

Let $(l,i) \in I$. We equip the corresponding generating system with a total ordering $Y_{l,i} = \{y_r \mid r = 1,\cdots,|Y_{l-1,i}|\}$, and we assemble the matrices $A_{l,i}, B_{l,i}$ by

$$(A_{l,i})_{r,s} := a_{l,i}(y_r, y_s), \quad \text{and } (B_{l,i})_{r,s} := b_{l,i}(y_r, y_s).$$

To obtain the approximation space $S_{l,i}$, we solve the generalized eigenproblem

$$B_{l,i} x_{l,i,k} = \lambda_{l,i,k} A_{l,i} x_{l,i,k},$$

which will typically have $\ker(A_{l,i} + B_{l,i}) \neq \{0\}$. To avoid working in the quotient by this kernel, we perform an $LDL^\top$ factorization $A_{l,i} = L_{l,i} D_{l,i} L_{l,i}^\top$, then regularize

$$D_{l,i,\epsilon} := D_{l,i} + \epsilon I, \quad \text{for small } \epsilon > 0.$$

We then solve the regularized eigenproblem:

$$L_{l,i}^{-\top} D_{l,i,\epsilon}^{-1} L_{l,i}^{-1} B_{l,i} x_{l,i,k} = \lambda_{l,i,k} x_{l,i,k}.$$

Vectors in $\ker B_{l,i}$ yield eigenvalue $\lambda = 0$; vectors in $\ker A_{l,i} \cap \mathrm{Im}(B_{l,i})$ yield eigenvalues $\sim \epsilon^{-1}$. This approach is numerically stable as long as $B_{l,i}x = 0$ is detected reliably in finite precision. See [3] for a similar discussion including an analysis of this regularization.

5.3 Numerical Results

This section presents numerical experiments from [10], validating the theoretical findings from Theorem 5.21. The implementation makes use of the techniques discussed in the previous section. The multilevel additive Schwarz preconditioner is tested for multilevel GenEO elements within the Krylov subspace projection methods CG and GMRES, as described in Section 1.4. We consider the linear system

$$Ax = b,$$

where A is the stiffness matrix of a PDE discretization. The solver is terminated once

$$\|b - Ax^m\| < 10^{-8} \|b - Ax^0\|,$$

and the number of iterations #IT $= m$ is reported.

5.3.1 Software and Hardware

All experiments were carried out using the DUNE software framework [9, 11]. Arpack [41] (via Arpack++) was used in shift-invert mode to solve the generalized eigenproblems. Cholmod [17] served as the subdomain solver during the iterative solve phase, and UMFPack [20] in the eigensolver. ParMETIS [39] was employed as the graph partitioner. Runtimes, reported in seconds, were measured on an Intel(R) Xeon(R) Silver 4114 CPU @ 2.20GHz.

5.3.2 Islands Problem

We begin with the classical diffusion equation posed on the unit square or unit cube. The coefficient field illustrated in Figure 5.1 features highly heterogeneous islands. In 3D, the coefficient remains independent of the z-coordinate. Dirichlet boundary conditions are imposed on the two faces orthogonal to the x-axis; the rest of the boundary has homogeneous Neumann conditions.

Two-level Experiments

We begin with a series of basic experiments that illustrate the behaviour of two-level spectral domain decomposition (DD) methods. As shown in [32], isolated regions with large diffusion coefficients can give rise to very large eigenvalues that are well separated from the rest of the spectrum.

When constructing the coarse space for spectral elements, one generally has two options: (i) include a fixed number n_{ev} of eigenvectors per subdomain, or (ii) include all eigenvectors with eigenvalue below a specified threshold $\lambda_{\max}$. The former allows control over the size of the coarse space, while the latter controls the condition number

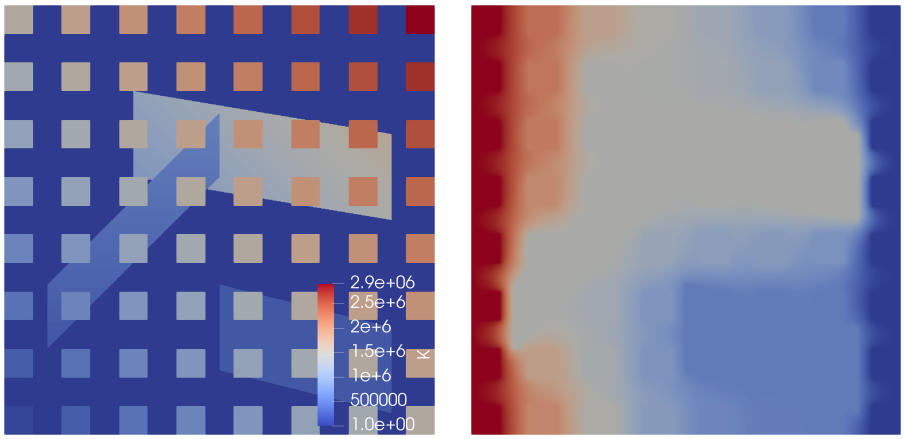


FIGURE 5.1: Permeability field (left) and numerical solution (right) for the 2D islands problem. [10]

TABLE 5.1: Number of iterations #IT and coarse space size n_0 for fixed subdomain count, varying fine mesh size h and overlap δ. CG solver, Q_1 FEM in 2D, 16 subdomains, $\lambda_{\max} = 0.15$. [10]

	Laplace				Islands			
	$\delta \sim h$		$\delta \sim H$		$\delta \sim h$		$\delta \sim H$	
h^{-1}	#IT	n_0	#IT	n_0	#IT	n_0	#IT	n_0
320	30	56	30	56	31	65	31	65
640	29	114	32	50	27	103	30	59
1280	27	236	25	54	27	240	25	78
2560	26	488	26	54	25	481	24	55

of the resulting operator. In most of the experiments presented below, we adopt the threshold-based strategy and select coarse space basis functions according to a fixed eigenvalue threshold $\lambda_{\max}$.

Table 5.1 investigates how the size of the coarse space in the two-level method depends on the overlap δ. For the classical two-level Schwarz method, it is well known that the convergence rate depends on the ratio H/δ [61]. With a fixed number of subdomains (i.e. fixed H) and decreasing mesh size h, the number of iterations remains stable when $\delta \sim H$, but increases when $\delta \sim h$. In contrast, for the additive GenEO preconditioner fixed eigenvalue threshold $\lambda_{\max}$, Table 5.1 shows that the number of iterations remains essentially constant regardless of the overlap size. However, the size of the coarse space n_0 increases significantly when the overlap is small ($\delta \sim h$), whereas it stays relatively stable when $\delta \sim H$. The table also highlights that the transition from a homogeneous diffusion coefficient (Laplace) to a heterogeneous one (Islands) leads to only a modest increase in the coarse space size, further demonstrating the robustness of the method with respect to coefficient variation.

Table 5.2 examines the performance of the two-level method applied to the DG discretization of the Islands problem. We employ the Neumann variant as outlined in Example 4.49. In this experiment, both the number of subdomains and the mesh size are fixed, while the polynomial degree p is varied. As a result, the number of degrees of freedom on the fine level, denoted by n_1, increases accordingly. The table presents results for two different strategies for constructing the coarse space: one based on a fixed eigenvalue threshold $\lambda_{\max} = 0.15$, and the other using a fixed

TABLE 5.2: Iteration numbers #IT and coarse space sizes n_0 for a fixed number of subdomains in the two-level method, when varying polynomial degree p and overlap δ (using Conjugate Gradient, DG finite elements in 2d, 384^2 elements and 256 subdomains). [10]

		$\lambda_{\max} = 0.15$					$n_{ev} = 20$				
		$\delta = 2$		δ var			$\delta = 2$		δ var		
p	n_1	#IT	n_0	δ	#IT	n_0	#IT	n_0	δ	#IT	n_0
1	589824	28	1457	2	28	1457	18	5120	2	18	5120
2	1327104	21	3171	3	22	1901	19	5120	3	18	5120
3	2359296	20	5026	3	21	2991	20	5120	3	19	5120
4	3686400	18	8217	4	21	3322	21	5120	4	20	5120
5	5308416	17	13596	4	21	5078	23	5120	4	21	5120
6	7225344	17	17029	5	22	5234	24	5120	5	22	5120

TABLE 5.3: Conjugate Gradients, $\mathbb{Q}_1$ conforming finite elements, $\delta = 3h$, $\lambda_{\max} = 0.3$. [10]

subdomains	64	256	1024	4096	16384
levels	degrees of freedom				
finest total	410881	1640961	6558721	26224641	104878081
2 lvl n_0	306	1348	5523	22673	91055
3 lvl n_0		130	431	1319	3890
4 lvl n_0			207	436	891
levels	iterations #IT				
2	25	26	27	26	26
3		32	31	31	33
4			40	38	38

number of eigenvectors per subdomain $n_{\text{ev}} = 20$. Each setup is tested with both fixed and varying overlap δ. When using a fixed threshold, the number of iterations remains constant — or even decreases slightly — as the polynomial degree increases. However, this comes at the cost of a growing coarse space dimension n_0. In contrast, using a fixed number of eigenvectors leads to a constant coarse space size, but a modest increase in the number of iterations. Overall, the results demonstrate that the spectral preconditioner remains highly effective across all tested configurations, maintaining robust performance as the polynomial degree increases.

Weak Scaling

We now investigate the multilevel method in a weak scaling scenario by increasing the number of subdomains while keeping the number of degrees of freedom per subdomain fixed. Table 5.3 presents results for the two-dimensional Islands problem

TABLE 5.4: Islands problem in 3d. Fixed problem size 320^3 mesh, 32768000 degrees of freedom. On the finest level $n_{ev} = 15$ eigenvectors are taken per subdomain, while for the three level calculation threshold $\lambda_{\max} = 0.4$ is used on the intermediate level. Iteration numbers are for the hybrid form of the preconditioner using multiplicative subspace correction over levels and restricted additive Schwarz in each level used within GMRES (restart not reached). Times are in seconds. [10]

P_L	P_{L-1}	#IT	n_0	T_{seq}	T_{par}	$T_{i,min}$	$T_{i,max}$	T_{coarse}
two level method								
512	1	12	7680	63613	191.3	70.3	176.2	0.47
1024	1	12	15360	35817	58.4	18.2	49.8	1.3
2048	1	14	30720	18781	25.2	4.9	13.2	5.1
4096	1	13	61441	19982	33.5	2.2	7.0	20.1
three level method								
4096	32	15	1387	21168	55.9	9.8	42.3	0.27
4096	64	15	1817	20725	27.7	2.5	15.1	0.18
4096	128	16	2569	20549	18.4	0.59	6.2	0.15

using $\mathbb{Q}_1$ continuous finite elements with fixed overlap $\delta = 3h$ and an eigenvalue threshold $\lambda_{\max} = 0.3$.

From left to right, the number of subdomains grows from 64 to 16,384. The row labelled "finest total" shows the total number of degrees of freedom on the finest level. The following rows report n_0, the size of the level-0 coarse space, when using two, three, or four levels (intermediate level sizes are omitted for brevity). These results clearly demonstrate that introducing additional levels can significantly reduce the size of the coarsest space.

The final rows of the table show the number of iterations required for convergence when using two, three, or four levels. Within each row, we observe that the iteration counts remain nearly constant as the number of subdomains increases, indicating robustness of the method with respect to domain partitioning. Across each column, we note a moderate growth in iteration numbers with increasing levels, much milder than the exponential increase predicted by Theorem 5.21. The data in Table 5.3 instead suggest a scaling behaviour of $\kappa(BA) = O(L^2)$.

Strong Scaling

Table 5.4 presents results for the three-dimensional Islands problem using a cell-centered finite volume discretization with two-point flux approximation. In this experiment, the mesh is fixed while the number of subdomains and the number of levels are varied.

The first block of rows corresponds to the two-level method. Notably, the sequential setup time T_{seq} for building the preconditioner decreases by nearly a factor of three as the number of subdomains increases from 512 to 4096. This improvement arises because both the direct solver and the eigensolver scale non-linearly with the number of degrees of freedom per subdomain and thus smaller subdomains reduce computational cost. However, increasing the number of subdomains also leads to a larger coarse problem, as reflected by the growing coarse space dimension n_0. As

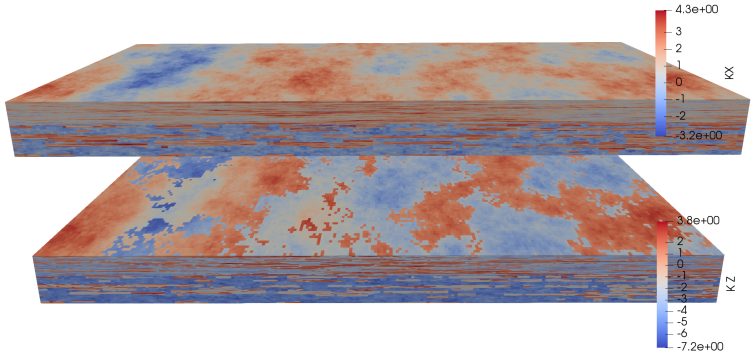


FIGURE 5.2: Permeability field for the SPE10 problem [10].

a result, the estimated parallel runtime reaches a minimum of 25.2 seconds at 2048 subdomains. At 4096 subdomains, the sequential time required for factorizing the coarse problem, denoted T_{coarse}, becomes the dominant cost.

Additionally, we observe a notable spread in the time required to solve the local eigenproblems. The minimum and maximum times across subdomains are shown in columns $T_{i,\min}$ and $T_{i,\max}$, respectively. This variability suggests that distributing more subdomains than available processors may improve efficiency by averaging runtimes over multiple subdomains.

The last three rows of the table report results for a three-level method using 4096 subdomains on the finest level, with varying numbers of subdomains on the intermediate level. Among these, the best performance is achieved when using 128 intermediate subdomains, which yields a lower total parallel runtime than the two-level method.

Note that all reported times in Table 5.4 refer to the preconditioner setup phase. The actual time spent solving the linear system is only about one-twentieth of the setup time.

5.3.3 SPE10

Next, we consider the SPE10 benchmark problem [18], which was originally designed to evaluate multiscale methods but is now frequently used as a test case for preconditioners as well. The underlying diffusion equation is solved on a box-shaped domain discretized with an axis-aligned, equidistant hexahedral mesh consisting of 1,122,000 elements. The diffusion coefficient $A(x)$ is diagonal and highly heterogeneous: the x- and y-components are identical and vary across 7 orders of magnitude, while the z-component varies across 11 orders. Figure 5.2 visualizes the resulting permeability field.

Table 5.5 presents results for this problem, comparing the performance of different discretization schemes: continuous Q_1 finite elements on a refined mesh, cell-centered finite volumes on a refined mesh, and DG-Q_1 on the original mesh. All configurations result in approximately 9 million degrees of freedom. Results are reported for both two- and three-level preconditioners using up to 2048 subdomains.

The preconditioner used is in hybrid form, combining multiplicative subspace correction across levels with restricted additive Schwarz within each level. GMRES is employed as the iterative solver (restart was not reached). For each setup, we report

TABLE 5.5: SPE10 problem. Discretization schemes: $\mathbb{Q}_1$ conforming finite elements (CG), cell-centered finite volumes (FV) and DG-$\mathbb{Q}_1$ are compared for two and three levels. Times are in seconds. [10]

	CG, $n_L = 9124731$			CCFV, $n_L = 8976000$			DG, $n_L = 8976000$		
P_L, P_{L-1}	#IT	n_0	T_{par}	#IT	n_0	T_{par}	#IT	n_0	T_{par}
two levels, $\lambda_{\max} = 0.3$									
256	24	7237	133.8	25	7502	49.2	22	8366	233.5
512	24	9830	60.4	23	10600	21.6	21	13690	124.9
1024	28	21881	22.3	25	25753	12.3	24	31637	42.4
2048	25	29023	15.7	25	35411	11.2	25	46844	36.7
three levels, $\lambda_{\max} = 0.3$									
256, 16	29	1222	151.4	29	1364	70.3	31	1683	273.2
512, 16	27	1228	77.8	28	1446	47.4	28	1762	186.8
1024, 32	36	3145	46.3	34	3487	47.3	33	5476	231.4
2048, 32	31	3120	40.1	35	3421	49.9	36	5359	204.8

the number of iterations, the size of the coarsest space, and the estimated parallel runtime required to set up the preconditioner.

Key observations include:

- The number of iterations is essentially independent of both the number of subdomains and the discretization scheme.
- Introducing a third level leads to a moderate increase in iteration count.
- However, due to the relatively small problem size, using more than two levels does not result in improved total runtime.

5.3.4 Composites Problem

We conclude with results on the simulation of carbon fibre composite materials arising in aerospace engineering, as described in detail in [52, 16]. The setup closely follows the one presented in [52, p. 271], with the exception that the domain is flattened and consists of 9 ply layers and 8 interface (resin) layers.

The equations of linear elasticity are solved in three dimensions using $\mathbb{Q}_2$ serendipity elements, resulting in a total of 10,523,067 degrees of freedom. Table 5.6 reports results for domain decompositions into 1024 subdomains, employing 2, 3, or 4 levels. The hybrid variant of the spectral preconditioner is used, combining multiplicative subspace correction across levels with restricted additive Schwarz within each level. GMRES is used as the iterative solver (no restarts occurred during the experiments).

