Operator preconditioning and space-time methods for parabolic evolution equations

van Venetië, R.

Publication date
2021

Document Version
Final published version

Operator preconditioning and Space-time methods for parabolic evolution equations

Raymond van Venetië

Operator preconditioning
and
Space-time methods for parabolic evolution equations

Raymond van Venetië

About the cover: the displayed mesh is found by running the adaptive method from Chapter 9 on the 'boundary' of this book. The solution is strongly singular for $t = 0$, which explains the refinements towards the bottom of the cover.

This research was funded by the Netherlands Organization for Scientific Research (NWO) under contract. no. 613.001.652.

Printed by GVO drukkers & vormgevers
ISBN: 978-94-6332-788-6

Operator preconditioning
and
Space-time methods for parabolic evolution equations

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties
ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op dinsdag 28 september 2021, te 15.00 uur
door

Raymond van Venetië
geboren te Zoetermeer

Promotiecommissie

Promotor:	prof. dr. R.P. Stevenson	Universiteit van Amsterdam
Copromotor:	dr. J.H. Brandts	Universiteit van Amsterdam
Overige leden:	prof. dr. H. Peters	Universiteit van Amsterdam
	prof. dr. A.J. Homburg	Universiteit van Amsterdam
	dr. C.C. Stolk	Universiteit van Amsterdam
	dr. G. Gantner	Universiteit van Amsterdam
	prof. dr. U. Langer	Johannes Kepler University Linz
	dr. C.A. Urzúa-Torres	Technische Universiteit Delft

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

Contents

1 Introduction

Partial differential equations are used for the modeling of a wide variety of (natural) phenomena appearing in biology, physics, chemistry, engineering, finance and many more fields. Typically, closed-form solutions to these differential equations are unknown or do not exist, despite their importance in practice. A remedy is provided by numerical methods that construct *approximations* of the solutions to these differential equations. Ideally, these numerical methods provide good approximations at a small computational cost. This is the main topic of investigation in this thesis, where we will focus on *linear* partial differential equations of elliptic or parabolic type.

The contents of this thesis can be roughly divided into two parts. In the first part we study *preconditioning*, a technique that is used to accelerate solvers for systems of linear equations arising in our approximation schemes. In the second part, we focus on (adaptive) numerical methods for parabolic evolution equations in a simultaneous *space-time* approach.

1.1 Numerical methods for operator equations

We start with a brief description of the general setting that we study in this thesis. For a thorough introduction, we refer the reader to the literature, e.g. [Bra01, EG04, Ste08a].

1.1.1 Operator equations

We consider *linear* operator equations of the following type. For some Hilbert space $\mathcal{V}$, a linear map $A\colon \mathcal{V} \to \mathcal{V}'$ and data $f \in \mathcal{V}'$, we seek $u \in \mathcal{V}$ that solves

$$Au = f \quad \text{or equivalently} \quad (Au)(v) = f(v) \quad (v \in \mathcal{V}). \tag{1.1}$$

We require the problem (1.1) to be *well-posed*, meaning that there is a unique solution that depends continuously on the given data. Typically, $\mathcal{V}$ is a Sobolev space of functions on some bounded domain, and the operator A corresponds to the variational formulation of some partial differential equation, but we also encounter situations where A is a boundary integral operator.

Let us restrict ourselves to A being a *bounded* and *coercive* operator, meaning that for all $v \in \mathscr{V}$ we have $\|Av\|_{\mathscr{V}'} \leq C\|v\|_{\mathscr{V}}$ and $(Av)(v) \geq \alpha\|v\|_{\mathscr{V}}^2$, for some $C < \infty$ and $\alpha > 0$. In this setting, the Lax–Milgram theorem asserts that the problem (1.1) is well-posed with $\|A^{-1}\|_{\mathcal{L}(\mathscr{V}',\mathscr{V})} \leq 1/\alpha$.

Example 1.1.1. As a model problem for (elliptic) partial differential equations, one may look at Poisson's equation with homogeneous Dirichlet boundary conditions. For some bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$, this is the problem of finding u that solves

$$-\Delta u = f \quad \text{on } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega,$$

for given forcing data f.

In variational formulation this problem reads as finding u from the Sobolev space $H^1_0(\Omega) := \{v \in H^1(\Omega) : v|_{\partial\Omega} = 0\}$ that satisfies

$$(Au)(v) := \langle \nabla u, \nabla v\rangle_{L_2(\Omega)} = \langle f, v\rangle_{L_2(\Omega)} \quad (v \in H^1_0(\Omega)). \tag{1.2}$$

The operator A is a bounded linear map $H^1_0(\Omega) \to H^1_0(\Omega)'$. Moreover, the Poincaré inequality shows that A is coercive. We conclude that (1.2) is a well-posed operator equation for forcing data $f \in H^1_0(\Omega)'$, and falls in the abstract setting of (1.1).

Example 1.1.2. Another model problem of interest is Laplace's equation with inhomogeneous Dirichlet boundary conditions. For some bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$, this is the problem of finding u that solves

$$\Delta u = 0 \quad \text{on } \Omega, \qquad u = g \quad \text{on } \partial\Omega, \tag{1.3}$$

for given Dirichlet boundary data g. This problem admits an (alternative) weak formulation in terms of boundary integral operators.

This requires some notation. Set $\Gamma := \partial\Omega$, and consider the fractional Sobolev space $H^{\frac{1}{2}}(\Gamma)$, which can be seen as the trace space of $H^1(\Omega)$, with its dual that we denote by $H^{-\frac{1}{2}}(\Gamma)$. Let G be the fundamental solution to the Laplace equation (1.3), e.g., for $d = 2$ we have $G(\cdot) := -\frac{1}{2\pi}\log|\cdot|$. We can now introduce the *Single Layer* integral operator $A\colon H^{-\frac{1}{2}}(\Gamma) \to H^{\frac{1}{2}}(\Gamma)$ by

$$(A\tilde{u})(\tilde{v}) := \int_\Gamma \tilde{v}(\boldsymbol{x}) \int_\Gamma G(\boldsymbol{x}-\boldsymbol{y})\tilde{u}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{x} \quad \big(\tilde{u}, \tilde{v} \in H^{-\frac{1}{2}}(\Gamma)\big).$$

The Single Layer operator A is bounded and coercive, so for $g \in H^{\frac{1}{2}}(\Gamma)$ we may consider the well-posed problem of finding $\tilde{u} \in H^{-\frac{1}{2}}(\Gamma)$ that solves

$$A\tilde{u} = g.$$

It turns out that this problem is equivalent to the original problem (1.3), in that we can recover the weak solution $u \in H^1(\Omega)$ via the representation formula $u(\boldsymbol{x}) = \int_\Gamma G(\boldsymbol{x}-\boldsymbol{y})\tilde{u}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}$.

Finally, we note that solving the Laplace problem with Neumann boundary data—so replacing the boundary condition by $\frac{\partial u}{\partial \boldsymbol{n}} = g$ in (1.3)—allows for a similar weak formulation in terms of boundary integral operators. Here one solves $Bu = g$, with the *Hypersingular* integral operator $B\colon H^{\frac{1}{2}}(\Gamma) \to H^{-\frac{1}{2}}(\Gamma)$. This operator maps in the 'opposite direction' of A, which will be important later in this introduction.