The results show that the two-level method achieves convergence in just 13 iterations. However, the three-level and four-level methods require 31 and 35 iterations, respectively. While the iteration count increases with the number of levels, the benefit lies in a significantly reduced maximum number of degrees of freedom per coarse-level subdomain.

TABLE 5.6: Carbon fibre composite problem. 9 ply layers (thickness 0.23mm), 8 resin layers (thickness 0.02mm), discretized with 256 × 64 × 52 mesh using $\mathbb{Q}_2$ serendipity elements. Threshold $\lambda_{\max} = 0.35$. [10]

	subdomains				max dofs/subdomain				#IT
level	3	2	1	0	3	2	1	0	
			1024	1			28791	21565	13
		1024	32	1		28791	1260	914	31
	1024	128	16	1	28791	546	515	273	35

5.3.5 Conclusions

The numerical results from [10] presented above verify applicability of the theoretical findings to real-world PDE problems. They in particular investigate robustness with respect to mesh size and number of subdomains. The decay of the convergence rate with increasing number of levels predicted by the theory is found to be moderate. For the largest problem with 30 million degrees of freedom, a speed-up could be achieved by going from two to three levels. Notably the latter two used a restricted hybrid Schwarz scheme which has become popular for spectral methods such as GenEO. We will analyse this in detail in the following chapter.

The numerical results presented above, originally reported in [10], confirm that the theoretical framework developed in this work is applicable to large-scale, real-world PDE problems. In particular, the experiments systematically assess the robustness of the multilevel spectral preconditioners with respect to both mesh refinement and the number of subdomains.

The theoretical prediction of a gradual degradation in convergence rate with increasing levels is supported by the results: although the number of iterations does grow, the increase is moderate and remains acceptable in practical scenarios. For example, in the largest test case involving over 30 million degrees of freedom, a clear speed-up is observed when transitioning from a two-level to a three-level scheme.

Importantly, the best-performing configurations employed the restricted hybrid Schwarz method, which combines multiplicative correction across levels with restricted additive Schwarz on each level. This approach has emerged as particularly effective in conjunction with spectral coarse spaces such as GenEO.

In the following chapter, we will examine this restricted hybrid Schwarz strategy in greater detail.

Chapter 6

Restricted Hybrid Schwarz for Spectral Elements

MS-GFEM was originally phrased as an approximation space within the GFEM framework. However, it also lends itself for use as the coarse space of a two-level Schwarz method. This chapter presents the MS-GFEM preconditioner for the diffusion equation with continuous finite elements as developed in [58]. As a first step, the MS-GFEM approximation map is defined as an automorphism on the finite element space. This can then be applied in a Richardson-type iteration. Writing the resulting method in matrix notation reveals that the MS-GFEM approximation map is exactly one application of the restricted hybrid two-level Schwarz scheme (Definition 1.57).

We analyse the resulting preconditioner, showing that even the simple Richardson-type iteration converges with an arbitrarily fast convergence rate. It then follows that employing this method with GMRES will converge at least as fast. This is notably a far stronger result than what is currently available for GenEO and its variants; however, this does not necessarily justify the larger setup cost of MS-GFEM in practice, as we pick up on in the conclusions of this thesis.

Next, the MS-GFEM preconditioner is generalized to the presheaf framework in Section 6.2. This immediately casts it as a multilevel method, for which we prove robustness — again with arbitrarily fast convergence, although here deteriorating with the number of levels. Phrasing the method in this generality clarifies its relation to GenEO. In principle, the provided bounds also apply to Full GenEO. However, the significance of the results relies on achieving an eigenvalue threshold $\lambda_{\max} < 1$, which is unavailable in this case. Still, since theoretical bounds are often pessimistic, the given proofs provide some insight as to why employing GenEO in a restricted hybrid additive Schwarz scheme leads to better results than in non-restricted versions.

The presented two-level experiments from [58] illustrate that the asymptotic result of arbitrarily fast convergence is achievable with the MS-GFEM preconditioner for typical PDE examples. They also highlight the main drawback of the method — its costly setup. Experiments for a diffusion problem with a skyscraper coefficient test the method across a wide range of parameters, showing that there is a sweet spot between optimal coarse space accuracy and computational efficiency. In particular, it can be cheaper to iterate MS-GFEM a few times rather than reduce the approximation error to very high accuracy, even when setup costs are neglected.

A linear elasticity example of a composite material problem from aerospace engineering then demonstrates the feasibility of the method for real-world applications. The MS-GFEM approximation for this problem was previously studied in [12]. Here, it is compared to the hybrid restricted Schwarz preconditioner with MS-GFEM and GenEO coarse spaces. Both experiments confirm the theoretical results, especially the arbitrarily fast convergence rate for MS-GFEM. However, the extreme setup cost will be difficult to justify in practice, unless the coarse space can be recycled for subsequent solves of the same or similar problems.

Section 6.4 reports on numerical results applying the MS-GFEM preconditioner to the Helmholtz equation. These were presented by the author at ENUMATH 2023 in Lisbon. Similar ideas for the Helmholtz problem have been published in [43, 37, 31]. A key feature in this setting is that the bilinear form associated with the Helmholtz problem is indefinite. As such, it goes beyond the theory presented in this thesis. MS-GFEM approximation results, however, are given for this case in [44]. In essence, everything works as in the positive definite case, except that approximation and convergence results deteriorate with increasing wavenumber. This is not surprising, since the wavenumber essentially measures how far the problem is from being positive definite. Experiments confirm convergence of the iterative method, as expected from the corresponding approximation theory.

6.1 MS-GFEM as a Two-Level Schwarz Method

In this section, we show how MS-GFEM — originally phrased as an approximation method — naturally gives rise to a two-level Schwarz preconditioner. These results have been previously published in [58]. To keep the presentation concrete, we begin with the two-level setting introduced in Section 2.2.3, using the discretized heterogeneous diffusion equation as our model problem. We will investigate a more general formulation in Section 6.2.

Here, we consider the suitable presheaf (H, H_0, a) of continuous finite elements over a Lipschitz domain $\Omega \subset \mathbb{R}^d$ and with the bilinear form $a(\cdot,\cdot)$ from (1.15). Let $\{\omega_i\}_{i\leq M}$ be a cover of Ω allowing ξ^*-colouring, and let $\{\chi_i\}_{i\leq M}$ be a subordinate partition of unity allowing ξ-colouring. As the coarse space, we use the global MS-GFEM approximation space

$$S := \mathcal{H}(\Omega)$$

where $(\mathcal{H}, \mathcal{H}_0, a)$ is the presheaf of MS-GFEM elements for (H, H_0, a) with respect to the given cover, partition of unity, and local approximation spaces S_i, and with threshold $\lambda_{\max}$.

We then interpret the local particular functions $u^p_{h,i}$ from Section 2.2.3 as local solves, and couple these with a coarse correction in the space S. This naturally leads to a restricted hybrid two-level Schwarz method.

6.1.1 Richardson-Type MS-GFEM Iteration

We now describe the most direct way to use MS-GFEM as a preconditioner, leading to a Richardson-type iteration. This produces a sequence $\{u^j\}_{j\in\mathbb{N}} \subset H(\Omega)$ converging to the solution $u \in H(\Omega)$ of the variational problem

$$a(u, v) = \ell(v) \quad \text{for all } v \in H(\Omega).$$

At each step, we compute an MS-GFEM approximation of the residual and use it to update the current iterate.

As before, denote for any subspace $W \subset H(\Omega)$ the corresponding a-orthogonal projection by

$$\pi_W : H(\Omega) \mapsto W.$$

In particular, we denote by π_S the projection onto the coarse space $S = \mathcal{H}(\Omega)$, and by $\pi_i := \pi_{H_0(\omega_i)}$ the projections onto the local fine spaces.

Definition 6.1 (MS-GFEM approximation map)**.** The MS-GFEM approximation map $G : H(\Omega) \mapsto H(\Omega)$ is defined by

$$G(v) := \sum_{i=1}^{M} \chi_i \pi_i(v) + \pi_S\left(v - \sum_{i=1}^{M} \chi_i \pi_i(v)\right).$$

Remark 6.2. Note that $G(v)$ is the MS-GFEM approximation of any $v \in H(\Omega)$ (cf. Section 2.2.3). In particular, it satisfies the bound

$$\|v - G(v)\|_a \leq \Lambda \|v\|_a,$$

where $\Lambda := (\xi\xi^*\lambda_{\max})^{1/2}$ by Theorem 2.8.

Corollary 6.3. *For all $v \in H(\Omega)$, the MS-GFEM map satisfies*

$$\|v - G(v)\|_a \leq \Lambda \|v\|_a.$$

This motivates the following iterative scheme.

Definition 6.4 (Richardson-type iteration)**.** Given an initial guess $u^{(0)} \in H(\Omega)$, define the sequence $\{u^{(j)}\}_{j\in\mathbb{N}}$ recursively by

$$u^{(j+1)} := u^{(j)} + G\left(u - u^{(j)}\right).$$

Each iterate $u^{(j+1)}$ is computed by applying the MS-GFEM approximation map to the current residual. Crucially, this does not require knowledge of u itself, but only of the functional $\ell = a(u, \cdot) \in H(\Omega)'$, which can be evaluated in practice.

Corollary 6.5. *The Richardson-type iteration converges with rate Λ, that is,*

$$\|u^{(j+1)} - u\|_a \leq \Lambda \|u^{(j)} - u\|_a.$$

Proof. Using the definition of the iteration and Corollary 6.3, we get

$$\begin{aligned} \|u - u^{(j+1)}\|_a &= \left\|(u - u^{(j)}) - G(u - u^{(j)})\right\|_a \\ &\leq \Lambda \|u - u^{(j)}\|_a. \end{aligned}$$

□

6.1.2 Matrix Formulation

To clarify the computational steps required for computing the next iterate in the Richardson-type MS-GFEM method, we now express the iteration from Definition 6.4 in matrix form, by fixing a basis for each of the relevant spaces. Matrix representations are understood with respect to these bases. For any function $v \in H(\Omega)$, we write $\boldsymbol{v}$ for its coefficient vector. In particular, we denote the vector representation of the exact solution $u \in H(\Omega)$ by $\boldsymbol{u}$.

Let $\boldsymbol{A}$ be the matrix representation of the bilinear form $a(\cdot, \cdot)$, so that

$$a(v, w) = \boldsymbol{v}^\top \boldsymbol{A} \boldsymbol{w} \quad \text{for all } v, w \in H(\Omega).$$

Let $\boldsymbol{R}_i^\top$ be the matrix representation of the inclusion $H_0(\omega_i) \hookrightarrow H(\Omega)$. Similarly, let $\boldsymbol{R}_S^\top$ represent the inclusion $S \hookrightarrow H(\Omega)$.

Remark 6.6. In the standard Lagrange basis, the matrix

$$\boldsymbol{R}_i := (\boldsymbol{R}_i^\top)^\top$$

corresponds to the representation of the restriction operator R_{Ω,ω_i}. Generally, this can not be expected, however. Restrictions and extensions are defined to be adjoint with respect to the bilinear form $a(\cdot, \cdot)$. They are only by coincidence also adjoint with respect to the Euclidean inner product, unless the given basis is a-orthogonal. The notation in terms of $\boldsymbol{R}$, rather than $\boldsymbol{\iota}$, is thus somewhat unnatural. Since it is engrained into the preconditioner literature, we still adopt it here for consistency.

The local and coarse stiffness matrices, representing a_{ω_i} and $a_\Omega|_{S\times S}$ respectively, are given by

$$\boldsymbol{A}_i := \boldsymbol{R}_i \boldsymbol{A} \boldsymbol{R}_i^\top, \qquad \boldsymbol{A}_S := \boldsymbol{R}_S \boldsymbol{A} \boldsymbol{R}_S^\top.$$

Based on these, the projections π_i and π_S are represented by

$$\boldsymbol{\pi}_i = \boldsymbol{A}_i^{-1} \boldsymbol{R}_i \boldsymbol{A}, \qquad \boldsymbol{\pi}_S = \boldsymbol{A}_S^{-1} \boldsymbol{R}_S \boldsymbol{A}.$$

Let $\boldsymbol{\chi}_i$ denote the matrix representing the partition of unity operator χ_i. Then the matrix representation of the MS-GFEM approximation map G reads

$$\boldsymbol{G} = \sum_{i=1}^{M} \boldsymbol{R}_i^\top \boldsymbol{\chi}_i \boldsymbol{A}_i^{-1} \boldsymbol{R}_i \boldsymbol{A} + \boldsymbol{R}_S^\top \boldsymbol{A}_S^{-1} \boldsymbol{R}_S \boldsymbol{A} \left(\boldsymbol{I} - \sum_{i=1}^{M} \boldsymbol{R}_i^\top \boldsymbol{\chi}_i \boldsymbol{A}_i^{-1} \boldsymbol{R}_i \boldsymbol{A} \right).$$

This can also be written as $\boldsymbol{G} = \boldsymbol{B}\boldsymbol{A}$ with

$$\boldsymbol{B} := \sum_{i=1}^{M} \boldsymbol{R}_i^\top \boldsymbol{\chi}_i \boldsymbol{A}_i^{-1} \boldsymbol{R}_i + \boldsymbol{R}_S^\top \boldsymbol{A}_S^{-1} \boldsymbol{R}_S \left(\boldsymbol{I} - \boldsymbol{A} \sum_{i=1}^{M} \boldsymbol{R}_i^\top \boldsymbol{\chi}_i \boldsymbol{A}_i^{-1} \boldsymbol{R}_i \right). \tag{6.1}$$

This matrix $\boldsymbol{B}$ is referred to as the MS-GFEM preconditioner, and coincides with the restricted hybrid two-level Schwarz scheme from Definition 1.57. The Richardson-type iteration then becomes

$$\boldsymbol{u}^{(j+1)} = \boldsymbol{u}^{(j)} + \boldsymbol{B}\boldsymbol{A}(\boldsymbol{u} - \boldsymbol{u}^{(j)}).$$

In practice, this iteration is implemented via the following algorithm.

Algorithm 1 MS-GFEM as Richardson-type iteration

1: Start with initial vector $\boldsymbol{v} \in \mathbb{R}^n$
2: Compute residual: $\boldsymbol{r} = \boldsymbol{A}(\boldsymbol{u} - \boldsymbol{v})$
3: Compute local correction:

$$\boldsymbol{e}^p = \sum_{i=1}^{M} \boldsymbol{R}_i^\top \boldsymbol{\chi}_i \boldsymbol{A}_i^{-1} \boldsymbol{R}_i \boldsymbol{r}$$

4: Compute coarse correction:

$$\boldsymbol{e}^S = \boldsymbol{R}_S^\top \boldsymbol{A}_S^{-1} \boldsymbol{R}_S (\boldsymbol{r} - \boldsymbol{A}\boldsymbol{e}^p)$$

5: Update: $\boldsymbol{v} \leftarrow \boldsymbol{v} + \boldsymbol{e}^p + \boldsymbol{e}^S$
6: Repeat until convergence

6.1.3 The MS-GFEM Preconditioner

The matrix $\boldsymbol{B}$, introduced in (6.1), defines the MS-GFEM preconditioner. In this section, we investigate the conditioning of the preconditioned system

$$\boldsymbol{B}\boldsymbol{A}\boldsymbol{u} = \boldsymbol{B}\boldsymbol{f}, \tag{6.2}$$

as well as the convergence of a corresponding GMRES iteration. The MS-GFEM theory guarantees that $\boldsymbol{B}\boldsymbol{A}$ is close to the identity with respect to the a-norm, as quantified in the following result.

Proposition 6.7. *It holds that*

$$\|\boldsymbol{I} - \boldsymbol{BA}\|_a \leq \Lambda. \tag{6.3}$$

Moreover, if $\Lambda < 1$, *then*

$$\|\boldsymbol{BA}\|_a \cdot \|(\boldsymbol{BA})^{-1}\|_a \leq \frac{1+\Lambda}{1-\Lambda}. \tag{6.4}$$

Proof. From Corollary 6.3, we know that

$$\|(\boldsymbol{I} - \boldsymbol{BA})\boldsymbol{v}\|_a \leq \Lambda \|\boldsymbol{v}\|_a \quad \text{for all } \boldsymbol{v} \in H(\Omega),$$

which implies $\|\boldsymbol{I} - \boldsymbol{BA}\|_a \leq \Lambda$, and hence also $\|\boldsymbol{BA}\|_a \leq 1 + \Lambda$. Now, if $\Lambda < 1$, then the Neumann series

$$(\boldsymbol{BA})^{-1} = \sum_{i=0}^{\infty} (\boldsymbol{I} - \boldsymbol{BA})^i$$

converges, and we obtain the bound

$$\|(\boldsymbol{BA})^{-1}\|_a \leq \frac{1}{1-\Lambda},$$

which completes the proof. □

While the plain Richardson-type iteration offers arbitrarily fast convergence when ideal coarse spaces are used, it may be preferable in practice to employ a more sophisticated linear solver. Due to the presence of the partition of unity matrices $\boldsymbol{\chi}_i$, the preconditioner $\boldsymbol{B}$ is generally non-symmetric. As a result, symmetric solvers such as Conjugate Gradient are not applicable, and a non-symmetric method like GMRES must be used instead. A comprehensive treatment of GMRES is provided in [53]; here, we summarize only the aspects relevant for the given convergence analysis.

GMRES is defined relative to an inner product $b(\cdot, \cdot)$ on $H(\Omega)$, which may differ from $a(\cdot, \cdot)$. In practice, b is often taken to be the Euclidean inner product in the chosen basis for reasons of computational efficiency.

Notation 6.8. For the remainder of this section, let $b(\cdot, \cdot)$ be an inner product on $H(\Omega)$. Suppose there exist constants $b_1, b_2 > 0$ such that for all $\boldsymbol{v} \in \mathbb{R}^n$,

$$b_1 \|\boldsymbol{v}\|_b \leq \|\boldsymbol{v}\|_a \leq b_2 \|\boldsymbol{v}\|_b.$$

Remark 6.9. Consider a family $\{\tau_h\}_h$ of quasi-uniform meshes with mesh size h, and let $H(\Omega)$ be the corresponding finite element space of piecewise linear functions. Choosing the standard Lagrange basis, define the inner product $b_h(v, w) := \boldsymbol{v}^\top \boldsymbol{w}$. Then, there exist constants $\underline{C}, \overline{C} > 0$ independent of h and the coefficient contrast α, β, such that

$$\underline{C} \beta^{1/2} h^{-d/2+1} \|\boldsymbol{v}\|_{b_h} \leq \|\boldsymbol{v}\|_a \leq \overline{C} \alpha^{1/2} h^{-d/2} \|\boldsymbol{v}\|_{b_h}.$$

Proof. This is a standard result for the energy norm in finite element discretizations [27]. □

Notation 6.10. Let $\boldsymbol{x}_0 \in \mathbb{R}^n$ be an initial guess, and let $\boldsymbol{x}_j$ denote the j-th GMRES iterate for solving $\boldsymbol{BAu} = \boldsymbol{Bf}$ in the b-inner product.

Recall that the GMRES iterate is defined to minimize the residual over the affine Krylov subspace, i.e.

$$\boldsymbol{x}_j = \arg \min_{\boldsymbol{v} \in \boldsymbol{x}_0 + \mathcal{K}_j} \|\boldsymbol{BA}(\boldsymbol{u} - \boldsymbol{v})\|_b,$$

where

$$\mathcal{K}_j := \operatorname{span}\left\{ \boldsymbol{BA}(\boldsymbol{u} - \boldsymbol{x}_0), \dots, (\boldsymbol{BA})^j(\boldsymbol{u} - \boldsymbol{x}_0) \right\}.$$

This leads to the following observation.

Proposition 6.11. *If both, the Richardson-type iterates $\{\boldsymbol{y}_j\}_j$ and the GMRES iterates $\{\boldsymbol{x}_j\}_j$, start from the same initial guess $\boldsymbol{y}_0 = \boldsymbol{x}_0$, then for all $j \in \mathbb{N}$,*

$$\|\boldsymbol{BA}(\boldsymbol{u} - \boldsymbol{x}_j)\|_b \leq \|\boldsymbol{BA}(\boldsymbol{u} - \boldsymbol{y}_j)\|_b.$$

Proof. GMRES minimizes the residual norm over $\boldsymbol{x}_0 + \mathcal{K}_j$. The Richardson-type iterate $\boldsymbol{y}_j$ lies in this space by induction. □

So GMRES converges at least as fast as the Richardson-type iteration. With different inner products $a \neq b$ in play, this is made precise below.

Theorem 6.12. *If $\Lambda < 1$, then for all $j \in \mathbb{N}$,*

$$\|\boldsymbol{BA}(\boldsymbol{u} - \boldsymbol{x}_j)\|_b \leq \Lambda^j \left(\frac{1+\Lambda}{1-\Lambda} \cdot \frac{b_2}{b_1} \right) \|\boldsymbol{BA}(\boldsymbol{u} - \boldsymbol{x}_0)\|_b.$$

Proof. Combining Proposition 6.11 and Corollary 6.5 yields the estimate

$$\begin{aligned}
\|\boldsymbol{BA}(\boldsymbol{u} - \boldsymbol{x}_j)\|_b &\leq \|\boldsymbol{BA}(\boldsymbol{u} - \boldsymbol{y}_j)\|_b \\
&\leq \frac{1}{b_1} \|\boldsymbol{BA}(\boldsymbol{u} - \boldsymbol{y}_j)\|_a \\
&\leq \frac{1}{b_1} \|\boldsymbol{BA}\|_a \cdot \|\boldsymbol{u} - \boldsymbol{y}_j\|_a \\
&\leq \frac{b_2}{b_1} \Lambda^j \|\boldsymbol{BA}\|_a \|(\boldsymbol{BA})^{-1}\|_a \cdot \|\boldsymbol{BA}(\boldsymbol{u} - \boldsymbol{x}_0)\|_b,
\end{aligned}$$

and the result follows by inserting the bound from Proposition 6.7. □

Corollary 6.13. *Under the assumptions of Remark 6.9 and with $\Lambda < 1$, GMRES achieves a residual bound*

$$\|\boldsymbol{BA}(\boldsymbol{u} - \boldsymbol{x}_{j+k})\|_b \leq \Lambda^j \|\boldsymbol{BA}(\boldsymbol{u} - \boldsymbol{x}_0)\|_b$$

after $k = \mathcal{O}(\log(h^{-1}) + \log(\alpha/\beta))$ additional steps, where α, β are the lower and upper bound of the diffusion coefficient.

Proof. By Theorem 6.12, the claimed inequality holds true for any k with

$$\Lambda^k \leq \left(\frac{1+\Lambda}{1-\Lambda} \cdot \frac{b_2}{b_1} \right).$$

The dependence on h and α/β comes in through Remark 6.9. □

Recall from Theorem 4.38 that MS-GFEM can achieve arbitrarily small eigenvalue thresholds $\lambda_{\max} > 0$. Since the approximation constant Λ satisfies $\Lambda \sim \lambda_{\max}$, this yields a preconditioner with arbitrarily fast convergence. Note that the dependency on the coefficient contrast and mesh size only affects the additional k steps. In particular, it does not affect the convergence rate. If $b = a$ is used, this issue disappears altogether.

6.2 A Multilevel Hybrid Schwarz Preconditioner

In this section, we generalize the two-level MS-GFEM preconditioner introduced previously to the abstract framework of suitable presheaves. The result is a multilevel, restricted hybrid Schwarz method applicable to a broad class of elliptic PDEs. We begin by introducing notation that aligns with the sheaf-theoretic setting of Chapter 3.

Notation 6.14. Throughout this section, all vector spaces are assumed to be separable Banach spaces over $\mathbb{R}$, unless otherwise stated. All maps between vector spaces are assumed to be continuous and linear. In particular, all presheaves are presheaves of separable Banach spaces with continuous linear restriction maps.

Notation 6.15. Let (H, H_0, a) be a suitable presheaf over a topological space Ω, and let $\{(\mathcal{H}_l, \mathcal{H}_{l,0}, a)\}_{l\leq L}$ denote the multilevel generalized elements for (H, H_0, a), constructed with respect to a domain decomposition $\{\omega_{l,i}\}_{(l,i)\in I}$, an initial partition of unity $\{\chi_{L,i}\}_{i\leq M_L}$, an index partition $\{\phi_{l,i}\}_{(l,i)\in I}$, and local approximation spaces $\{S_{l,i}\}_{(l,i)\in I}$. Let $\{\chi_{l,i}\}_{(l,i)\in I}$ denote the canonical partition of unity induced by the index partition $\{\phi_{l,i}\}_{(l,i)\in I}$ during the construction of the multilevel elements.

In particular, $\{(\mathcal{H}_l, \mathcal{H}_{l,0}, a)\}_{l\leq L}$ is a hierarchy of suitable presheaves. As previously, we use the abbreviations

$$\begin{aligned} V &:= \mathcal{H}_L(\Omega) = \mathcal{H}_{L,0}(\Omega), \\ V_l &:= \mathcal{H}_{l,0}(\Omega), \\ V_{l,i,0} &:= \mathcal{H}_{l,0}(\omega_{l,i}). \end{aligned}$$

For each $(l,i) \in I$, let $\mathrm{Id}_l : V_l \to V_l$ denote the identity on V_l, and define the a-orthogonal projections

$$\hat{\pi}_{l-1} : V_l \to V_{l-1}, \qquad \hat{\pi}_{l,i} : V_l \to V_{l,i,0}.$$

We also set

$$\hat{\pi}_{-1} = 0, \qquad \hat{\pi}_L = \mathrm{Id}_L$$

for convenience. Note that these projections differ from the projections $\pi_{l,i}$ used in the previous chapter, which were defined with domain V_L, rather than V_l.

As usual, we consider the problem of computing $u \in V$, given the linear form $a(u,\cdot) \in V^*$. Our goal is to construct an approximation map

$$G : V \to V$$

that approximates the identity and is composed of the projections $\hat{\pi}_{l,i}$ and $\hat{\pi}_l$. As seen in the matrix formulation in Section 6.1.2, such a map can be evaluated without knowing u explicitly, relying instead on $a(u,\cdot)$. This follows from the fact that for any a-orthogonal projection $\pi : V \to W$, it holds by definition

$$a(\pi(v),\cdot) = a(v,\cdot) \quad \text{as linear forms on } W.$$

Recall from Proposition 3.24 that each projection $\hat{\pi}_{l,i}$ can be computed using only local information, i.e. within $\mathcal{H}_l(\omega_{l,i})$. By contrast, evaluating $\hat{\pi}_l$ in general requires global information, which we aim to avoid except on the coarsest level. This consideration will guide the construction of the multilevel hybrid Schwarz method in the abstract setting below.

6.2.1 Abstract Two-Level Hybrid Schwarz

Under our assumptions, for each $0 < l \leq L$, the suitable presheaf $(\mathcal{H}_{l-1}, \mathcal{H}_{l-1,0}, a)$ is the presheaf of generalized elements for $(\mathcal{H}_l, \mathcal{H}_{l,0}, a)$ with respect to the cover $\{\omega_{l,i}\}_i$, partition of unity $\{\chi_{l,i}\}_i$, and local approximation spaces $\{S_{l,i}\}_i$. Hence, we may define the same two-level operators as in the previous section:

$$G_l : V_l \to V_l, \qquad G_l := \left(\sum_{i=1}^{M_l} \chi_{l,i} \hat{\pi}_{l,i}\right) + \hat{\pi}_{l-1}\left(\mathrm{Id}_l - \sum_{i=1}^{M_l} \chi_{l,i} \hat{\pi}_{l,i}\right),$$

where Id_l denotes the identity on V_l. In particular, we have

$$G_0 = \hat{\pi}_{0,0} = \mathrm{Id}_0,$$

since we assume $M_0 = 1$ and have previously set $\hat{\pi}_{-1} = 0$ for convenience.

Proposition 6.16. *Suppose that $(\mathcal{H}_{l-1}, \mathcal{H}_{l-1,0}, a)$ additionally satisfies the conditions of multilevel spectral elements for $(\mathcal{H}_l, \mathcal{H}_{l,0}, a)$ with respect to the MS-GFEM or Full GenEO eigenproblems, using threshold $\lambda_{\max}$. Further, assume the domain decomposition allows ξ^*-colouring and that the partition of unity admits ξ-colouring. Then, for all $0 < l \leq L$, it holds that*

$$\|Id_l - G_l\|_{a_\Omega} \leq (\xi\xi^*\lambda_{\max})^{1/2}.$$

Proof. Let $v \in V_l = \mathcal{H}_l(\Omega)$. Using Proposition 5.13 and the a-orthogonality of $\hat{\pi}_{l-1}$, it holds that

$$\begin{aligned}
\|(\mathrm{Id}_l - G_l)v\|_{a_\Omega}^2 &= \left\|\sum_{i=1}^{M_l} \chi_{l,i}(v - \hat{\pi}_{l,i}v) - \hat{\pi}_{l-1}\left(v - \sum_{i=1}^{M_l} \chi_{l,i}\hat{\pi}_{l,i}v\right)\right\|_{a_\Omega}^2 \\
&\leq \xi \sum_{i=1}^{M_l} \|\chi_{l,i}(v - \hat{\pi}_{l,i}v) - \hat{\pi}_{l-1}(v - \hat{\pi}_{l,i}v)\|_{a_\Omega}^2 \\
&= \xi \sum_{i=1}^{M_l} \inf_{s \in V_{l-1}} \|\chi_{l,i}(v - \hat{\pi}_{l,i}v - s)\|_{a_\Omega}^2.
\end{aligned}$$

Since $v - \hat{\pi}_{l,i}v \in \widetilde{V}_{l,i} = H_0(\omega_i)^{\perp_a}$, we can apply Theorem 4.54 to obtain

$$\begin{aligned}
\sum_{i=1}^{M_l} \inf_{s \in V_{l-1}} \|\chi_{l,i}(v - \hat{\pi}_{l,i}v - s)\|_{a_\Omega}^2 &\leq \sum_{i=1}^{M_l} \|\chi_{l,i}((v - \hat{\pi}_{l,i}v) - p_{l,i}(v - \hat{\pi}_{l,i}v))\|_{a_{\omega_{l,i}}}^2 \\
&\leq \lambda_{\max} \sum_{i=1}^{M_l} \|v - \hat{\pi}_{l,i}v\|_{a_{\omega_{l,i}}}^2 \\
&\leq \lambda_{\max} \sum_{i=1}^{M_l} \|v\|_{a_{\omega_{l,i}}}^2 \leq \xi^*\lambda_{\max}\|v\|_{a_\Omega}^2,
\end{aligned}$$

where the final step uses Proposition 3.17. □

6.2.2 Abstract Multilevel Hybrid Schwarz

As discussed above, we aim to avoid global projections $\hat{\pi}_l$ for $l > 0$ — only $\hat{\pi}_0$ on the coarsest level is assumed computable. Hence, among the operators G_l, only G_1 is directly available. For finer levels, we instead define approximate maps $\hat{G}_l$, in which

the global projection $\hat{\pi}_{l-1}$ is concatenated with the recursively defined composition $\hat{G}_{l-1} \circ \hat{\pi}_{l-1}$.

Definition 6.17. Let $\hat{G}_0 := G_0 := \mathrm{Id}_0$. Then, for $0 < l \leq L$, we recursively define the multilevel hybrid maps by

$$\hat{G}_l := \left(\sum_{i=1}^{M_l} \chi_{l,i} \hat{\pi}_{l,i} \right) + \hat{G}_{l-1} \circ \hat{\pi}_{l-1} \left(\mathrm{Id}_l - \sum_{i=1}^{M_l} \chi_{l,i} \hat{\pi}_{l,i} \right).$$

We refer to $\hat{G}_L$ as the hybrid multilevel approximation map associated with the multilevel hierarchy $\{(\mathcal{H}_l, \mathcal{H}_{l,0}, a)\}_{l \leq L}$.

Remark 6.18. For any $v \in V_l$, the value $\hat{G}_l(v)$ can be computed using only local linear solves, provided that the residual $a(v, \cdot) \in V_l^*$ is available. Here, we define a linear system as local if it is posed in one of the local spaces $V_{l,i,0}$, which includes the coarsest space $V_0 = V_{0,1,0}$.

Proof. The case $\hat{G}_0$ is trivial. Assume the claim holds for $\hat{G}_{l-1}$ and that $a(v, \cdot)$ is available for $v \in V_l$. We write

$$\hat{G}_l(v) = \underbrace{\sum_{i=1}^{M_l} \chi_{l,i} \hat{\pi}_{l,i} v}_{=:X} + \hat{G}_{l-1} \hat{\pi}_{l-1}(v - X).$$

The term X is computable using standard local solves and evaluations of the partition functions $\chi_{l,i}$. Given X, the residual

$$a(v - X, \cdot) = a(v, \cdot) - a(X, \cdot) \in V_l^*$$

is available, and thus, so is $a(\hat{\pi}_{l-1}(v - X), \cdot) \in V_{l-1}^*$. Then, $\hat{G}_{l-1}(\hat{\pi}_{l-1}(v - X))$ can be computed from local solves by assumption on $\hat{G}_{l-1}$. □

Theorem 6.19. *Under the assumptions from Notation 6.15, suppose that each presheaf $(\mathcal{H}_{l-1}, \mathcal{H}_{l-1,0}, a)$ satisfies the conditions for multilevel spectral elements for $(\mathcal{H}_l, \mathcal{H}_{l,0}, a)$, using either MS-GFEM or Full GenEO with eigenvalue threshold $\lambda_{\max}$. Assume also that the domain decomposition permits ξ^*-colouring, and the partition of unity functions admit ξ-colouring. Then, for each level $0 < l \leq L$, we have the estimate*

$$\|Id_l - \hat{G}_l\|_a \leq (\xi \xi^* \lambda_{\max})^{1/2} \sum_{j=0}^{l-1} \left(\chi_{\max} (\xi \xi^*)^{1/2} \right)^j =: \Lambda_l,$$

where $\chi_{\max} := \max_{(l,i) \in I} \|\chi_{l,i}\|_a$ and Id_l denotes the identity on V_l.