1.1.2 Numerical approximation

For the numerical approximation of the operator equation (1.1) we can use the Ritz–Galerkin method. For some finite-dimensional subspace $\mathscr{V}_\mathcal{T} \subset \mathscr{V}$, also called the *trial space*, we consider the discretized operator $A_\mathcal{T}\colon \mathscr{V}_\mathcal{T} \to \mathscr{V}_\mathcal{T}'$ given by $(A_\mathcal{T} u)(v) := (Au)(v)$ $(u, v \in V_\mathcal{T})$ and the naturally embedded right hand side $f \in \mathscr{V}_\mathcal{T}'$. The Galerkin approximation of (1.1) is then the function $u_\mathcal{T} \in \mathscr{V}_\mathcal{T}$ that solves

$$A_\mathcal{T} u_\mathcal{T} = f. \tag{1.4}$$

Since $\mathscr{V}_\mathcal{T}$ is a closed subspace of $\mathscr{V}$, the operator $A_\mathcal{T}$ is again coercive, and hence the Lax–Milgram theorem asserts that this discretized problem is well-posed.

The advantage of the Galerkin method is that it provides us with a concrete and tight error bound. Indeed, Céa's lemma shows that the error satisfies

$$\|u - u_\mathcal{T}\|_\mathscr{V} \leq \|A\|_{\mathcal{L}(\mathscr{V},\mathscr{V}')} \|A^{-1}\|_{\mathcal{L}(\mathscr{V}',\mathscr{V})} \inf_{v \in \mathscr{V}_\mathcal{T}} \|u - v\|_\mathscr{V},$$

which tells us that $u_\mathcal{T}$ is a *quasi-best* approximation to u from $\mathscr{V}_\mathcal{T}$.

To design a convergent numerical method, we actually have to construct a family of trial spaces $(\mathscr{V}_\mathcal{T})_{\mathcal{T} \in \mathbb{T}}$. Here we wish to achieve two things simultaneously. On the one hand, we want our method to have a good approximation rate, that is, we want the approximations to satisfy $\|u - u_\mathcal{T}\|_\mathscr{V} \leq C(\dim \mathscr{V}_\mathcal{T})^{-\gamma} \|u\|_\mathscr{V}$ for all $\mathcal{T} \in \mathbb{T}$ with some large $\gamma > 0$ (and some constant $C > 0$). On the other hand, we want the number of operations required to calculate $u_\mathcal{T}$ from (1.4) to be small, so ideally of order $\mathcal{O}(\dim \mathscr{V}_\mathcal{T})$. Both these motives play a central role in this thesis.

In the *finite element method*, the trial spaces $\mathscr{V}_\mathcal{T}$ are constructed as spaces of (dis)continuous piecewise polynomials of fixed degree with respect to meshes $\mathcal{T}$ of the underlying domain. If the solution u is *smooth*, i.e., its derivatives of sufficiently high order are bounded in a suitable norm, then optimal approximation rates are often obtained by simply considering a family of trial spaces with respect to *quasi-uniform* meshes $\mathcal{T}$, i.e., meshes having a uniform mesh width. The situation changes drastically when the (unknown) solution u contains *singularities*, which may be induced by the geometry or the data. In such cases, the approximation rate offered by trial spaces with respect to quasi-uniform meshes can drop significantly, making the numerical scheme

converge slowly. Luckily, in many situations the approximation rate can be improved significantly by considering trial spaces with respect to meshes that are *locally refined* at these singularities.

Throughout this thesis we will focus on this latter situation, where the solution u contains singularities. Clearly, in terms of accuracy, using trial spaces adapted to the singularities is preferable. This comes at a price however, as the mathematics and implementation of such adaptive numerical methods is often more difficult than for the quasi-uniform case.

1.1.3 Solving the discretized system

An imported question is how to *efficiently* solve (1.4). To analyze this, we first reformulate the problem in coordinates. With $\Phi_{\mathcal{T}} := \{\phi_1, \dots, \phi_n\}$ being a basis for $\mathscr{V}_{\mathcal{T}}$, one infers that solving (1.4) for $u_{\mathcal{T}} = \boldsymbol{u}^\top \Phi_{\mathcal{T}}$, is equivalent to finding $\boldsymbol{u} \in \mathbb{R}^n$ that solves

$$\boldsymbol{A}_{\mathcal{T}} \boldsymbol{u} = \boldsymbol{f}, \tag{1.5}$$

with the *system matrix* $\boldsymbol{A}_{\mathcal{T}} \in \mathbb{R}^{n \times n}$ and right hand side $\boldsymbol{f} \in \mathbb{R}^n$ given by

$$\boldsymbol{A}_{\mathcal{T}} := (A_{\mathcal{T}} \Phi_{\mathcal{T}})(\Phi_{\mathcal{T}}) = [(A_{\mathcal{T}} \phi_j)(\phi_i)]_{ij}, \qquad \boldsymbol{f} := f(\Phi_{\mathcal{T}}) = [f(\phi_i)]_i.$$

The computational complexity of solving the above matrix-vector system using a direct method, e.g. using LU-factorization, is $\mathcal{O}(n^3)$. This is prohibitively expensive, as we are interested in constructing methods that run in optimal (linear) time. Another problem is that direct methods require the matrix $\boldsymbol{A}_{\mathcal{T}}$ to be available, which is not the case for *matrix-free* methods, where one has access to only the application of $\boldsymbol{A}_{\mathcal{T}}$. An example of such a matrix-free method will be given in this thesis, where we devise an algorithm to apply a system matrix $\boldsymbol{A}_{\mathcal{T}}$ in linear complexity, even though the matrix itself is not sparsely populated.

A first efficiency gain can be obtained by solving (1.5) approximately yielding some $\mathbb{R}^{\boldsymbol{n}} \ni \hat{\boldsymbol{u}} \approx \boldsymbol{u}$. For $\hat{u}_{\mathcal{T}} := \hat{\boldsymbol{u}}^\top \Phi_{\mathcal{T}}$ we have

$$\|u - \hat{u}_{\mathcal{T}}\|_{\mathscr{V}} \leq \underbrace{\|u - u_{\mathcal{T}}\|_{\mathscr{V}}}_{\text{discretization error}} + \underbrace{\|u_{\mathcal{T}} - \hat{u}_{\mathcal{T}}\|_{\mathscr{V}}}_{\text{algebraic error}},$$

so as long as the algebraic error is dominated by the discretization error, our approximation $\hat{u}_{\mathcal{T}}$ is still quasi-optimal. Such approximate solutions to (1.5) are given by *iterative solvers*, being methods that produce a sequence of vectors $(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots)$ converging to the solution $\boldsymbol{u}$.