Proof. We proceed by induction on l. For $l = 0$, we have $\hat{G}_0 = \mathrm{Id}_0$, and thus $\hat{G}_1 = G_1$. Applying Proposition 6.16, we obtain

$$\|\mathrm{Id}_1 - \hat{G}_1\|_{a_\Omega} = \|\mathrm{Id}_1 - G_1\|_{a_\Omega} \leq (\xi \xi^* \lambda_{\max})^{1/2}.$$

Assume the result holds for $\hat{G}_{l-1}$. We expand

$$\begin{aligned}\mathrm{Id}_l - \hat{G}_l = \mathrm{Id}_l - \left(\sum_{i=1}^{M_l} \chi_{l,i}\hat{\pi}_{l,i} + \hat{\pi}_{l-1}\left(\mathrm{Id}_l - \sum_{i=1}^{M_l}\chi_{l,i}\hat{\pi}_{l,i}\right)\right) \\ + (\hat{G}_{l-1} - \mathrm{Id}_{l-1})\hat{\pi}_{l-1}\left(\mathrm{Id}_l - \sum_{i=1}^{M_l}\chi_{l,i}\hat{\pi}_{l,i}\right).\end{aligned}$$

Using the triangle inequality and submultiplicativity of the operator norm, we obtain

$$\|\mathrm{Id}_l - \hat{G}_l\|_{a_\Omega} \leq \|\mathrm{Id}_l - G_l\|_{a_\Omega} + \Lambda_{l-1}\cdot\|\mathrm{Id}_l - \sum_{i=1}^{M_l}\chi_{l,i}\hat{\pi}_{l,i}\|_{a_\Omega}.$$

The first term $\|\mathrm{Id}_l - G_l\|_{a_\Omega}$ can be bounded with Proposition 6.16. The remaining term yields, using Proposition 3.17 and Proposition 5.13,

$$\begin{aligned}\left\|v - \sum_{i=1}^{M_l}\chi_{l,i}\hat{\pi}_{l,i}v\right\|_{a_\Omega}^2 &\leq \xi\sum_{i=1}^{M_l}\|\chi_{l,i}(v-\hat{\pi}_{l,i}v)\|_{a_{\omega_{l,i}}}^2 \\ &\leq \chi_{\max}^2\xi\sum_{i=1}^{M_l}\|v\|_{a_{\omega_{l,i}}}^2 \\ &\leq \chi_{\max}^2\xi^*\xi\|v\|_{a_\Omega}^2.\end{aligned}$$

Combining these estimates gives

$$\|\mathrm{Id}_l - \hat{G}_l\|_{a_\Omega} \leq (\xi\xi^*\lambda_{\max})^{1/2} + \Lambda_{l-1}\chi_{\max}(\xi\xi^*)^{1/2} = \Lambda_l.$$

□

Remark 6.20. In particular, Theorem 6.19 states that

$$\|\mathrm{Id}_L - \hat{G}_L\|_a \leq \Lambda_L. \tag{6.5}$$

So $\hat{G}_L : V_L \to V_L$ is an approximation to the identity, evaluable using only local solves. For $L = 1$, this recovers the two-level construction from earlier.

For the given result to be useful, it is crucial that $\Lambda_L < 1$. To see this explicitly, note that with $\Lambda_L \geq 1$, the inequality (6.5) would also be satisfied by choosing $\hat{G}_L = 0$. Section 4.2.3 provides sufficient conditions under which an arbitrarily small threshold $\lambda_{\max}$ and thus $\Lambda_L < 1$ can be achieved for MS-GFEM. In contrast, (Full) GenEO does not permit arbitrarily small threshold due to eigenvalue accumulation near $\lambda = 1$, see also Figure 4.1.

However, when employed in GMRES, the multilevel restricted hybrid preconditioner can be effective also for (Full) GenEO, as the experiments in Section 5.3.3 and Section 5.3.4 show. The convergence analysis given here can give an intuition of why this is the case, but is not directly applicable to these methods.

6.2.3 Abstract Iterative Methods

If the eigenvalues in a multilevel MS-GFEM setup exhibit exponential decay — as established in Theorem 4.38 — we can achieve $\Lambda_L < 1$, enabling us to apply the same iteration strategy introduced in Section 6.1.

Definition 6.21 (Multilevel Richardson-type iteration)**.** Let $y_0 \in V_L$ be an arbitrary initial guess. Define the sequence $\{y_j\}_{j\in\mathbb{N}}$ recursively by

$$y_{j+1} := y_j + \hat{G}_L \left(u - y_j\right), \tag{6.6}$$

where $u \in V_L$ is the desired solution. We refer to this as the *(multilevel) Richardson-type iteration*.

The key property required for convergence is that $\|\mathrm{Id} - \hat{G}_L\|_a \leq \Lambda_L < 1$, which mirrors the condition in the two-level setting. As a result, the convergence guarantees established earlier remain valid.

Corollary 6.22. *The multilevel Richardson-type iteration satisfies*

$$\|y_{j+1} - u\|_a \leq \Lambda_L \|y_j - u\|_a.$$

In particular, the method converges with rate Λ_L, provided $\Lambda_L < 1$.

We can also define a corresponding GMRES iteration, which always guarantees convergence at least as fast as the Richardson method.

Notation 6.23. Let $b(\cdot,\cdot)$ be a symmetric, coercive bilinear form on V_L, and let $x_0 \in V_L$ be an initial guess. Define the GMRES iterates $\{x_j\}_{j\in\mathbb{N}}$ by

$$x_j := \arg\min_{v\in x_0+\mathcal{K}_j} \|\hat{G}_L(u-v)\|_b,$$

where

$$\mathcal{K}_j := \mathrm{span}\{\hat{G}_L(u-x_0), \ldots, \hat{G}_L^j(u-x_0)\}.$$

We refer to this as (abstract) GMRES performed in the inner product $b(\cdot,\cdot)$.

Theorem 6.24. *The abstract GMRES iteration $\{x_j\}_{j\in\mathbb{N}}$ converges to u with rate at least Λ_L. More precisely, if $\Lambda_L < 1$, then for all $j \in \mathbb{N}$,*

$$\|\hat{G}_L(u-x_j)\|_b \leq \Lambda_L^j \left(\frac{1+\Lambda_L}{1-\Lambda_L} \cdot \frac{b_2}{b_1}\right) \|\hat{G}_L(u-x_0)\|_b,$$

where $b_1, b_2 > 0$ are constants satisfying the norm equivalence

$$b_1\|v\|_b \leq \|v\|_a \leq b_2\|v\|_b \quad \textit{for all } v \in V_L.$$

Proof. The argument follows exactly as in the proof of Theorem 6.12, replacing $\boldsymbol{BA}$ with $\hat{G}_L$, and Λ with Λ_L. □

In summary, we have constructed a multilevel hybrid preconditioner that achieves arbitrarily fast convergence under suitable conditions. The formulation is fully abstract and applicable to a broad class of elliptic PDEs. In the following chapters, we demonstrate the practical utility of its two-level version, including application to the Helmholtz equation — a problem beyond the current theoretical framework.

6.3 Numerical Experiments

This section presents numerical experiments from [58] that validate the theoretical results of the MS-GFEM preconditioner and demonstrate its practical effectiveness. Two examples are considered.

The first is a two-dimensional heterogeneous diffusion problem with high-contrast coefficients, similar to the islands problem described in Section 5.3.2. This example is implemented in MATLAB and is designed to verify the predicted convergence rates, assess computational cost across a broad range of parameters, and compare the MS-GFEM preconditioner with related methods.

The second example is a three-dimensional elasticity problem arising in the simulation of composite aero-structures. It is implemented using the Distributed and Unified Numerics Environment (DUNE) software package [11], with the goal of demonstrating the applicability of MS-GFEM in large-scale, real-world problems on modern high-performance computing systems.

To assess the performance of the methods, we monitor the reduction of the residual norm, as the exact solution is generally not available. Specifically, for any sequence $\{\boldsymbol{u}_j\}_j$ approximating the solution $\boldsymbol{u}$ of a linear system $\boldsymbol{A}\boldsymbol{u} = \boldsymbol{f}$, we define the residual reduction at step j by

$$\text{residual reduction}(j) := \frac{\|\boldsymbol{A}\boldsymbol{u}_j - \boldsymbol{f}\|_2}{\|\boldsymbol{A}\boldsymbol{u}_0 - \boldsymbol{f}\|_2}. \tag{6.7}$$

As stopping criteria a residual reduction of 10^{-10} is used for the diffusion example and of 10^{-6} for the linear elasticity experiment. The average convergence rate per iteration is computed as

$$\text{convergence rate} := (\text{residual reduction}(J))^{1/J},$$

where J is the number of iterations until the stopping criterion is reached, minus 1. We discard the last iterate for this calculation since there the residual reduction can be affected by machine precision for very fast convergence rates. This can be observed in Figure 6.5 for the purple line, representing a convergence rate of around 10^{-9}.

6.3.1 Numerical Solution of Local Eigenproblems

The dominant computational cost of the MS-GFEM method typically lies in solving the local restricted eigenproblems. For the following discussion we use the notation from Section 6.1 and assume that

$$\ker(a_{\omega_i}(\cdot,\cdot) + a_{\omega_i}(\chi_i\cdot, \chi_i\cdot)) = 0,$$

which applies to all experiments presented in this chapter. Moreover, we fix $i \in I$, and we write $\omega = \omega_i$, $\chi = \chi_i$ for short.

To facilitate the computation, the a-harmonic constraint in $H(\omega)^{\perp_a}$ is relaxed using a Lagrange multiplier. This allows the reformulation of the MS-GFEM eigenproblem as a saddle-point system. Specifically, we seek $\lambda \in \mathbb{R} \cup \{+\infty\}$, $\psi \in H(\omega)$, and $p \in H(\omega)$ such that

$$\begin{aligned} \lambda\,(a_\omega(\psi, v) + a_\omega(v, p)) &= a_\omega(\chi(\psi), \chi(v)) \quad && \forall\, v \in H(\omega), \\ a_\omega(\psi, w) &= 0 && \forall\, w \in H_0(\omega). \end{aligned} \tag{6.8}$$

This eigenproblem is equivalent to the original problem (2.14).

Let $\{\psi_1, \ldots, \psi_{n_1+n_2}\}$ give a basis for $H(\omega)$ such that $\{\psi_1, \ldots, \psi_{n_1}\}$ span $H_0(\omega)$. Then the saddle-point formulation (6.8) leads to the following matrix eigenvalue problem: find $\lambda \in \mathbb{R} \cup \{+\infty\}$, $\psi = (\psi_1, \psi_2) \in \mathbb{R}^{n_1+n_2}$, and $\boldsymbol{p} \in \mathbb{R}^{n_1}$ such that

$$\begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \mathbf{0} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \boldsymbol{p} \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{11} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{21} \\ \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \boldsymbol{p} \end{bmatrix}, \tag{6.9}$$

where, with index sets $I_1 = \{1, \ldots, n_1\}$ and $I_2 = \{n_1 + 1, \ldots, n_1 + n_2\}$:

$$\mathbf{A}_{ij} = (a_\omega(\psi_k, \psi_l))_{k \in I_i, l \in I_j},$$
$$\mathbf{P}_{ij} = (a_\omega(\chi_i \psi_k, \chi_i \psi_l))_{k \in I_i, l \in I_j}.$$

This mixed formulation preserves all eigenvectors associated with finite eigenvalues from the original formulation (2.14). It is important, to ensure that the chosen eigensolver is capable of handling the presence of eigenvalues $\lambda = \infty$, which correspond to vectors in the kernel of the right-hand side.

As shown in [45], the special structure of the system matrix allows for efficient block elimination to extract eigenvalues and eigenvectors. This strategy is generalized in [42] for broader classes of elliptic problems. In the MATLAB implementation discussed below, the block elimination technique significantly reduces the computational cost compared to solving the full eigenproblem directly. The computed eigenpairs were accurate to approximately 10^{-5}, which is sufficient for many applications. However, since the objective of this experiment is to verify the theoretical convergence results precisely, we solve the eigenproblem (6.9) directly to avoid introducing additional numerical approximation. An exception is made in Figure 6.4, where we report setup times that take advantage of the block elimination strategy as described in [45].

6.3.2 Example 1: Heterogeneous Diffusion Equation

This example solves the diffusion equation with heterogeneous coefficient (cf. Figure 6.1) on the unit square Ω with the mixed boundary conditions

$$\begin{aligned} A\nabla u \cdot \mathbf{n} &= 1 \quad \text{on } (0,1) \times \{1\}, \quad u = 10 \quad \text{on } \{0\} \times [0,1], \\ A\nabla u \cdot \mathbf{n} &= -1 \quad \text{on } (0,1) \times \{0\}, \quad u = -10 \quad \text{on } \{1\} \times [0,1]. \end{aligned}$$

The source term f is given by

$$f(x) := 1000 \exp((-(x_1 - 0.15)(x_1 - 0.15) - 10(x_2 - 0.55)(x_2 - 0.55))).$$

For the discretization, we use continuous Q_1 finite elements on a uniform quadrilateral mesh with mesh size $h = 1/700$. The computational domain is partitioned into 7×7 rectangular, non-overlapping subdomains. To create overlapping subdomains $\{\text{supp}(\chi_i)\}_i$, we extend each subdomain by two layers of mesh elements in all directions.

The partition of unity is based on multiplication with patch-wise constant functions $\phi_i : \Omega \to [0,1]$, given by

$$\phi_i(x) = \frac{\mathbf{1}_{\text{supp}(\chi_i)}(x)}{\sum_{j=1}^{M} \mathbf{1}_{\text{supp}(\chi_j)}(x)} \quad \text{for } x \in \mathring{\text{supp}}(\chi_i) \cup \left(\partial\Omega \cap \overline{\text{supp}(\chi_i)}\right),$$

and $\phi_i(x) = 0$ else. Here, $\mathbf{1}_{\text{supp}(\chi_j)}$ denotes the indicator function of $\text{supp}(\chi_j)$. The partition of unity operators are then defined as

$$\chi_i(v) = \mathcal{I}(\phi_i \cdot v),$$

where $\phi_i \cdot v$ denotes point-wise multiplication and $\mathcal{I}$ the Lagrange interpolation operator.

Each $\text{supp}(\chi_i)$ is extended further by adding a few layers of elements to generate the corresponding oversampling domain ω_i. We refer to the number of these additional layers of elements as oversampling layers or "Ovsp" for short in the following. For the coarse space we use a fixed number of eigenfunctions on each subdomain, denoted by "#Eig". The experiment was run on a desktop PC with an AMD Ryzen 5 2600 processor. For the eigensolves we use Matlab's eigs(), a wrapper for Arpack. Local solves and coarse solves are implemented with the built-in backslash, a direct solver based on a multifrontal method. The iteration is stopped once a residual reduction of 10^{-10} is attained. The experiment was carried out in a sequential setting. However, to simulate parallel execution time, for any local computation the reported timings reflect only the maximum time taken across all subdomains. Unless otherwise noted, the results refer to the Richardson-type iteration with restricted hybrid Schwarz as described in Definition 6.4.

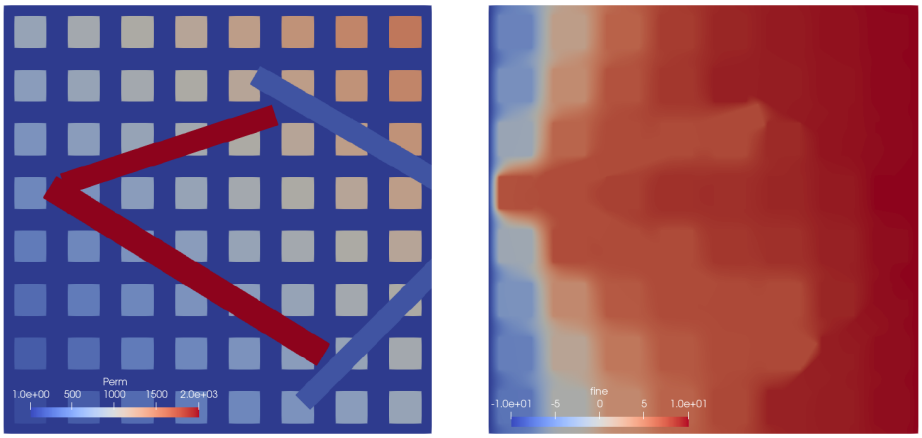


FIGURE 6.1: Diffusion equation example. Left: skyscraper coefficient. Right: solution. [58]

In Figure 6.2, we evaluate the performance of MS-GFEM used as a preconditioner in the Richardson-type iteration, for varying numbers of oversampling layers and local eigenfunctions. As predicted by theory, increasing either parameter significantly reduces the iteration count. Convergence rates as low as 10^{-9} per step are attainable with moderate parameter choices.

The fastest convergence rate does not correspond to the lowest computation time for the following reasons. Larger oversampling increases the cost of eigensolves and local solves. Adding more eigenfunctions enlarges the coarse space, thus raising the cost of coarse system assembly and coarse solves. As seen in Figure 6.2, a practical trade-off exists, leading to an optimal parameter region where computation time is minimized.

To test robustness under high contrast, we repeat the experiment using the square of the skyscraper coefficient from Figure 6.1 as the diffusion coefficient, yielding a

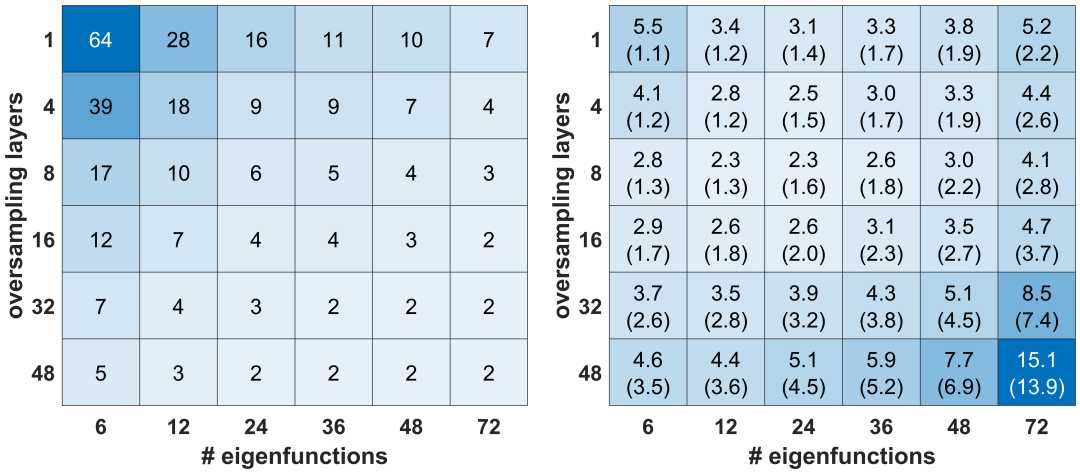


FIGURE 6.2: Diffusion problem solved with MS-GFEM. Left: number of iterations. Right: total computational time in seconds, with time spent on eigenproblems shown in brackets. Results vary by oversampling layers and number of eigenfunctions. [58]

contrast of $4 \cdot 10^6$. The corresponding results in Figure 6.3 show comparable performance to the lower contrast case — or even better when few eigenfunctions are used.

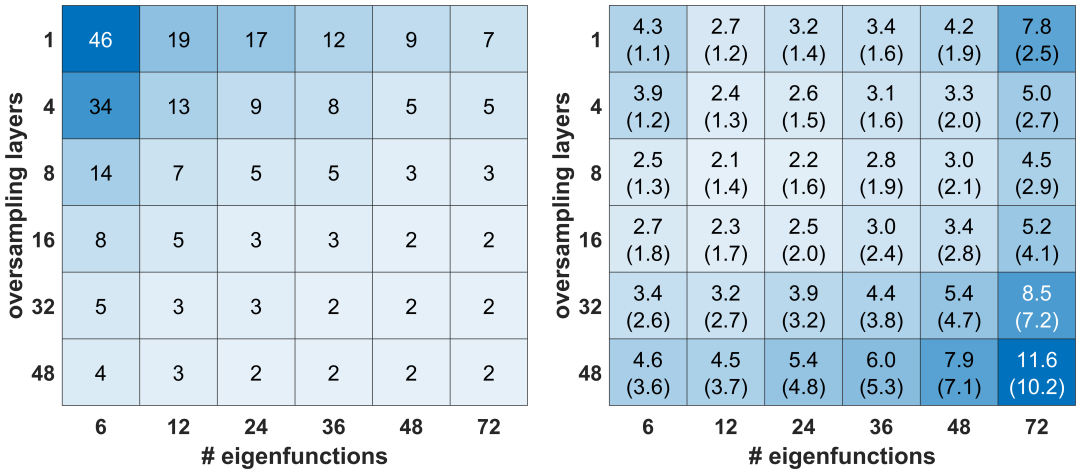


FIGURE 6.3: Same setup as in Figure 6.2, but with the square of the coefficient from Figure 6.1, increasing contrast to $4 \cdot 10^6$. The method remains robust under this change. [58]

Next, we investigate the block elimination strategy for the local eigenproblems, as proposed in [45]. The results in Figure 6.4 show a clear trade-off: while this approach significantly reduces eigenproblem cost, its lower accuracy increases the number of iterations — especially for cheaper coarse spaces. Thus, although this strategy may be preferable in many practical settings, we solve the eigenproblem directly for all other numerical results in this chapter to avoid introducing additional numerical error.

Figure 6.5 provides a more detailed view of the convergence behaviour for selected parameter choices. As expected, larger coarse spaces yield faster per-iteration convergence, but not always lower total runtime. The fastest overall time is achieved with 8 oversampling layers and 12 eigenfunctions per subdomain. This emphasizes that minimizing iteration count does not necessarily minimize computational time.

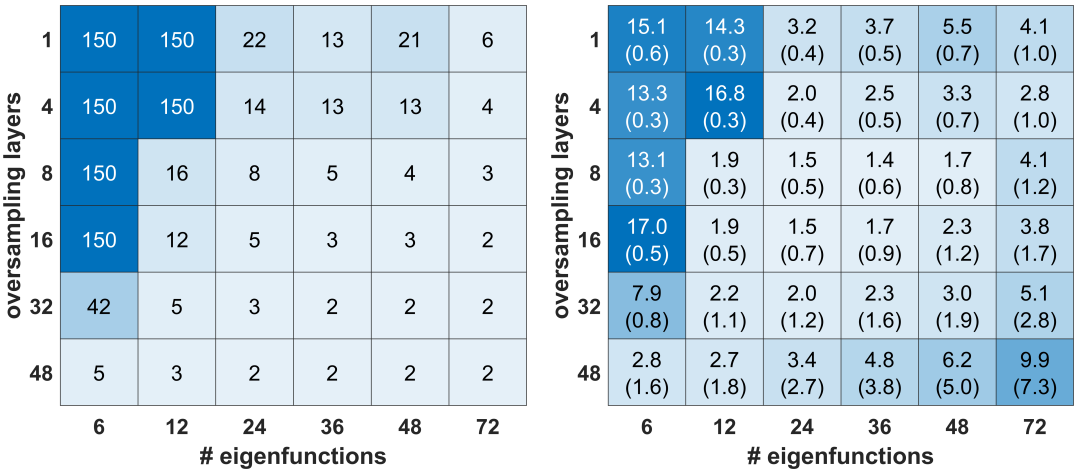


FIGURE 6.4: Same setup as in Figure 6.2, but using the block elimination eigenvalue solver from [45]. While eigenproblems are significantly cheaper, convergence slows due to reduced accuracy of the eigenvectors. The iteration count was capped at 150.

This remains true, even when ignoring setup costs, although in this case the sweet spot shifts towards more costly coarse spaces. Constructing a "perfect" coarse space is suboptimal in either scenario.

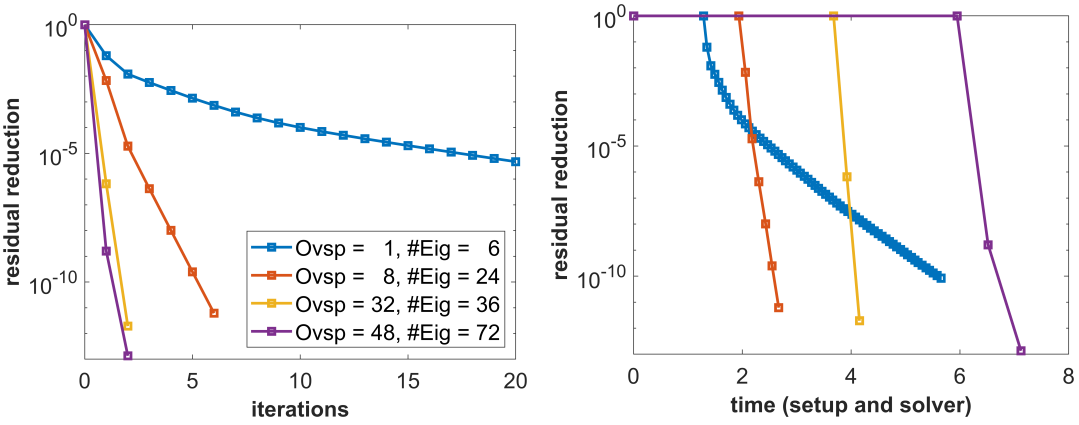


FIGURE 6.5: Convergence for selected parameter settings. Left: residual reduction per iteration. Right: residual reduction over runtime. [58]

Figure 6.6 (left) compares different Schwarz preconditioner schemes using a fixed MS-GFEM coarse space. Among the four methods the hybrid RAS scheme analysed in this chapter shows the best performance. Notably, this scheme converges rapidly even with the simple Richardson-type iteration. The others require GMRES to converge at all, while their corresponding Richardson-type iterations diverge (not shown).

In Figure 6.6 (right) we repeat the comparison using a GenEO coarse space. For comparability, we apply the same oversampling in eigenproblems and local solves. MS-GFEM consistently outperforms GenEO, which highlights the benefits of defining eigenproblems on a-harmonic subspaces. Nonetheless, GenEO likewise performs best with hybrid RAS which has become a popular setting in applications.

Figure 6.7 (top) explores convergence rate trends under varying oversampling layers and eigenfunction counts. Hybrid RAS behaves as expected, while the benefits of more expensive coarse spaces are not matched when employing other two-level

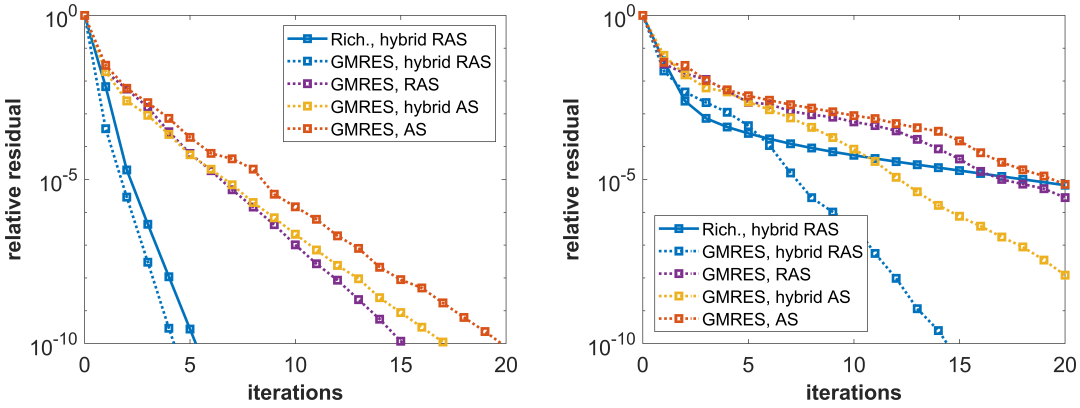


FIGURE 6.6: Comparison of preconditioner schemes for MS-GFEM (left) and GenEO (right). Hybrid RAS achieves the fastest convergence in both cases, and is the only scheme for which the Richardson-type iteration converges at all. Parameters: Ovsp=8, #Eig=24. [58]

Schwarz schemes such as (non-restricted) RAS. The bottom row analyses runtime trade-offs in more detail. Increasing oversampling or number of eigenfunctions reduces iteration count, but this is offset by higher setup and per-iteration costs. In both cases, an optimal region emerges where overall runtime is minimized.

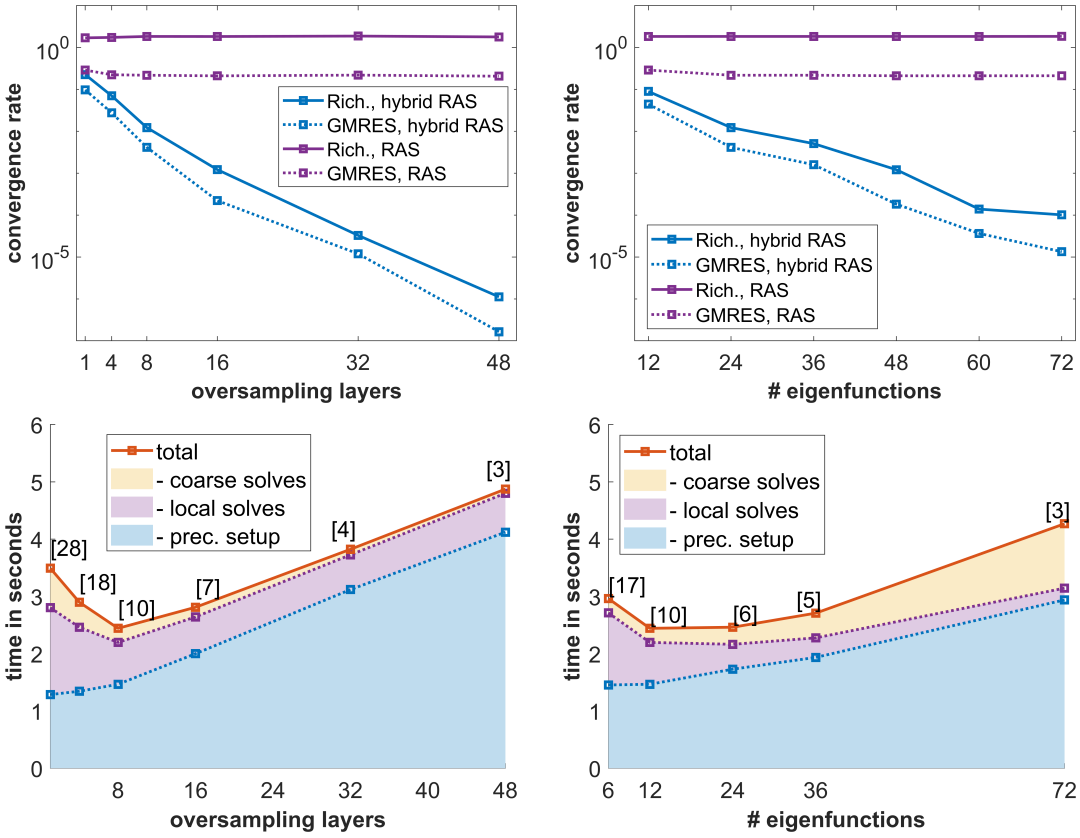


FIGURE 6.7: Top: convergence rate vs. oversampling layers (left, #Eig = 24) and number of eigenfunctions (right, Ovsp=8). Bottom: total runtime vs. oversampling (left, #Eig = 24) and eigenfunctions (right, Ovsp=8) for the MS-GFEM preconditioner with hybrid RAS in the Richardson-type iteration. Bracketed numbers indicate iteration counts. [58]

In summary, the MS-GFEM preconditioner demonstrates convergence rates approaching machine precision with moderate parameters. However, the fastest convergence is not always the most efficient in terms of runtime — even when setup costs are neglected. Hence, the MS-GFEM preconditioner is more practical than the corresponding full approximation approach.

Compared to GenEO and similar methods, MS-GFEM shows extreme gains in convergence rate. Still, if the coarse space cannot be reused, its higher setup cost complicates the overall efficiency trade-off.

6.3.3 Example 2: Linear Elasticity for Composite Aero-Structures

As a real-world application, we apply the MS-GFEM preconditioner to the system of three-dimensional linear elasticity equations posed on a composite structure from aero-engineering. The geometry is a C-shaped wing spar (C-spar), measuring 500 mm in length and featuring a central joggle region. The material is a laminated composite consisting of 24 uni-directional plies, each with a thickness of 0.2 mm, composed of carbon fibres embedded in resin. For a detailed description of the model, we refer to [52, 16]. The implementation draws significantly from [12].

We use continuous hexahedral finite elements of degree 1, with one mesh element in the thickness direction per ply. The mesh is partitioned into subdomains using ParMETIS [39]. Implementation is carried out in the DUNE framework [11, 52]. The left panel of Figure 6.8 illustrates the composite structure, the coefficient field (elasticity tensor), and the finite element mesh.

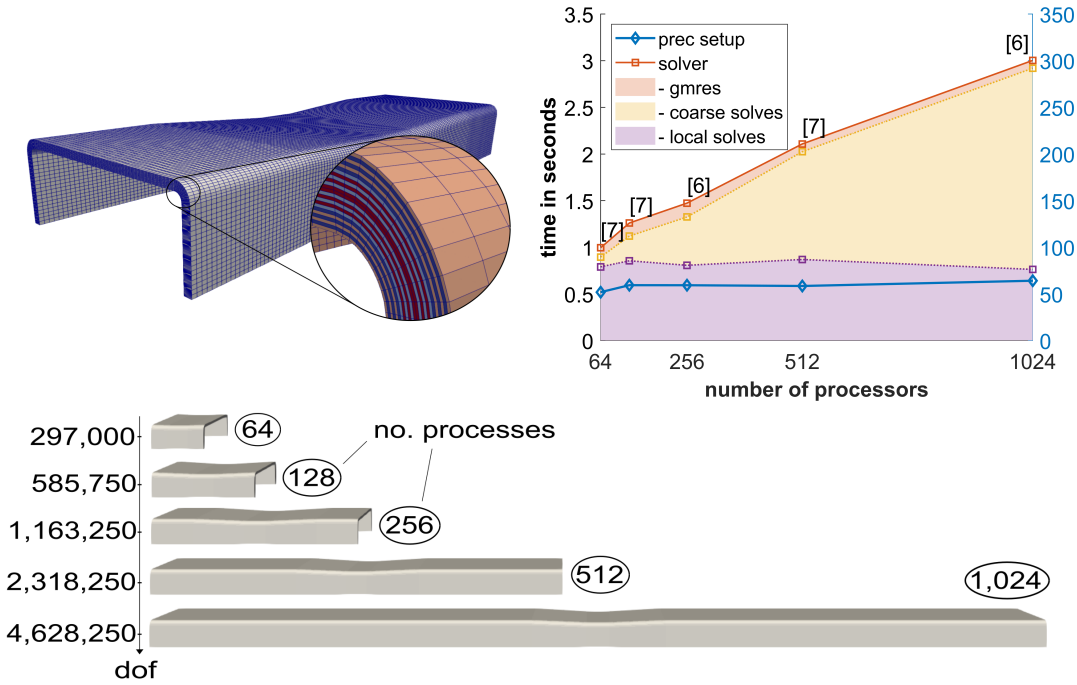


FIGURE 6.8: Composite C-spar experiment. Top left: illustration of the C-spar with laminate layers and mesh. Top right: results of a weak scaling test across five models described in Table 6.1, each using one oversampling layer and 30 eigenfunctions per subdomain. Values in brackets denote iteration counts. The preconditioner setup time (prec setup) corresponds to the right-hand y-axis. Bottom: visualizations of the five composite structures used for the weak scaling test. [58]

The experiment was run in parallel on the HPC bwForCluster Helix, which provides compute nodes equipped with 64 CPU cores (2x AMD EPYC 7513 Milan) and 236 GB available RAM per node. We use standard GMRES (restart after 10 iterations) to solve the preconditioned system, while local eigenproblems are handled by the Arpack interface in DUNE with the shift-invert method. A relative residual reduction of 10^{-6} is used as the convergence criterion for GMRES.