If the matrix $\boldsymbol{A}_{\mathcal{T}}$ is symmetric and positive definite, as when the operator A is self-adjoint ($A = A'$) and coercive, then of particular interest is the *Conjugate Gradient* (CG) method. This iterative solver has favourable convergence properties and has minimal computation costs, i.e., the cost of a single iteration is dominated by the cost of a single application of $\boldsymbol{A}_{\mathcal{T}}$. Write $\rho(\cdot)$ for

the spectral radius of an operator, and denote the spectral condition number by $\kappa_S(\boldsymbol{A}_\mathcal{T}) = \rho(\boldsymbol{A}_\mathcal{T})\rho(\boldsymbol{A}_\mathcal{T}^{-1})$. Starting from an initial guess, the number of iterations required by CG to reduce the initial algebraic error by a factor ϵ is bounded by $\sqrt{\kappa_S(\boldsymbol{A})}\log(1/\epsilon)$.

Unfortunately, for standard finite (or boundary) element bases, the system matrices $\boldsymbol{A}_\mathcal{T}$ are (generally) ill-conditioned, in the sense that the condition number $\kappa_S(\boldsymbol{A}_\mathcal{T})$ increases for decreasing mesh width. This implies that the number of iterations required by CG to reach a certain accuracy increases upon mesh refinement.

One way of fixing this conditioning issue, is to make use of *preconditioning*: instead of solving (1.5), one considers $\boldsymbol{G}_\mathcal{T}\boldsymbol{A}_\mathcal{T}\boldsymbol{u} = \boldsymbol{G}_\mathcal{T}\boldsymbol{f}$ where $\boldsymbol{G}_\mathcal{T}$ is some preconditioner, i.e., an approximation of $\boldsymbol{A}_\mathcal{T}^{-1}$ that can be applied efficiently. The number of iterations required by CG applied to this preconditioned system to reduce the initial algebraic error by a factor ϵ, is bounded by $\sqrt{\kappa_S(\boldsymbol{G}_\mathcal{T}\boldsymbol{A}_\mathcal{T})}\log(1/\epsilon)$.

1.2 About Part I: Operator preconditioning

In the first part of this thesis we focus on constructing *uniformly optimal* preconditioners $\boldsymbol{G}_\mathcal{T}$ for the family $\mathcal{T} \in \mathbb{T}$. This means that we want the condition number of the preconditioned matrix $\kappa_S(\boldsymbol{G}_\mathcal{T}\boldsymbol{A}_\mathcal{T})$, and therefore also the number of iterations required by CG, to be bounded independently of the trial space $\mathscr{V}_\mathcal{T}$.

One technique for constructing such preconditioners is so-called *operator preconditioning* ([Hip06]). This approach hinges on the availability of an *opposite order* operator B, being a bounded and coercive linear map from $\mathscr{V}'$ to $\mathscr{V}$. On a continuous level we have that BA is a boundedly invertible map $\mathscr{V} \to \mathscr{V}$, suggesting that B may be used to construct a preconditioner for $A_\mathcal{T}\colon \mathscr{V}_\mathcal{T} \to \mathscr{V}_\mathcal{T}'$.

Let $\mathscr{W}_\mathcal{T} \subset \mathscr{W} := \mathscr{V}'$ be some finite dimensional subspace, and consider the discretized operator $B_\mathcal{T}\colon \mathscr{W}_\mathcal{T} \to \mathscr{W}_\mathcal{T}'$ given by $(B_\mathcal{T}u)(v) := (Bu)(v)\,(u, v \in \mathscr{W}_\mathcal{T})$. If one additionally has a boundedly invertible operator $D_\mathcal{T}\colon \mathscr{V}_\mathcal{T} \to \mathscr{W}_\mathcal{T}'$, then the composition $D_\mathcal{T}^{-1}B_\mathcal{T}(D_\mathcal{T}')^{-1}\colon \mathscr{V}_\mathcal{T}' \to \mathscr{V}_\mathcal{T}$ is a bounded and coercive mapping, and therefore serves as a preconditioner for $A_\mathcal{T}$. See also the diagram:

$$\begin{array}{ccc} \mathscr{V}_\mathcal{T} & \xrightarrow{A_\mathcal{T}} & \mathscr{V}_\mathcal{T}' \\ {\scriptstyle D_\mathcal{T}^{-1}}\uparrow & & \downarrow{\scriptstyle (D_\mathcal{T}')^{-1}} \\ \mathscr{W}_\mathcal{T}' & \xleftarrow[B_\mathcal{T}]{} & \mathscr{W}_\mathcal{T} \end{array}. \tag{1.6}$$

The typical example to keep in mind is where $A\colon H^{-\frac{1}{2}}(\Gamma) \to H^{\frac{1}{2}}(\Gamma)$ is the Single Layer operator and $B\colon H^{\frac{1}{2}}(\Gamma) \to H^{-\frac{1}{2}}(\Gamma)$ the Hypersingular operator from Example 1.1.2.

Now, for the construction of a suitable $D_{\mathcal{T}}$, we assume $\mathscr{H}$ to be some Hilbert space (we take $L_2(\Gamma)$ in the above example) for which we have a Gelfand triple

$$\mathscr{W} \hookrightarrow \mathscr{H} \simeq \mathscr{H}' \hookrightarrow \mathscr{W}'.$$

If the trial spaces satisfy $\mathscr{V}_{\mathcal{T}} \subset \mathscr{H}$, we can consider the operator $D_{\mathcal{T}} \colon \mathscr{V}_{\mathcal{T}} \to \mathscr{W}_{\mathcal{T}}'$ defined by $(D_{\mathcal{T}} v)(w) := \langle v, w \rangle_{\mathscr{H}} \quad (v \in \mathscr{V}_{\mathcal{T}}, w \in \mathscr{W}_{\mathcal{T}})$. This is a uniformly boundedly invertible operator if the subspaces $\mathscr{W}_{\mathcal{T}}$ satisfy

$$\dim \mathscr{W}_{\mathcal{T}} = \dim \mathscr{V}_{\mathcal{T}} \tag{1.7}$$

and

$$\inf_{\mathcal{T} \in \mathbb{T}} \inf_{0 \neq v \in \mathscr{V}_{\mathcal{T}}} \sup_{0 \neq w \in \mathscr{W}_{\mathcal{T}}} \frac{\langle v, w \rangle_{\mathscr{H}}}{\|v\|_{\mathscr{V}} \|w\|_{\mathscr{W}}} > 0. \tag{1.8}$$

Assume these constraints to be satisfied. By equipping $\mathscr{V}_{\mathcal{T}}$ and $\mathscr{W}_{\mathcal{T}}$ with bases $\Phi_{\mathcal{T}}$ and $\Psi_{\mathcal{T}}$, respectively, the matrix representation of the preconditioned system reads as

$$\boldsymbol{D}_{\mathcal{T}}^{-1} \boldsymbol{B}_{\mathcal{T}} \boldsymbol{D}_{\mathcal{T}}^{-\top} \boldsymbol{A}_{\mathcal{T}}, \tag{1.9}$$

with system matrices $\boldsymbol{A}_{\mathcal{T}} := (A_{\mathcal{T}} \Phi_{\mathcal{T}})(\Phi_{\mathcal{T}})$ and $\boldsymbol{B}_{\mathcal{T}} := (B_{\mathcal{T}} \Psi_{\mathcal{T}})(\Psi_{\mathcal{T}})$, and 'generalized mass matrix' $\boldsymbol{D}_{\mathcal{T}} := \langle \Phi_{\mathcal{T}}, \Psi_{\mathcal{T}} \rangle_{\mathscr{H}}$. The spectral condition number of this preconditioned matrix system equals that of $D_{\mathcal{T}}^{-1} B_{\mathcal{T}} (D_{\mathcal{T}}')^{-1} A_{\mathcal{T}}$, and is thus uniformly bounded.