As partition of unity, we use distance-based functions which continuously drop from 1 to 0. The construction is described in detail in [12]. Throughout this example, we always use one processor per subdomain and thus use these terms interchangeably. Unlike in the previous experiment where we compared different Schwarz schemes, here we exclusively use hybrid RAS.

Figure 6.8 (right) shows the outcome of a weak scaling study over five models of increasing length, from 125 mm to 2000 mm. The mesh size is kept constant while the number of DOFs and processor cores scale proportionally (see Table 6.1). As expected, preconditioner setup time dominates overall cost and scales favourably with the number of processors. The cost of the coarse solve, however, increases significantly for larger problems. This bottleneck can be mitigated with multilevel methods as discussed previously.

Length in mm	DOFs	Cores
125	297,000	64
250	585,750	128
500	1,163,250	256
1000	2,318,250	512
2000	4,628,250	1024

TABLE 6.1: Details of the C-spar models used in the weak scaling test: domain length, number of degrees of freedom (DOFs), and number of processor cores. Unless otherwise noted, results refer to the 500 mm model on 256 cores. [58]

Figure 6.9 shows how both iteration counts and total time vary with the oversampling size and number of local eigenfunctions. As in the diffusion case (cf. Figure 6.2), increasing either parameter reduces the number of iterations, but raises computational cost for both, setup and iteration. The sweet spot for total runtime is achieved with minimal oversampling and around 10 eigenfunctions per subdomain. If setup cost is ignored, the best time is reached with one oversampling layer and 20 eigenfunctions.

Figure 6.10 (top) compares convergence rates of the hybrid RAS preconditioner with MS-GFEM and GenEO coarse spaces, across varying oversampling and eigenfunction parameters. MS-GFEM shows substantial improvement with more costly coarse spaces, while GenEO stagnates. This is consistent with expectations from theory, due to the accumulation of eigenvalues around 1, which effectively makes this a lower bound for the achievable threshold $\lambda_{\max}$. Figure 6.10, bottom, presents detailed runtime breakdowns for MS-GFEM. Oversampling increases eigensolve and local solve costs, while increasing the number of eigenfunctions primarily affects coarse assembly and coarse solve. Note that the x-axes range differs between examples: oversampling layers go only up to 4 here (vs. 48 in the diffusion experiment), and setup times are on a larger scale than solver times.

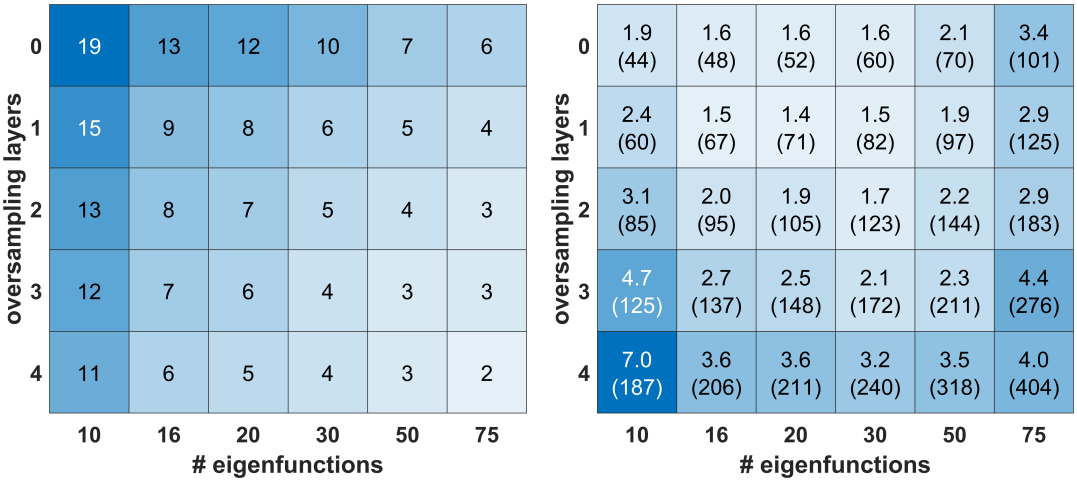


FIGURE 6.9: Composite C-spar experiment. Iteration counts (left) and solver time (right) for different numbers of oversampling layers and local eigenfunctions. On the right, setup times are shown inside brackets. [58]

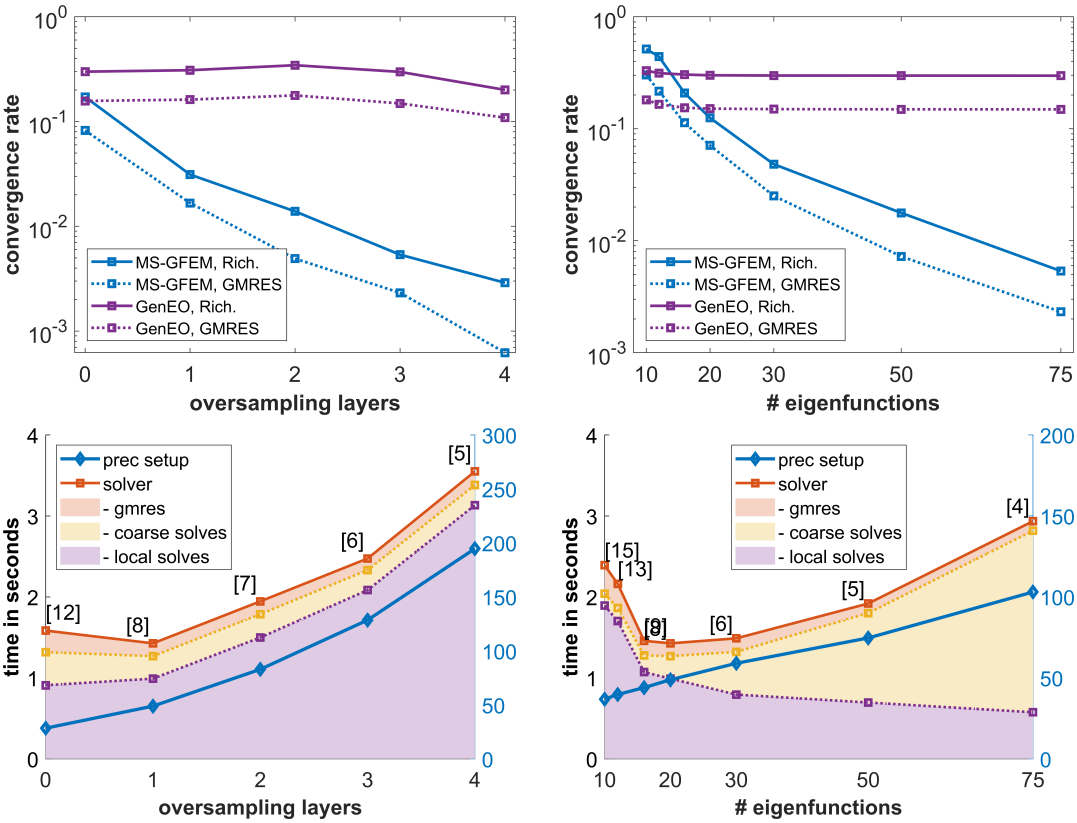


FIGURE 6.10: Composite C-spar experiment. Top: convergence rates for MS-GFEM and GenEO with oversampling (left) and eigenfunctions (right) varied. Bottom: total runtime of MS-GFEM preconditioned GMRES under the same variations.[58]

6.3.4 Discussion

The presented experiments confirm that the theoretical results of this chapter are applicable in practice — on a test case in Section 6.3.2, as well as in a real-world application in Section 6.3.3. The MS-GFEM preconditioner can be tuned to achieve arbitrarily fast convergence, with residual reductions approaching machine precision at levels around 10^{-9}. However, such extreme accuracy is generally not practical, as gains in convergence rate are offset by increased setup costs and greater computational effort per iteration.

This trade-off can be understood by examining the three main levers available to reduce the error: enlarging the coarse space (i.e. increasing the number of eigenfunctions), increasing the number of oversampling layers, and performing more iterations. According to Theorem 4.38, the first two strategies lead to nearly exponential improvements in the approximation error. The same is true for the third: each additional iteration also exponentially reduces the residual. It is therefore unsurprising that optimal performance is achieved not by pushing any one of these parameters to its extreme, but rather through a balanced combination of all three.

One notable difference between the two experiments is the behaviour with respect to oversampling. In the diffusion problem, oversampling significantly improves convergence, as predicted. In contrast, the linear elasticity example benefits very little from additional oversampling. This is in large part due to the nature of the subdomain partitioning: in the elasticity case, the domain is partitioned using an unstructured `ParMETIS` decomposition applied to a complex 3D geometry. This results in a significantly less favourable surface-to-volume ratio compared to the regular rectangular decomposition in 2D. Consequently, even a modest amount of oversampling can double the size of an unfavourably shaped subdomain, or worse. Notably, one such unfavourable subdomain is enough to slow down the entire computation in the given implementation. This can be alleviated, to some extent, by oversubscribing more than a single subdomain to each processor — allowing local computations with varying cost to even out. We will come back to these considerations in the conclusions of this thesis.

6.4 Outlook: The Helmholtz Equation

The MS-GFEM approximation has been shown to extend to the Helmholtz equation [44], and more broadly to indefinite bilinear forms in [42]. In this section, we report on numerical experiments illustrating that the MS-GFEM hybrid RAS preconditioner is also effective in this indefinite setting. These results were originally presented at the ENUMATH conference in Lisbon (2023). Since then, several theoretical advances have supported this approach. The hybrid RAS preconditioner for the Helmholtz equation has been analysed in [43], and related ideas have appeared in [37], [51], [33], and [31].

Throughout this section, we consider the Helmholtz problem in the frequency domain using the notation introduced in Section 1.3.3. Specifically, we seek $u \in H^1_{\Gamma_D}$ such that

$$\mathcal{B}(u,v) = F(v) \quad \text{for all } v \in H^1_{\Gamma_D}$$

with

$$\begin{aligned}\mathcal{B}(v,w) &:= \int_\Omega (A\nabla u \cdot \nabla \bar{v} - k^2 V^2 u\bar{v})\mathrm{d}x - ik\int_{\Gamma_R} \beta u\bar{v}\mathrm{d}s \\ F(v) &:= \int_{\Gamma_R} g\bar{v}\mathrm{d}s + \int_\Omega f\bar{v}\,\mathrm{d}x.\end{aligned}$$

Note that while $\mathcal{B}(\cdot,\cdot)$ is indefinite, it is still bounded from below in a certain sense: if we add a sufficiently large multiple of the L^2 inner product, we obtain a coercive form. When developing methods for this problem one thus often turns to the bilinear forms

$$a(u,v) := \int_\Omega A\nabla u \cdot \nabla \bar{v}\,\mathrm{d}x, \quad \text{and} \quad a_k(u,v) := \int_\Omega (A\nabla u \cdot \nabla \bar{v} + k^2 V^2 u\bar{v})\,\mathrm{d}x \tag{6.10}$$

for constructions where coercivity is needed. We will mostly use the a_k-norm in the following. As usual, we will denote their restrictions to a subdomain $\omega \subset \Omega$ by $\mathcal{B}_\omega$, a_ω, $a_{k,\omega}$. Further, since we will use Robin boundary conditions for the local solves, we define the bilinear forms

$$\mathcal{B}_{R,\omega}(v,w) := \mathcal{B}_\omega(v,w) - ik\int_{\partial\omega\cap\Omega} Vu\bar{v}\mathrm{d}s.$$

6.4.1 MS-GFEM in the Helmholtz Setting

Here, we define the MS-GFEM preconditioner for the Helmholtz equation, following the approximation theory introduced in [44]. The construction closely parallels the positive definite case discussed previously. Ideally, we would simply replace the positive definite form $a(\cdot,\cdot)$ by the indefinite form $\mathcal{B}(\cdot,\cdot)$. However, due to the lack of coercivity, this substitution is not always possible, and we must instead revert to one of the positive definite forms $a(\cdot,\cdot)$ or $a_k(\cdot,\cdot)$, introduced in (6.10).

As the wavenumber k increases, these auxiliary bilinear forms deviate more significantly from $\mathcal{B}$. It is therefore unsurprising that the preconditioner's performance deteriorates for larger k. This is a common challenge for solvers targeting the Helmholtz problem, and the dependency on k for MS-GFEM has been explicitly characterized in [44, 43].

We proceed analogously to the positive definite case described in Section 6.1, with adjustments to handle indefiniteness. Let H be the sheaf of continuous finite elements over a Lipschitz domain Ω, and let $\{\omega_i\}_{i\leq M}$ be an open cover resolved by the given

mesh which is admitting ξ^*-colouring. Let further $\{\chi_i\}_{i\leq M}$ be a subordinate partition of unity admitting ξ-colouring. We consider the $\mathcal{B}$-harmonic subspaces

$$H(\omega_i)^{\perp_{\mathcal{B}}} := \{v \in H(\omega_i) : \mathcal{B}_{\omega_i}(v,w) = 0 \quad \forall w \in H_0(\omega_i)\}.$$

The local eigenproblems read: find eigenvalues $\lambda_{i,1} \leq \cdots \leq \lambda_{i,m_i}$ and eigenfunctions $\phi_{i,j} \in H(\omega_i)^{\perp_{\mathcal{B}}}$, for $1 \leq j \leq m_i$, satisfying

$$a_{k,\omega_i}(\chi_i\phi_{i,j}, \chi_i v) = \lambda_{i,j}\, a_{\omega_i}(\phi_{i,j}, v) \quad \forall v \in H(\omega_i)^{\perp_{\mathcal{B}}}. \tag{6.11}$$

Note that here we use the coercive bilinear forms a and a_k rather than the indefinite form $\mathcal{B}$. The local approximation spaces are then defined as

$$S_i := \operatorname{span}\{\phi_{i,1}, \ldots, \phi_{i,m_i}\},$$

and the global coarse space $S = \sum_i \chi_i S_i$ is obtained by gluing these local spaces via the partition of unity.

Next, define the local and coarse solve operators analogously to the previous section — but now using the Helmholtz bilinear form $\mathcal{B}$ — by

$$\begin{aligned} &\pi_i : H(\Omega) \to H(\omega_i), && \mathcal{B}_{R,\omega_i}(\pi_i(v), w) = \mathcal{B}(v,w) \quad \forall w \in H(\omega_i), \\ &\pi_S : H(\Omega) \to S, && \mathcal{B}_{R,\Omega}(\pi_S(v), w) = \mathcal{B}(v,w) \quad \forall w \in S, \end{aligned}$$

where $\mathcal{B}_{R,\omega}$ incorporates suitable Robin conditions to ensure well-posedness, as detailed in [44]. The MS-GFEM approximation operator G and its associated preconditioner $\boldsymbol{B}$ are then defined exactly as in Section 6.1, that is, as the hybrid RAS method using the MS-GFEM coarse space.

6.4.2 Theoretical Results and Remarks

The following theorem provides an approximation estimate for the Helmholtz MS-GFEM preconditioner.

Theorem 6.25. *[44] Let $\mathcal{C}_a$ be a constant that bounds $a(\cdot,\cdot)$ in the a_k-norm as follows*

$$|a(v,w)| \leq \mathcal{C}_a \|v\|_{a_k} \|w\|_{a_k} \text{ for all } v, w \in H(\Omega),$$

and let C_{stab} be a constant such that

$$\|u\|_{a_k} \leq C_{stab} \left(\|f\|_{L^2(\Omega)} + \|g\|_{L^2(\Gamma_R)} \right).$$

Let further $C_{stab,i}$ be constants such that

$$\|u|_{\omega_i}\|_{a_k} \leq C_{stab} \left(\|f|\omega_i\|_{L^2} + \|g|\omega_i\|_{L^2(\Gamma_R)} \right)$$

and set $C_{\max} = \max_i C_{stab,i}$. Then we have the estimate

$$\|u - \boldsymbol{B}\boldsymbol{A}u\|_{a_k} \leq \mathcal{C}_a \sqrt{2\xi\xi^*\lambda_{\max}} \left(C_{stab} + \sqrt{2} C_{\max} \right) \left(\|f\|_{L^2(\Omega)} + \|g\|_{L^2(\Gamma_R)} \right),$$

where $\lambda_{\max}$ is the largest eigenvalue whose eigenvector was not added to the coarse space, as before.