The real challenge is constructing suitable subspaces $\mathscr{W}_{\mathcal{T}} \subset \mathscr{W}$ that satisfy (1.7) and (1.8). Moreover, we require a basis $\Psi_{\mathcal{T}}$ for $\mathscr{W}_{\mathcal{T}}$ that allows the preconditioner (1.9) to be applied efficiently.

Special attention has to be paid to the inverse matrix $\boldsymbol{D}_{\mathcal{T}}^{-1}$ appearing in the preconditioner. If the matrix $\boldsymbol{D}_{\mathcal{T}}$ is not diagonal, its inverse has to be approximated, and it can generally be expected that, in order to obtain a uniform preconditioner, the accuracy with which $\boldsymbol{D}_{\mathcal{T}}^{-1}$ has to be approximated increases with a decreasing minimal mesh-size. As a result, an application of $\boldsymbol{D}_{\mathcal{T}}^{-1}$ cannot be expected to execute in linear time.

In this thesis, we will propose suitable spaces $\mathscr{W}_{\mathcal{T}} \subset \mathscr{W}$, and circumvent this latter issue by constructing bases $\Psi_{\mathcal{T}}$ that are $\mathscr{H}$-orthogonal to $\Phi_{\mathcal{T}}$, making $\boldsymbol{D}_{\mathcal{T}}$ a diagonal matrix whose inverse can be exactly evaluated.

1.2.1 Contributions and Outline of Part I

The contents of Chapters 2–5 is essentially that of the following papers:

[SvV20a] R.P. Stevenson and R. van Venetië. Uniform preconditioners for problems of negative order. *Mathematics of Computation*, 89(322):645–674, 2020.

[SvV21a] R.P. Stevenson and R. van Venetië. Uniform preconditioners of linear complexity for problems of negative order. *Computational Methods in Applied Mathematics*, 21(2):469–478, 2021.

[SvV20b] R.P. Stevenson and R. van Venetië. Uniform preconditioners for problems of positive order. *Computers & Mathematics with Applications*, 79(12):3516–3530, 2020.

[SvV21b] R.P. Stevenson and R. van Venetië. Operator preconditioning: the simplest case. Submitted to *Applied Numerical Mathematics*, 2021.

On average, the authors contributed equally to these works.

Chapter 2 ([SvV20a])

For some domain (or manifold) Ω and $s \in [0,1]$, we consider the fractional Sobolev space $H^s(\Omega)$ (possibly with homogeneous Dirichlet boundary conditions incorporated) and its dual that we denote here by $H^{-s}(\Omega)$.

In this chapter we consider operator preconditioning for a bounded and coercive operator $A\colon H^{-s}(\Omega) \to H^s(\Omega)$, so of *negative* order $-2s$, discretized by a family of trial spaces $\mathscr{V}_\mathcal{T} \subset H^{-s}(\Omega)$ being discontinuous piecewise constants w.r.t. $\mathcal{T}$. We consider a family $\mathbb{T}$ of uniformly shape regular, possibly locally refined, meshes of Ω. In order to apply the aforementioned framework, we assume availability of a suitable opposite order operator $B\colon H^s(\Omega) \to H^{-s}(\Omega)$.

We propose a family of subspaces $\mathscr{W}_\mathcal{T} = \operatorname{span}\Psi_\mathcal{T} \subset H^s(\Omega)$ satisfying both (1.7) and (1.8). We achieve this by constructing $\Psi_\mathcal{T}$ as a collection that is $L_2(\Omega)$-biorthogonal to $\Phi_\mathcal{T}$, the (canonical) piecewise constant basis for $\mathscr{V}_\mathcal{T}$. As a consequence, the matrix $\boldsymbol{D}_\mathcal{T} := \langle \Phi_\mathcal{T}, \Psi_\mathcal{T}\rangle_{L_2(\Omega)}$ is *diagonal*, and thus its inverse, which appears in the preconditioner (1.9), can be evaluated exactly.

The functions $\Psi_\mathcal{T}$ are constructed in $\mathscr{S}_\mathcal{T} \oplus \mathscr{B}_\mathcal{T}$, where $\mathscr{S}_\mathcal{T}$ is the space of continuous piecewise linears w.r.t. $\mathcal{T}$, and $\mathscr{B}_\mathcal{T}$ is a space containing bubble functions. This allows us to construct a bounded and coercive operator $B_\mathcal{T}\colon \mathscr{W}_\mathcal{T} \to \mathscr{W}_\mathcal{T}'$ as the sum of $B^\mathscr{S}_\mathcal{T}\colon \mathscr{S}_\mathcal{T} \to \mathscr{S}_\mathcal{T}'$, being the opposite order operator B discretized on $\mathscr{S}_\mathcal{T} \subset H^s(\Omega)$, and an invertible diagonal scaling operator on the bubble space. Besides the cost of the discretized operator $B^\mathscr{S}_\mathcal{T}$, the cost of the resulting preconditioner scales linearly in $\dim \mathscr{V}_\mathcal{T}$.

Our approach has a few advantages over earlier proposals: it does not require the inverse of a non-diagonal matrix; it applies without any mildly grading assumption on the mesh; and it does not require a barycentric refinement of the mesh underlying the trial space. Furthermore, we will show that our approach extends to the general case where $\mathscr{V}_\mathcal{T}$ is chosen as a space of (dis)continuous piecewise polynomials of any order.

Chapter 3 ([SvV21a])

We continue with the setting from the previous chapter. In this chapter we construct a multi-level type operator that both fulfills the role of the opposite order operator B *and* can be applied in optimal *linear complexity*. For this construction, we require $\mathbb{T}$ to be a family of conforming partitions created by newest vertex bisection. Together with the results from the previous chapter, this provides uniformly optimal preconditioners for negative order operators A discretized on a family of possibly locally refined meshes, that can be applied in linear complexity.

Chapter 4 ([SvV20b])

In this chapter consider the operator preconditioning framework for *positive* order operators, that is, we switch the roles of A and B.

More precisely, we construct preconditioners for a bounded and coercive operator $A\colon H^s(\Omega) \to H^{-s}(\Omega)$, being of positive order $2s$, discretized by $\mathscr{V}_\mathcal{T} \subset H^s(\Omega)$ being continuous piecewise linears w.r.t $\mathcal{T}$. Again, the mesh family $\mathbb{T}$ is supposed to be uniformly shape regular, which allows for locally refined meshes. Assuming the availability of an opposite order operator $B\colon H^{-s}(\Omega) \to H^s(\Omega)$, we explore the operator preconditioning framework, and aim to get results similar to the setting studied in Chapter 2.

We introduce a family of subspaces $\mathscr{W}_\mathcal{T} \subset H^{-s}(\Omega)$ satisfying both (1.7) and (1.8). Similarly as before, we do this by constructing $\mathscr{W}_\mathcal{T}$ as the span of a collection $\Psi_\mathcal{T}$ that is $L_2(\Omega)$-biorthogonal to $\Phi_\mathcal{T}$, the Lagrange basis for $\mathscr{V}_\mathcal{T}$. Besides an easy proof of the inf-sup condition (1.8), this biorthogonality has the advantage that the matrix $\boldsymbol{D}_\mathcal{T} := \langle \Phi_\mathcal{T}, \Psi_\mathcal{T} \rangle_{L_2(\Omega)}$ is *diagonal*, and thus its inverse, which appears in the preconditioner (1.9), can be evaluated exactly.

Since $\mathscr{W}_\mathcal{T}$ is a non-standard discretization space, we wish to simplify the implementation of an operator $B_\mathcal{T}\colon \mathscr{W}_\mathcal{T} \to \mathscr{W}_\mathcal{T}'$, similarly as the construction given in Chapter 2. We achieve this by constructing $\mathscr{W}_\mathcal{T}$ as a subspace of $\mathscr{U}_\mathcal{T} \oplus \mathscr{B}_\mathcal{T}$, where $\mathscr{U}_\mathcal{T} \subset H^{-s}(\Omega)$ is the space of discontinuous piecewise constants w.r.t. $\mathcal{T}$, and $\mathscr{B}_\mathcal{T}$ is some bubble space for which the $H^{-s}(\Omega)$-norm is equivalent to a weighted $L_2(\Omega)$-norm. This allows us to construct a suitable $B_\mathcal{T}\colon \mathscr{W}_\mathcal{T} \to \mathscr{W}_\mathcal{T}'$ as the sum of $B_\mathcal{T}^\mathscr{U}\colon \mathscr{U}_\mathcal{T} \to \mathscr{U}_\mathcal{T}'$, being our opposite order operator discretized on $\mathscr{U}_\mathcal{T}$, and an invertible diagonal scaling operator on the bubble space. Besides the cost of the discretized operator $B_\mathcal{T}^\mathscr{U}$, the cost of the resulting preconditioner scales linearly in $\dim \mathscr{V}_\mathcal{T}$.

This construction has the same advantages as that of Chapter 2: it avoids the inverse of a non-diagonal matrix, it circumvents the need for a barycentric refinement of the mesh, it applies without any mildly grading assumption on the mesh, and it can be extended to higher order trial spaces.

Chapter 5 ([SvV21b])

By restricting the setting from the previous chapters, we are able to construct uniform preconditioners with an even simpler implementation. Consider some closed manifold (or domain) Ω and trial spaces $\mathscr{V}_\mathcal{T}$ that are *continuous* piecewise polynomials of some fixed degree w.r.t. $\mathcal{T}$. Let some bounded and coercive operators $A\colon H^{-s}(\Omega) \to H^s(\Omega)$ and $B\colon H^s(\Omega) \to H^{-s}(\Omega)$ be given, and consider their corresponding discretizations $A_\mathcal{T}\colon \mathscr{V}_\mathcal{T} \to \mathscr{V}_\mathcal{T}'$ and $B_\mathcal{T}\colon \mathscr{V}_\mathcal{T} \to \mathscr{V}_\mathcal{T}'$.

In this chapter we introduce a uniformly boundedly invertible operator $D_\mathcal{T}\colon \mathscr{V}_\mathcal{T} \to \mathscr{V}_\mathcal{T}'$, allowing us to take $\mathscr{W}_\mathcal{T}$ equal to $\mathscr{V}_\mathcal{T}$ in the operator preconditioning framework (1.6). The resulting preconditioned system $D_\mathcal{T}^{-1}B_\mathcal{T}(D_\mathcal{T}')^{-1}A_\mathcal{T}$ and $(D_\mathcal{T}')^{-1}A_\mathcal{T}D_\mathcal{T}^{-1}B_\mathcal{T}$ are uniformly boundedly invertible. Moreover, the matrix representation of $D_\mathcal{T}$ with respect to the Lagrange basis of $\mathscr{V}_\mathcal{T}$ is diagonal, making the implementation of this preconditioner surprisingly simple.

1.3 About Part II: Parabolic evolution equations

The second topic that will be discussed in this thesis is the (adaptive) numerical solution of parabolic evolution equations written in a simultaneous space-time variational formulation.

As an illustrative example, let us introduce the model problem of the *heat equation* with homogeneous Dirichlet boundary conditions. For some time interval $I := (0,T)$ and some spatial domain $\Omega \subset \mathbb{R}^d$, the heat equation reads as finding $u : I \times \Omega \to \mathbb{R}$ that solves

$$\begin{aligned} \partial_t u - \Delta_{\boldsymbol{x}} u &= g && \text{on } I \times \Omega, \\ u &= 0 && \text{on } I \times \partial\Omega, \\ u &= u_0 && \text{on } \{0\} \times \Omega, \end{aligned} \tag{1.10}$$

for some forcing function $g\colon I \times \Omega \to \mathbb{R}$ and initial data $u_0\colon \Omega \to \mathbb{R}$.

In order to apply our approximation scheme we first need to derive a weak formulation of the above differential equation. We multiply the first equation by a test function v that vanishes on $I \times \partial\Omega$, integrate over $I \times \Omega$, and apply integration by parts in space to find

$$(Bu)(v) := \int_{I\times\Omega} (\partial_t u)v + \nabla_{\boldsymbol{x}} u \cdot \nabla_{\boldsymbol{x}} v \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t = \int_{I\times\Omega} gv \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t.$$

To enforce the initial condition, we introduce the trace map $\gamma_0\colon u \mapsto u(0,\cdot)$, and test it against some additional test function w to find that

$$\int_\Omega (\gamma_0 u)w \,\mathrm{d}\boldsymbol{x} = \int_\Omega u_0 w \,\mathrm{d}\boldsymbol{x}.$$

Clearly, we must yet find suitable function spaces for u, v and w. Define $X := L_2(I; H_0^1(\Omega)) \cap H^1(I; H^{-1}(\Omega))$, $Y := L_2(I; H_0^1(\Omega))$, and $H := L_2(\Omega)$, assume that $g \in Y'$ and $u_0 \in H$, and denote $\mathcal{B} := \begin{bmatrix} B & \gamma_0 \end{bmatrix}^\top$. Finding $u \in X$ that solves

$$\mathcal{B}u = \begin{bmatrix} g \\ u_0 \end{bmatrix} \tag{1.11}$$

is then a well-posed variational formulation of (1.10), meaning that the operator $\mathcal{B}$ is a boundedly invertible map $X \to Y' \times H$ ([SS09]).