Remark 6.26. The estimate above depends explicitly on the problem data f, g rather than on u itself. An estimate directly in terms of u has been derived in [37, 43], although these references use slightly different definitions for the local solves.

Remark 6.27. It is known from [47] that the constant C_a can be chosen independently of k. However, the stability constants C_{stab} and $C_{\max}$ do depend on k.

6.4.3 Numerical Experiment

We now consider a numerical experiment using the Marmousi problem, a widely-used benchmark example from geophysics. Specifically, we solve the Helmholtz equation as described in Section 1.3.3 with wavenumber $k = 40\pi$, constant coefficient $A(x) = \boldsymbol{I}$, a velocity field $1/V(x)$ as depicted in Figure 6.11, and Robin boundary parameter $\beta = kV$. The computational domain is rectangular, measuring 9 km $\times$ 3 km. By relative residual and convergence rate we refer to the corresponding quantities described in Section 6.3.

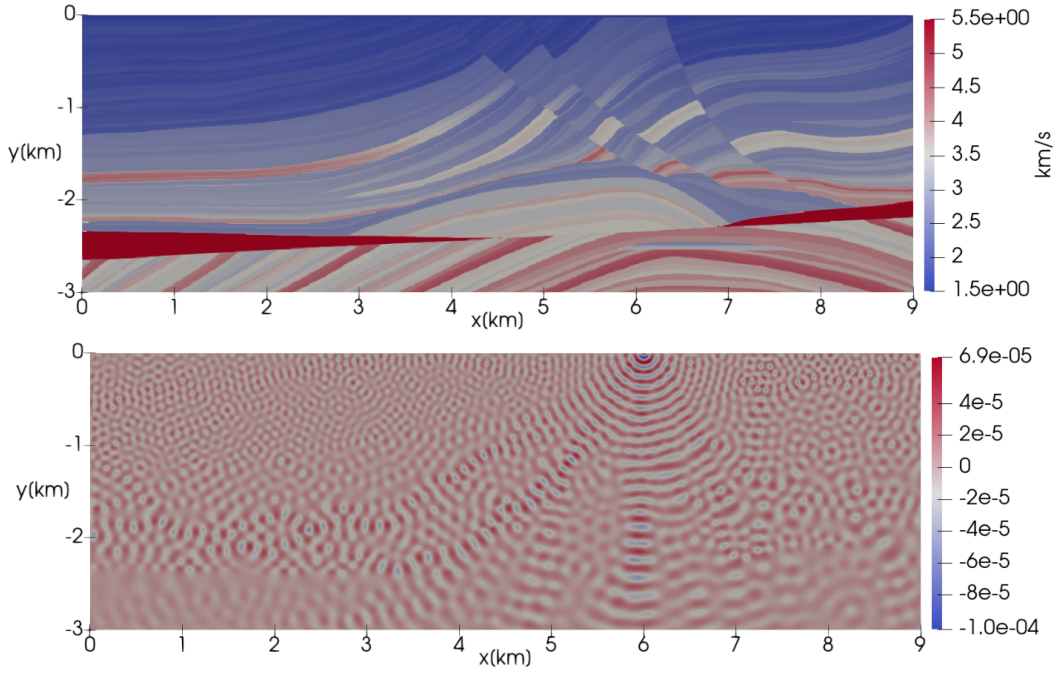


FIGURE 6.11: The Marmousi experiment: velocity field coefficient $1/V$ (top) and the corresponding numerical solution u (bottom).

We discretize the domain using a structured simplicial mesh containing 12000 $\times$ 4000 vertices, subdivided into 18 $\times$ 6 subdomains, each with size $H = 0.5$ km. Minimal overlap and an oversampling region of $H/6$ are used. For constructing the coarse space, we compute a fixed number of local eigenfunctions per subdomain. We then retain only those eigenfunctions whose eigenvalues exceed a threshold defined as the largest minimal eigenvalue across all subdomains, of those previously computed. Eigenfunctions corresponding to smaller eigenvalues do not further reduce $\lambda_{\max}$ in Theorem 6.25 and thus are excluded. We report the resulting coarse space dimension as a key parameter in the following plots.

The experiment was performed sequentially on two AMD Milan EPYC 7513 CPUs, with 32 GB of shared memory. Reported timings are estimates of parallel

execution, counting only the maximum computation time among any local computations that can be parallelized. Additionally, we measure the computational cost for eigenproblems, local solves, and coarse solves explicitly. Other overheads, such as the assembly of coarse and local matrices, are not included here, as they are common to all two-level Schwarz methods.

Figure 6.12 compares the solver performance across various coarse space dimensions. Similar to the positive definite cases, achieving convergence in just one iteration is not ideal — even when setup time is disregarded. This phenomenon becomes increasingly pronounced at higher wavenumbers, as the required coarse space dimension grows approximately linearly with the wavenumber (not shown here). Even a linear increase is problematic, as the coarse system's non-zeroes — and hence computational complexity — increase quadratically with the number of eigenfunctions per subdomain. For further analysis on wavenumber dependency, we refer to [44, 43].

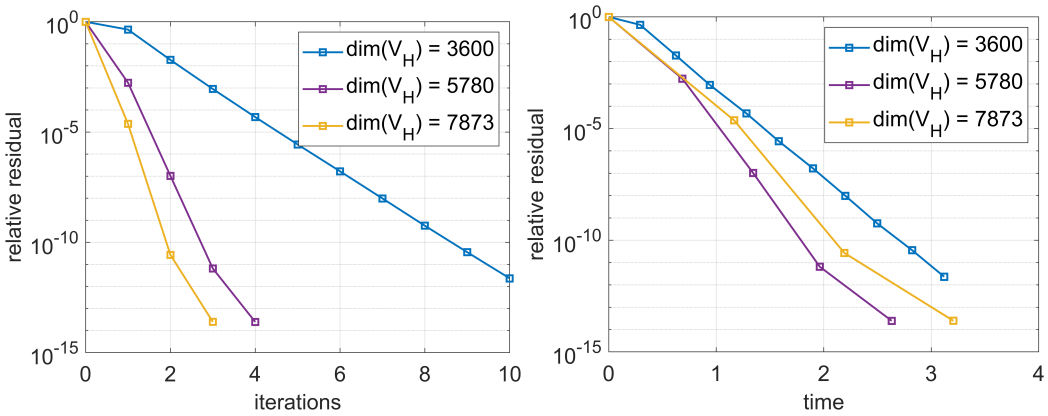


FIGURE 6.12: Relative residual reduction per iteration (left) and per computational time (right) for varying coarse space dimensions.

In Figure 6.13 (left) we examine the residual reduction as a function of total computation time, now including the time required for solving the local eigenproblems. The largest coarse space clearly looses appeal compared to smaller coarse spaces in this metric. On the right side, we compare convergence performance between Richardson-type and GMRES iterations. Here, GMRES provides minimal benefit.

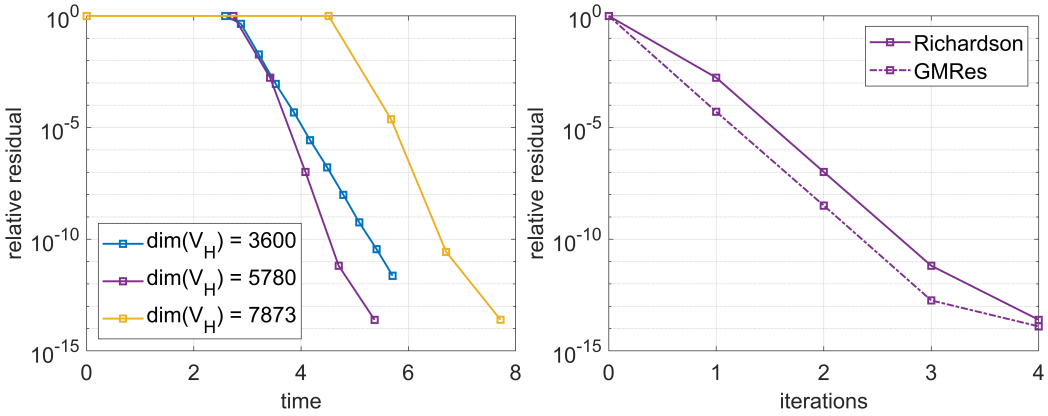


FIGURE 6.13: Relative residual reduction over total computation time (left). Comparison between Richardson-type and GMRES iterations (right).

In conclusion, the MS-GFEM preconditioner achieves very fast convergence for the Helmholtz equation, at least for moderate frequencies. For higher frequencies, rapid convergence demands increasingly large coarse spaces, suggesting potential benefits from employing a multilevel strategy.

Chapter 7

Conclusions

After establishing mathematical foundations and reviewing several spectral and related methods, this thesis introduced a general framework based on suitable presheaves and multilevel generalized elements. This abstract setting accommodates a broad class of PDE discretizations and spectral coarse space methods. We showed how GenEO and MS-GFEM naturally fit into this framework and demonstrated its applicability to both continuous finite elements and the discontinuous WSIP method. An abstract discussion of oversampling outlined how compact restrictions can lead to rapid eigenvalue decay in GMsFEM and MS-GFEM.

The core contributions are two abstract and robust spectral multilevel preconditioners: the non-restricted additive multilevel Schwarz method from Chapter 5, with robustness proven in Theorem 5.21, and the restricted hybrid multilevel Schwarz method from Chapter 6, with main result Theorem 6.24 and for which we demonstrated arbitrarily fast convergence when employed with MS-GFEM elements.

This thesis thus unifies and generalizes the author's previous work from [10] and [58], providing a wider theoretical foundation for the numerical methods developed therein. In particular, the GenEO and MS-GFEM variants for the discontinuous case presented here avoid the additional constant required by the Neumann variant of GenEO (cf. Example 4.49), as used in [10]. Moreover, we were able to easily extend the MS-GFEM preconditioner from [58] to a multilevel method. Finally, all theoretical results in this thesis are formulated within the framework of suitable presheaves, thus applying equally to continuous and discontinuous finite elements.

The numerical experiments underline the applicability of the developed preconditioners. They not only confirm the theoretical predictions but also demonstrate impressive robustness on both synthetic test cases and realistic applications from engineering practice. The experiments clearly highlight the advantages of the new methods, particularly their scalability and convergence speed. These merits come at the cost of an expensive setup phase, which we discuss in the following.

Generalizations of the Abstract Framework

The framework of suitable presheaves provides a unified perspective for a wide range of PDE discretizations, offering a multilevel formulation that bridges the gap between geometric and algebraic approaches. To this end, it draws on the established mathematical language of presheaves, rather than introducing entirely new terminology. Some foundational work is required in Chapter 3, where we introduce, among other concepts, the notions of locality of a bilinear form and the idea of extendables for a given presheaf H. These definitions are tailored to the assumptions used in the later chapters.

One limitation of the theory is that the canonical partition of unity used in the construction of multilevel generalized elements does not naturally fit into the presheaf framework. Ideally, one would like to define it as a map of presheaves $\chi : H \to H$. However, the canonical partition of unity on the presheaf of generalized elements fails to satisfy the necessary conditions. As a result, we resort to weaker assumptions in Section 3.1.4 to make the construct work.

The abstract framework covers both MS-GFEM and GenEO and can likely be extended to further spectral methods. We also briefly discuss application of the MS-GFEM preconditioner to the Helmholtz equation and provide numerical results for the two-level case. An even broader range of applications for MS-GFEM is discussed in [42]. Many of these require generalizing the framework to indefinite or non-symmetric problems, in which the local solves are no longer orthogonal projections.

This calls for a careful generalization of the presheaf setting which could be an avenue of future work.

Practical Considerations

From a theoretical perspective, MS-GFEM constitutes a clear improvement over the other methods discussed in this thesis. The MS-GFEM preconditioner guarantees arbitrarily fast convergence while simultaneously providing a bound on the coarse space size, as shown in Theorem 2.9. Comparable methods do not offer an analogous result. In particular, the best convergence rate that Theorem 2.6 can guarantee for Full GenEO in the hybrid additive Schwarz method is approximately 0.79 for a typical three-dimensional problem.

To see this, recall that Theorem 2.6 bounds the condition number of Full GenEO elements for the diffusion equation by

$$\kappa(\boldsymbol{B}_{\text{hy}}^{\text{GenEO}}\boldsymbol{A}) \leq \xi\,(1+\xi\lambda_{\max})$$

when used in the (non-restricted) hybrid Schwarz method. Since the GenEO eigenvalues accumulate around one, the best achievable scenario is $\lambda_{\max} = 1$. With $\xi = 8$, this leads to a bound on the condition number of 72. As this is a symmetric preconditioner, it is typically used with the Conjugate Gradient method, which then guarantees a convergence rate of $(\sqrt{\kappa}-1)/(\sqrt{\kappa}+1) \approx 0.7891$.

However, the improved convergence of MS-GFEM comes at the cost of a significantly more expensive setup. Oversampling and restricted eigenproblems introduce additional computational complexity. While it is possible to remove oversampling or simplify the eigenproblems, doing so sacrifices the theoretical guarantees of Theorem 2.9, effectively recovering GenEO. In this sense, GenEO can be viewed as a heuristic simplification of MS-GFEM. Historically, however, the development went in the opposite direction. GenEO was introduced first and the refinements leading to MS-GFEM were motivated by the goal of establishing a result on the decay of the local eigenvalues.

A major drawback of spectral methods, and thus of MS-GFEM in particular, is the high setup cost, which often constitutes a primary computational bottleneck. The extent of this issue varies by application, as illustrated by the 2D islands example in Section 6.3.2 and the 3D carbon fibre composite from Section 6.3.3. In the 2D case, the setup accounts for approximately 57–70% of the total runtime for optimal parameters, as shown in Figure 6.2. In contrast, the 3D example sees setup costs rise to around 96% of total time, even for the cheapest setup and thus optimal configuration tested (see Figure 6.9).

By comparison, the GenEO eigenproblems are generally much cheaper, unless the computation encounters the characteristic cluster of eigenvalues around 1. This can be problematic for iterative eigensolvers, since they are typically ill-conditioned when the final computed eigenvalues are nearly identical. This phenomenon can be observed for GenEO in Figure 7.1 where the setup time grows dramatically when enough eigenvectors are computed to enter the cluster of eigenvalues. Fortunately, the problematic eigenvectors are not required in the coarse space, so this issue of ill-conditioning is often avoidable in practice. When it is, GenEO's eigenproblems typically cost less than half of those in MS-GFEM.

Given that eigenvalue problems dominate the setup time, an obvious direction for improvement is their optimization. Several strategies were explored during implementation of the experiments, but ultimately the straightforward saddle point

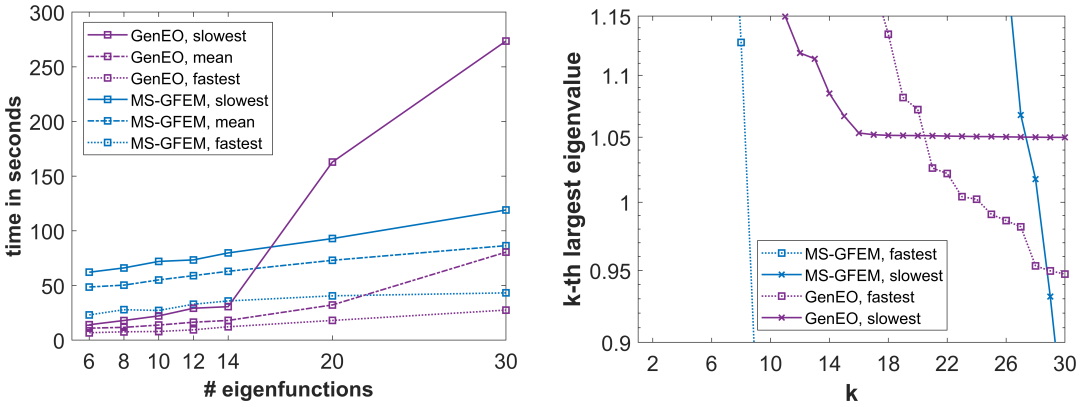


FIGURE 7.1: Left: cost comparison of eigensolves for MS-GFEM and GenEO in the C-Spar experiment from Section 6.3.3, here with length 500 mm, 128 subdomains and oversampling of 1 cell layer. As computation times vary across subdomains, slowest, fastest and average eigensolve times are plotted. GenEO remains significantly cheaper until it reaches an ill-conditioned eigenvalue cluster. Right: spectrum of the corresponding eigenproblems for the slowest and fastest eigensolves respectively. Scale on the y-axis is chosen so that the cluster near 1 becomes clearly visible.

formulation (6.8) was used. Alternative approaches such as the block elimination technique from [45] suffered from high memory usage and limited accuracy, while randomized solvers were slower than Arpack when targeting high accuracy. Nevertheless, substantial potential remains for improvement, and future work could explore this further. Notably, the strong scaling limit of spectral multilevel methods are of particular interest in that regard, since using very small subdomains opens up direct eigensolvers as possible solution strategies.