A difficulty of the operator equation (1.11) appears when we consider discretizations. As the function space on the trial side does not coincide with that of the test side, we cannot simply apply the Galerkin method like we did in (1.4). Suppose that we have a family of trial spaces $(X^\delta)_{\delta\in\Delta} \subset X$, the question is how to construct a family of test spaces $(Z^\delta)_{\delta\in\Delta} \subset Y \times H$ for which the discretized operator $\mathcal{B}^\delta \colon X^\delta \to {Z^\delta}'$, given by $(\mathcal{B}^\delta u)(v) := (\mathcal{B}u)(v)$ $((u, v) \in X^\delta \times Z^\delta)$, is uniformly boundedly invertible. For the latter to hold, it turns out that the pairs (X^δ, Z^δ) must satisfy $\dim X^\delta = \dim Z^\delta$ and $\inf_{\delta\in\Delta} \inf_{0\neq u\in X^\delta} \sup_{0\neq u\in Z^\delta} \frac{(\mathcal{B}u)(v)}{\|u\|_X \|v\|_{Y\times H}} > 0$; cf. (1.7)–(1.8). The construction of such a test space Z^δ is hard, and we do not proceed this way.

Instead, we note that an equivalent problem to (1.11) is to compute

$$u = \operatorname*{argmin}_{w\in X} \|Bw - g\|_{Y'}^2 + \|\gamma_0 w - u_0\|_H^2,$$

suggesting an approximation approach by restricting the minimization to some finite dimensional trial space $X^\delta \subset X$. Unfortunately this is not feasible in practice, because the Y'-norm cannot be computed. To resolve this, we consider some finite dimensional test space $Y^\delta \subset Y$ and replace the Y'-norm by the (computable) ${Y^\delta}'$-norm, yielding the approximation

$$u^\delta = \operatorname*{argmin}_{w\in X^\delta} \|Bw - g\|_{{Y^\delta}'}^2 + \|\gamma_0 w - u_0\|_H^2. \tag{1.12}$$

In [And13] it is shown that, if the family of pairs $(X^\delta, Y^\delta)_{\delta\in\Delta}$ satisfy

(1.13)
$$X^\delta \subseteq Y^\delta \quad (\delta \in \Delta) \quad \text{and} \quad \gamma_\Delta := \inf_{\delta\in\Delta} \inf_{0\neq w\in X^\delta} \sup_{0\neq v\in Y^\delta} \frac{(\partial_t w)(v)}{\|\partial_t w\|_{Y'} \|v\|_Y} > 0,$$

then the approximation u^δ satisfies $\|u - u^\delta\|_X \leq \gamma_\Delta^{-1} \inf_{w\in X^\delta} \|u - w\|_X$, making it is a quasi-best approximation to u from X^δ. The advantage of this approach is that Y^δ can be chosen larger in relation to X^δ, making it easier to satisfy the inf-sup condition.

1.3.1 Contributions and Outline of Part II

The contents of Chapters 6–9 is essentially that of the following papers:

[SvVW21] R.P. Stevenson, R. van Venetië, and J. Westerdiep. A wavelet-in-time, finite element-in-space adaptive method for parabolic evolution equations. Submitted to *Advances in Computational Mathematics*, 2021.

[vVW21b] R. van Venetië and J. Westerdiep. Efficient space-time adaptivity for parabolic evolution equations using wavelets in time and finite elements in space. Submitted to *Numerical Linear Algebra with Applications*, 2021.

[vVW21a] R. van Venetië and J. Westerdiep. A parallel algorithm for solving linear parabolic evolution equations. Accepted in *9th Parallel-in-Time Workshop*, 2021.

[GvV21] G. Gantner and R. van Venetië. Adaptive space-time BEM for the heat equation. Submitted to *Computers & Mathematics with Applications*, 2021.

On average, the authors contributed equally to these works.

Chapter 6 ([SvVW21])

For trial spaces X^δ that are *full* (or *sparse*) tensor products of finite element spaces in time and space, in [And13] it was shown how to construct corresponding test spaces $Y^\delta \subset Y$ such that (1.13) holds. Unfortunately, neither family allows for adaptive refinements both locally in time and space.

In this chapter we solve this issue by equipping X with a tensor product basis of a wavelet basis in time and a hierarchical finite element basis in space. We then construct X^δ as the span of a (finite) subset of this tensor product basis, and construct Y^δ of a similar type such that (1.13) holds, with the dimension of Y^δ being proportional to that of X^δ.

Using properties of the wavelets in time and applying multigrid preconditioners in space, we construct optimal preconditioners $K_X^\delta \colon X^{\delta\prime} \to X^\delta$ and $K_Y^\delta \colon Y^{\delta\prime} \to Y^\delta$, allowing to solve the discrete problem (1.12) efficiently.

We propose an adaptive algorithm using a standard solve-estimate-mark-refine loop. Let X^δ be the current trial space. In the solve step we find its corresponding approximation u^δ from (1.12). In the estimate step, we introduce a neighborhood $X^{\underline{\delta}} \supset X^\delta$ and evaluate the residual on a basis of $X^{\underline{\delta}} \setminus X^\delta$. In the mark step, we select the basis functions for which the residual is large. Finally, in the refine step, we build a new trial space $X^{\tilde{\delta}} \supset X^\delta$ containing all the marked basis functions. Under a saturation assumption, we prove that the adaptive loop produces an r-linearly converging sequence to the solution.

Chapter 7 ([vVW21b])

In this chapter we discuss an implementation of the aforementioned adaptive method in which every step is of linear complexity.

The downside of having bases for X^δ and Y^δ of wavelet-type is that the system matrices appearing in the implementation of (1.12) are not sparsely populated. By imposing a *double-tree* constraint on the index sets of X^δ and Y^δ, we are able to derive a matrix-free algorithm that can apply the system matrices in linear complexity.

In order to build an actual linear complexity implementation, we based our implementation on tree- and double-tree traversals. We conclude this chapter with extensive results that demonstrate the linear runtime of our code in practice.

Chapter 8 ([vVW21a])

One of the advantages of a simultaneous space-time solver over classical time-stepping approaches is that they are much better suited for a massively parallel implementation. In this chapter we investigate such a parallel algorithm.

We consider trial spaces X^δ being the tensor product of finite element spaces in time and space, and show how this tensor product assumption simplifies the implementation of (1.12). After introducing suitable preconditioners, we investigate the parallel complexity of the resulting algorithm. We illustrate our theoretical findings with massively time-parallel computations done in practice.

Chapter 9 ([GvV21])

Deviating from the previous setting, here we construct an adaptive space-time *boundary element method* for the heat equation. We consider the heat equation with homogeneous forcing data and prescribed Dirichlet data: for given initial condition $u_0 \colon \Omega \to \mathbb{R}$ and Dirichlet data $u_D \colon I \times \partial\Omega \to \mathbb{R}$, we seek u that solves

$$\begin{aligned} \partial_t u - \Delta_{\boldsymbol{x}} u &= 0 && \text{on } I \times \Omega, \\ u &= u_D && \text{on } I \times \partial\Omega, \\ u &= u_0 && \text{on } \{0\} \times \Omega. \end{aligned}$$

Since the fundamental solution for the heat equation is known, we can proceed similarly as in Example 1.1.2 to find an equivalent formulation of the problem by solving an integral equation on the lateral space-time boundary $I \times \partial\Omega$, see e.g. [AN87, Cos90]. In contrast to (1.11), this formulation is *coercive*.

In this chapter, we propose an a posteriori error estimator for the Galerkin approximation to this integral equation. We show that the estimator is a lower bound for the approximation error, and up to weighted L_2-terms, also an upper bound. In the numerical results, we let this error estimator drive an adaptive loop that allows for anisotropic refinement. We observe that this loop is able to effectively resolve singularities in time and space, and that it recovers the optimal error decay rate in all of our examples.

Part I

Operator preconditioning

2 Problems of negative order

2.1 Introduction

This chapter is about the construction of preconditioners for discretized boundedly invertible linear operators of negative order using the concept of 'operator preconditioning' ([Hip06]). The idea is to precondition the discretized operator by a discretized operator of opposite order. This is an appealing idea, but it turns out that in order to get a uniformly well-conditioned system, as well as a preconditioner that can be implemented efficiently, the second discretization has to be carefully chosen dependent on the first one.

For a Hilbert space $\mathscr{H}$, and a densely embedded reflexive Banach space $\mathscr{W} \hookrightarrow \mathscr{H}$, consider the Gelfand triple

$$\mathscr{W} \hookrightarrow \mathscr{H} \simeq \mathscr{H}' \hookrightarrow \mathscr{W}'.$$

For A being a boundedly invertible coercive linear operator $\mathscr{W}' \to \mathscr{W}$, and $\mathscr{V}_\mathcal{T} \subset \mathscr{H}$ being a finite dimensional subspace of $\mathscr{W}'$, let $(A_\mathcal{T} v)(\tilde{v}) := (Av)(\tilde{v})$ $(v, \tilde{v} \in \mathscr{V}_\mathcal{T})$. For B being a boundedly invertible coercive linear operator $\mathscr{W} \to \mathscr{W}'$, and $\mathscr{W}_\mathcal{T}$ being a finite dimensional subspace of $\mathscr{W}$, let $(B_\mathcal{T} w)(\tilde{w}) := (Bw)(\tilde{w})$ $(w, \tilde{w} \in \mathscr{W}_\mathcal{T})$.

A typical example is given by the case that for the boundary Γ of some domain, $\mathscr{H} = L_2(\Gamma)$, $\mathscr{W} = H^{\frac{1}{2}}(\Gamma)$, A is the single layer integral operator arising from the Laplacian, B is the corresponding hypersingular integral operator, $\mathcal{T}$ is a partition from an infinite collection of partitions $\mathbb{T}$, $\mathscr{V}_\mathcal{T}$ is a trial space of discontinuous piecewise polynomials w.r.t. $\mathcal{T}$, and $\mathscr{W}_\mathcal{T}$ is a suitable subspace of $\mathscr{W}$, which thus cannot be equal to $\mathscr{V}_\mathcal{T}$. Besides as boundary integral equations, coercive linear operators of order -1 also appear in various domain decomposition type methods in the equations for normal fluxes on interfaces.

Although less frequently, coercive linear operators of order -2 also appear in the literature (e.g. see [FH19]).

In order to precondition $A_\mathcal{T} : \mathscr{V}_\mathcal{T} \to \mathscr{V}_\mathcal{T}'$ with $B_\mathcal{T} : \mathscr{W}_\mathcal{T} \to \mathscr{W}_\mathcal{T}'$ we need to be able to 'identify' $\mathscr{V}_\mathcal{T}'$ with $\mathscr{W}_\mathcal{T}$, similar to the identification of $\mathscr{W}''$ with $\mathscr{W}$.

Let $\dim \mathscr{W}_\mathcal{T} = \dim \mathscr{V}_\mathcal{T}$ and

$$\inf_{\mathcal{T}\in\mathbb{T}} \inf_{0\neq v\in\mathscr{V}_\mathcal{T}} \sup_{0\neq w\in\mathscr{W}_\mathcal{T}} \frac{\langle v, w\rangle_{\mathscr{H}}}{\|v\|_{\mathscr{W}'}\|w\|_{\mathscr{W}}} > 0. \tag{2.1}$$

Then $D_\mathcal{T}$ defined by $(D_\mathcal{T} v)(w) := \langle v, w\rangle_{\mathscr{H}}$ ($v \in \mathscr{V}_\mathcal{T}$, $w \in \mathscr{W}_\mathcal{T}$) is a uniformly boundedly invertible linear map $\mathscr{V}_\mathcal{T} \to \mathscr{W}_\mathcal{T}'$, and so its adjoint $D_\mathcal{T}'$ is such a map $\mathscr{V}_\mathcal{T}' \to \mathscr{W}_\mathcal{T}$. We conclude that the preconditioned system $D_\mathcal{T}^{-1} B_\mathcal{T} (D_\mathcal{T}')^{-1} A_\mathcal{T}$ is uniformly boundedly invertible $\mathscr{V}_\mathcal{T} \to \mathscr{V}_\mathcal{T}$.

Equipping $\mathscr{V}_\mathcal{T}$ and $\mathscr{W}_\mathcal{T}$ with bases $\Xi_\mathcal{T}$ and $\Psi_\mathcal{T}$, respectively, the matrix representation of the preconditioned system reads as $\boldsymbol{D}_\mathcal{T}^{-1}\boldsymbol{B}_\mathcal{T}\boldsymbol{D}_\mathcal{T}^{-T}\boldsymbol{A}_\mathcal{T}$, with 'stiffness matrices' $\boldsymbol{A}_\mathcal{T} := (A_\mathcal{T}\Xi_\mathcal{T})(\Xi_\mathcal{T})$ and $\boldsymbol{B}_\mathcal{T} := (B_\mathcal{T}\Psi_\mathcal{T})(\Psi_\mathcal{T})$, and 'generalized mass matrix' $\boldsymbol{D}_\mathcal{T} := \langle \Xi_\mathcal{T}, \Psi_\mathcal{T}\rangle_{\mathscr{H}}$. Regardless of the choice of the bases, the spectral condition number of this matrix is equal to that of $D_\mathcal{T}^{-1} B_\mathcal{T} (D_\mathcal{T}')^{-1} A_\mathcal{T}$, and thus uniformly bounded.

After an earlier proposal from [Ste02], the currently commonly followed construction of a suitable pair $(\mathscr{V}_\mathcal{T}, \mathscr{W}_\mathcal{T})$ is the one from [BC07]. Here $\mathscr{V}_\mathcal{T}$ is the space of piecewise constants w.r.t. a partition $\mathcal{T}$ of a two-dimensional domain or manifold equipped with the usual basis $\Xi_\mathcal{T}$, and $\mathscr{W}_\mathcal{T}$, defined as the span of a collection $\Psi_\mathcal{T}$, is a subspace of the space of continuous piecewise linears w.r.t. a barycentric refinement of $\mathcal{T}$ constructed by subdividing each triangle into 6 subtriangles by connecting its vertices and midpoints with its barycenter. In [HUT16] the inf-inf-sup condition (2.1) was demonstrated for families of partitions including locally refined ones that satisfy a certain mildly-grading condition from [Ste03a].