As a general rule of thumb, spectral coarse spaces are rarely the method of choice when simpler heuristic approaches yield satisfactory convergence. Their primary value lies in scenarios where standard techniques fail or prove insufficient. In such cases, the added complexity of spectral methods can be justified, and they become a powerful tool. Here, GenEO often presents the more practical option compared to MS-GFEM, as long as only a single problem instance must be solved. This is because in practice GenEO often performs significantly better than its theoretical bound of 0.79 suggests, with observed convergence rates typically below 0.3. Such performance is usually "good enough", and faster convergence at the cost of more expensive eigensolves becomes difficult to justify.

Secondly, spectral methods become more attractive when the coarse space can be reused across multiple solves. For example, when solving the same linear system with different right-hand sides, the cost of setup can be amortized over repeated use. In such settings, the faster convergence of MS-GFEM will eventually offset its higher setup cost. A concrete example of this occurs in the context of nonlinear PDEs solved via Newton's method, where a sequence of similar linear systems arises. In such cases, MS-GFEM has the potential to outperform GenEO, depending on the specifics of the problem.

Third, MS-GFEM offers a systematic and generalizable theoretical framework that makes it well-suited for application to a broad range of PDE problems. This was illustrated in Section 6.4, where we applied the method to the Helmholtz equation. While the setup remains expensive, the difficulty of the PDE can make this cost more

acceptable. Theoretical results for this case have since been developed in [43, 31, 37]. Additional applications of MS-GFEM are surveyed in [42], many of which can be encompassed within the abstract multilevel framework introduced in this thesis, pending appropriate generalization as discussed above.

Lastly, even when not employed directly, spectral methods can inspire efficient heuristic simplifications, such as those explored in [2]. Moreover, the theoretical understanding gained from these methods sheds light on the behaviour of existing preconditioners. As demonstrated in this thesis, the convergence proof for MS-GFEM provides insight into why GenEO performs best in a restricted hybrid additive Schwarz method, despite often being analysed in the unrestricted version.

Bibliography

[1] Eduardo Abreu, Ciro Díaz, and Juan Galvis. "A convergence analysis of Generalized Multiscale Finite Element Methods". In: *Journal of Computational Physics* 396 (2019), pp. 303–324. DOI: https://doi.org/10.1016/j.jcp.2019.06.072.

[2] Hussam Al Daas and Laura Grigori. "A class of efficient locally constructed preconditioners based on coarse spaces". In: *SIAM Journal on Matrix Analysis and Applications* 40.1 (2019), pp. 66–91. DOI: 10.1137/18M1194365.

[3] Hussam Al Daas, Pierre Jolivet, and Jennifer A. Scott. "A Robust Algebraic Domain Decomposition Preconditioner for Sparse Normal Equations". In: *SIAM Journal on Scientific Computing* 44.3 (2022), A1047–A1068. DOI: 10.1137/21M1434891.

[4] Hussam Al Daas et al. "A Multilevel Schwarz Preconditioner Based on a Hierarchy of Robust Coarse Spaces". In: *SIAM Journal on Scientific Computing* 43.3 (2021), A1907–A1928. DOI: 10.1137/19M1266964.

[5] Christian Alber, Chupeng Ma, and Robert Scheichl. "A mixed multiscale spectral generalized finite element method". In: *Numerische Mathematik* (2025), pp. 1–40. DOI: https://doi.org/10.1007/s00211-024-01446-3.

[6] Robert Altmann, Patrick Henning, and Daniel Peterseim. "Numerical homogenization beyond scale separation". In: *Acta Numerica* 30 (2021), pp. 1–86. DOI: 10.1017/S0962492921000015.

[7] Ivo Babuška and Jens M Melenk. "The partition of unity method". In: *International Journal for Numerical Methods in Engineering* 40.4 (1997), pp. 727–758.

[8] Lori Badea. "On the Schwarz Alternating Method with More than two Subdomains for Nonlinear Monotone Problems". In: *SIAM Journal on Numerical Analysis* 28.1 (1991), pp. 179–204. DOI: 10.1137/0728010.

[9] Peter Bastian et al. "A generic grid interface for parallel and adaptive scientific computing. Part II: Implementation and tests in DUNE". In: *Computing* 82.2 (2008), pp. 121–138. DOI: 10.1007/s00607-008-0004-9.

[10] Peter Bastian et al. "Multilevel Spectral Domain Decomposition". In: *SIAM Journal on Scientific Computing* 45.3 (2023), S1–S26. DOI: 10.1137/21M1427231.

[11] Peter Bastian et al. "The Dune framework: Basic concepts and recent developments". In: *Computers & Mathematics with Applications* 81 (2021), pp. 75–112. DOI: 10.1016/j.camwa.2020.06.007.

[12] Jean Bénézech et al. "Scalable multiscale-spectral GFEM with an application to composite aero-structures". English. In: *Journal of Computational Physics* 508 (2024), p. 113013. DOI: 10.1016/j.jcp.2024.113013.

[13] Alain Bensoussan, Jacques-Louis Lions, and George Papanicolaou. *Asymptotic Analysis for Periodic Structures*. Contributions to Economic Analysis. North-Holland Publishing Company, 1978. URL: https://books.google.de/books?id=1b-vlAEACAAJ.

[14] Niall Bootland et al. "GenEO Coarse Spaces for Heterogeneous Indefinite Elliptic Problems". In: *Domain Decomposition Methods in Science and Engineering XXVI*. Springer International Publishing, 2022, pp. 117–125. DOI: 10.1007/978-3-030-95025-5_10.

[15] William L. Briggs and Van Emden Henson. "The FFT as a Multigrid Algorithm". In: *SIAM Review* 32.2 (1990), pp. 252–261. DOI: 10.1137/1032045.

[16] R. Butler et al. "High-performance dune modules for solving large-scale, strongly anisotropic elliptic problems with applications to aerospace composites". In: *Computer Physics Communications* 249 (2020), p. 106997. DOI: 10.1016/j.cpc.2019.106997.

[17] Yanqing Chen et al. "Algorithm 887: CHOLMOD, Supernodal Sparse Cholesky Factorization and Update/Downdate". In: *ACM Trans. Math. Softw.* 35.3 (2008). DOI: 10.1145/1391989.1391995.

[18] Michael Andrew Christie and Martin J Blunt. "Tenth SPE comparative solution project: A comparison of upscaling techniques". In: *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers. 2001. DOI: 10.2118/72469-PA.

[19] Richard Courant. "Variational methods for the solution of problems of equilibrium and vibrations". In: *Bulletin of the American Mathematical Society* 49.1 (1943), pp. 1 –23.

[20] Timothy A. Davis. "Algorithm 832: UMFPACK V4.3—an Unsymmetric-Pattern Multifrontal Method". In: *ACM Trans. Math. Softw.* 30.2 (2004), pp. 196–199. DOI: 10.1145/992200.992206.

[21] Victorita Dolean, Pierre Jolivet, and Frédéric Nataf. *An Introduction to Domain Decomposition Methods*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2015. DOI: 10.1137/1.9781611974065.

[22] Victorita Dolean et al. "Analysis of a Two-level Schwarz Method with Coarse Spaces Based on Local Dirichlet-to-Neumann Maps". In: *Computational Methods in Applied Mathematics* 12.4 (2012), pp. 391–414. DOI: doi:10.2478/cmam-2012-0027.

[23] Yalchin Efendiev, Juan Galvis, and Thomas Y. Hou. "Generalized multiscale finite element methods (GMsFEM)". In: *Journal of Computational Physics* 251 (2013), pp. 116–135. DOI: https://doi.org/10.1016/j.jcp.2013.04.045.

[24] Yalchin Efendiev, Juan Galvis, and Xiao-Hui Wu. "Multiscale finite element methods for high-contrast problems using local spectral basis functions". In: *Journal of Computational Physics* 230.4 (2011), pp. 937–955. DOI: https://doi.org/10.1016/j.jcp.2010.09.026.

[25] Yalchin Efendiev and Thomas Y Hou. *Multiscale Finite Element Methods. Theory and Applications*. Springer New York, NY, 2009. URL: https://link.springer.com/book/10.1007/978-0-387-09496-0.

[26] Yalchin Efendiev Eric Chung and Thomas Hou. *Multiscale Model Reduction*. Springer, 2023. URL: https://link.springer.com/content/pdf/10.1007/978-3-031-20409-8.pdf.

[27] Alexandre Ern and Jean-Luc Guermond. *Finite Elements I: Approximation and Interpolation*. Texts in Applied Mathematics. Springer International Publishing, 2021. URL: https://link.springer.com/book/10.1007/978-3-030-56341-7.

[28] Alexandre Ern and Jean-Luc Guermond. *Finite Elements II: Galerkin Approximation, Elliptic and Mixed PDEs*. Springer, 2021. URL: https://hal.science/hal-03226050.

[29] Alexandre Ern, Annette F. Stephansen, and Paolo Zunino. "A Discontinuous Galerkin method with weighted averages for advection–diffusion equations with locally small and anisotropic diffusivity". In: *IMA Journal of Numerical Analysis* 29.2 (2008), pp. 235–256. DOI: 10.1093/imanum/drm050.

[30] Joseph Fourier. *Théorie analytique de la chaleur*. Landmarks of Science. Didot, 1822. URL: https://books.google.de/books?id=1TUVAAAAQAAJ.

[31] Jeffrey Galkowski and Euan A. Spence. *Convergence theory for two-level hybrid Schwarz preconditioners for high-frequency Helmholtz problems*. 2025.

[32] Juan Galvis and Yalchin Efendiev. "Domain Decomposition Preconditioners for Multiscale Flows in High-Contrast Media". In: *Multiscale Modeling & Simulation* 8.4 (2010), pp. 1461–1483. DOI: 10.1137/090751190.

[33] Ivan G. Graham and Euan A. Spence. *Two-level hybrid Schwarz preconditioners with piecewise-polynomial coarse spaces for the high-frequency Helmholtz equation*. 2025. URL: https://arxiv.org/abs/2501.15976.

[34] George Green. *An Essay on the Application of mathematical Analysis to the theories of Electricity and Magnetism*. 2008.

[35] Thomas Y. Hou and Xiao-Hui Wu. "A Multiscale Finite Element Method for Elliptic Problems in Composite Materials and Porous Media". In: *J. Comput. Phys.* 134.1 (1997), pp. 169–189. DOI: 10.1006/jcph.1997.5682.

[36] Alexander Hrennikoff. "Solution of Problems of Elasticity by the Framework Method". In: *Journal of Applied Mechanics* 8.4 (2021), A169–A175. DOI: 10.1115/1.4009129.

[37] Qiya Hu and Ziyi Li. *A novel coarse space applying to the weighted Schwarz method for Helmholtz equations*. 2024. URL: https://arxiv.org/abs/2402.06905.

[38] Patrick Lechner Ivan Graham and Robert Scheichl. "Domain decomposition for multiscale PDEs". In: *Numerische Mathematik* 106 (2007), pp. 589–626. URL: https://doi.org/10.1007/s00211-007-0074-1.

[39] George Karypis and Vipin Kumar. "Multilevel *k*-way Partitioning Scheme for Irregular Graphs". In: *Journal of Parallel and Distributed Computing* 48.1 (1998), pp. 96 –129. DOI: 10.1006/jpdc.1997.1404.

[40] John Lee. *Introduction to Smooth Manifolds*. 2nd ed. Vol. 218. Graduate Texts in Mathematics. New York: Springer, 2013. URL: https://link.springer.com/book/10.1007/978-0-387-21752-9.

[41] Richard B Lehoucq, Danny C Sorensen, and Chao Yang. *Arpack Users' Guide*. Society for Industrial and Applied Mathematics, 1998. URL: https://epubs.siam.org/doi/abs/10.1137/1.9780898719628.

[42] Chupeng Ma. "A unified framework for multiscale spectral generalized FEMs and low-rank approximations to multiscale PDEs". In: *Foundations of Computational Mathematics* (2025), pp. 1–60. DOI: https://doi.org/10.1007/s10208-025-09711-z.

[43] Chupeng Ma, Christian Alber, and Robert Scheichl. *Two-level Restricted Additive Schwarz preconditioner based on Multiscale Spectral Generalized FEM for Heterogeneous Helmholtz Problems*. 2024. URL: https://arxiv.org/abs/2409.06533.

[44] Chupeng Ma, Christian Alber, and Robert Scheichl. "Wavenumber Explicit Convergence of a Multiscale Generalized Finite Element Method for Heterogeneous Helmholtz Problems". In: *SIAM Journal on Numerical Analysis* 61.3 (2023), pp. 1546–1584. DOI: 10.1137/21M1466748.

[45] Chupeng Ma and Robert Scheichl. "Error estimates for discrete generalized FEMs with locally optimal spectral approximations". In: *Mathematics of Computation* 91.338 (2022), pp. 2539–2569.

[46] Chupeng Ma, Robert Scheichl, and Tim Dodwell. "Novel Design and Analysis of Generalized Finite Element Methods Based on Locally Optimal Spectral Approximations". In: *SIAM Journal on Numerical Analysis* 60.1 (2022), pp. 244–273. DOI: https://doi.org/10.1137/21M1406179.

[47] Jens Markus Melenk. *On Generalized Finite Element Methods*. University of Maryland at College Park, 1995. URL: https://books.google.de/books?id=n9CRnQEACAAJ.

[48] Artem Napov and Yvan Notay. "An Algebraic Multigrid Method with Guaranteed Convergence Rate". In: *SIAM Journal on Scientific Computing* 34.2 (2012), A1079–A1109. DOI: 10.1137/100818509.

[49] Frédéric Nataf and Emile Parolin. *Coarse spaces for non-symmetric two-level preconditioners based on local generalized eigenproblems*. 2024. URL: https://arxiv.org/abs/2404.02758.

[50] Frédéric Nataf, Hua Xiang, and Victorita Dolean. "A two level domain decomposition preconditioner based on local Dirichlet-to-Neumann maps". In: *Comptes Rendus Mathematique* 348.21 (2010), pp. 1163–1167.

[51] Bowen Zheng Peipei Lu Xuejun Xu and Jun Zou. *Two-level hybrid Schwarz Preconditioners for The Helmholtz Equation with high wave number*. 2025. URL: https://arxiv.org/abs/2408.07669.

[52] Anne Reinarz et al. "Dune-composites – A new framework for high-performance finite element modelling of laminates". In: *Composite Structures* 184 (2018), pp. 269 –278. DOI: 10.1016/j.compstruct.2017.09.104.

[53] Y. Saad. *Iterative Methods for Sparse Linear Systems*. PWS Publishing Company, Boston, 1996. URL: https://www-users.cse.umn.edu/~saad/IterMethBook_2ndEd.pdf.

[54] Robert Scheichl, Panayot S Vassilevski, and Ludmil T Zikatanov. "Weak approximation properties of elliptic projections with functional constraints". In: *Multiscale Modeling & Simulation* 9.4 (2011), pp. 1677–1699. DOI: https://doi.org/10.1137/110821639.

[55] Barry Smith, Petter Bjorstad, and William Gropp. *Domain Decomposition – Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, 1996.

[56] N. Spillane et al. "Abstract Robust Coarse Spaces for Systems of PDEs via Generalized Eigenproblems in the Overlaps". In: *Numer. Math.* 126.4 (2014), pp. 741–770.

[57] Nicole Spillane. "Robust domain decomposition methods for symmetric positive definite problems". Theses. Université Pierre et Marie Curie - Paris VI, 2014. URL: https://theses.hal.science/tel-00958252.

[58] Arne Strehlow, Chupeng Ma, and Robert Scheichl. *Fast-convergent two-level restricted additive Schwarz methods based on optimal local approximation spaces*. 2024. URL: https://arxiv.org/abs/2408.16282.

[59] Theofanis Strouboulis, Ivo Babuška, and Kevin Copps. "The design and analysis of the Generalized Finite Element Method". In: *Computer Methods in Applied Mechanics and Engineering* 181.1 (2000), pp. 43–69. DOI: https://doi.org/10.1016/S0045-7825(99)00072-9.

[60] Barry R Tennison. *Sheaf Theory*. London Mathematical Society Lecture Note Series. Cambridge University Press, 1975.

[61] Andrea Toselli and Olof Widlund. *Domain Decomposition Methods - Algorithms and Theory*. Springer Berlin, Heidelberg, 2006. URL: https://link.springer.com/book/10.1007/b137868.

[62] Wei Xie et al. "CEM-GMsFEM for Poisson Equations in Heterogeneous Perforated Domains". In: *Multiscale Modeling & Simulation* 22.4 (2024), pp. 1683–1708. DOI: 10.1137/24M1641816.

[63] Jinchao Xu. "Iterative Methods by Space Decomposition and Subspace Correction". In: *SIAM Review* 34.4 (1992), pp. 581–613. DOI: 10.1137/1034116.

[64] Jinchao Xu and Ludmil Zikatanov. "Algebraic multigrid methods". In: *Acta Numerica* 26 (2017), pp. 591–721. DOI: 10.1017/S0962492917000083.

[65] Changqing Ye and Eric T. Chung. "Constraint Energy Minimizing Generalized Multiscale Finite Element Method for Inhomogeneous Boundary Value Problems with High Contrast Coefficients". In: *Multiscale Modeling & Simulation* 21.1 (2023), pp. 194–217. DOI: 10.1137/21M1459113.