A problem with the constructions from both [Ste02, BC07] is that the matrix $\boldsymbol{D}_\mathcal{T}$ is *not* diagonal, so that its inverse has to be approximated. Knowing that $\boldsymbol{D}_\mathcal{T}^{-1}\boldsymbol{B}_\mathcal{T}\boldsymbol{D}_\mathcal{T}^{-T}$ is *not* well-conditioned, because $\boldsymbol{A}_\mathcal{T}$ is not whereas their product is uniformly well-conditioned, the accuracy with which $\boldsymbol{D}_\mathcal{T}^{-1}$ has to be approximated such that it gives rise to a uniform preconditioner increases with an increasing (minimal) mesh-size.

2.1.1 Contributions

For the aforementioned $\mathscr{V}_\mathcal{T}$ and $\Xi_\mathcal{T}$, in this chapter a space $\mathscr{W}_\mathcal{T}$, given as the span of a collection $\Psi_\mathcal{T}$, will be constructed such that (2.1) is valid, and $\boldsymbol{D}_\mathcal{T} = \langle \Xi_\mathcal{T}, \Psi_\mathcal{T}\rangle_{\mathscr{H}}$ is *diagonal*. i.e., $\Xi_\mathcal{T}$ and $\Psi_\mathcal{T}$ are *biorthogonal*. Thanks to both $\Xi_\mathcal{T}$ and $\Psi_\mathcal{T}$ being 'locally supported', the corresponding biorthogonal projector onto $\mathscr{W}_\mathcal{T}$ is *local*, which allows to demonstrate the inf-inf-sup stability *without any mildly grading assumption* on the partitions.

Each function in $\Psi_\mathcal{T}$ equals a function from the space $\mathscr{S}^{0,1}_{\mathcal{T},0}$ of continuous piecewise linears[1] w.r.t. $\mathcal{T}$, plus a linear combination of 'bubble func-

[1]The subscript 0 in the notation $\mathscr{S}^{0,1}_{\mathcal{T},0}$ refers to possible boundary conditions that are incorporated.

tions' from a space denoted as $\mathscr{B}_\mathcal{T}$. Since the decomposition of $\mathscr{S}^{0,1}_{\mathcal{T},0} \oplus \mathscr{B}_\mathcal{T}$ into $\mathscr{S}^{0,1}_{\mathcal{T},0}$ and $\mathscr{B}_\mathcal{T}$ is stable w.r.t. the $\mathscr{W}$-norm, instead of simply defining $(B_\mathcal{T} w)(\tilde{w}) := (Bw)(\tilde{w})$ $(w, \tilde{w} \in \mathscr{W}_\mathcal{T})$, a suitable boundedly invertible linear operator $B_\mathcal{T} \colon \mathscr{W}_\mathcal{T} \to \mathscr{W}'_\mathcal{T}$ will be constructed from an invertible diagonal scaling on the bubble space and a boundedly invertible linear operator $B^\mathscr{S}_\mathcal{T} \colon \mathscr{S}^{0,1}_{\mathcal{T},0} \to (\mathscr{S}^{0,1}_{\mathcal{T},0})'$, e.g. $(B^\mathscr{S}_\mathcal{T} w)(\tilde{w}) := (Bw)(\tilde{w})$ with B the hypersingular operator. Other than in [Ste02, BC07], by this use of the stable decomposition there is no need to discretize the hypersingular operator on a refinement of $\mathcal{T}$. The whole approach relies on *existence* of bubble functions with certain properties, whereas these functions themselves do not enter the implementation.

The total cost of the resulting preconditioner is the sum of the cost of the application of $B^\mathscr{S}_\mathcal{T}$ plus cost that scales linearly in $\#\mathcal{T}$. For $\mathcal{T}$ being a uniform refinement of some initial coarse partition, a $B^\mathscr{S}_\mathcal{T}$ of multi-level type can be found whose cost scales linearly in $\#\mathcal{T}$ ([BPV00]). Such $B^\mathscr{S}_\mathcal{T}$ that also apply on locally refined partitions will be discussed in Chapter 3.

The construction of the biorthogonal collection $\Psi_\mathcal{T}$, and with that of the preconditioner, is based on a general principle. It applies in any space dimension, and, as we will see, it applies equally well when $\mathscr{V}_\mathcal{T}$ is the space of *continuous piecewise linears*. Higher order discretizations will be covered as well.

The construction applies equally well on manifolds. The coefficients of the functions from $\Psi_\mathcal{T}$ in terms of functions from $\mathscr{S}^{0,1}_{\mathcal{T},0}$ and the bubble functions are given as inner products between functions of $\mathscr{V}_\mathcal{T}$ and $\mathscr{S}^{0,1}_{\mathcal{T},0}$. Since in the manifold case, however, generally these inner products cannot be evaluated exactly, we present an alternative slightly modified construction in which the true L_2-inner product is replaced by a mesh-dependent one by an element-wise freezing of the Jacobian. It still yields a uniform preconditioner on general, possibly locally refined partitions, while the same explicit formula for the expansion coefficients of the functions of $\Psi_\mathcal{T}$ that was derived in the domain case, now also applies in the manifold case.

2.1.2 Notations

In this work, by $\lambda \lesssim \mu$ we will mean that λ can be bounded by a multiple of μ, independently of parameters which λ and μ may depend on, with the sole exception of the space dimension d, or in the manifold case, on the parametrization of the manifold that is used to define the finite element spaces on it. Obviously, $\lambda \gtrsim \mu$ is defined as $\mu \lesssim \lambda$, and $\lambda \eqsim \mu$ as $\lambda \lesssim \mu$ and $\lambda \gtrsim \mu$.

For normed linear spaces $\mathscr{Y}$ and $\mathscr{Z}$, in this work for convenience over $\mathbb{R}$, $\mathcal{L}(\mathscr{Y}, \mathscr{Z})$ will denote the space of bounded linear mappings $\mathscr{Y} \to \mathscr{Z}$ endowed with the operator norm $\|\cdot\|_{\mathcal{L}(\mathscr{Y},\mathscr{Z})}$. The subset of invertible operators in $\mathcal{L}(\mathscr{Y}, \mathscr{Z})$ with inverses in $\mathcal{L}(\mathscr{Z}, \mathscr{Y})$ will be denoted as $\mathcal{L}\mathrm{is}(\mathscr{Y}, \mathscr{Z})$. The *condition number* of a $C \in \mathcal{L}\mathrm{is}(\mathscr{Y}, \mathscr{Z})$ is defined as $\kappa_{\mathscr{Y},\mathscr{Z}}(C) := \|C\|_{\mathcal{L}(\mathscr{Y},\mathscr{Z})}\|C^{-1}\|_{\mathcal{L}(\mathscr{Z},\mathscr{Y})}$.

For $\mathscr{Y}$ a reflexive Banach space and $C \in \mathcal{L}(\mathscr{Y}, \mathscr{Y}')$ being *coercive*, i.e.,

$$\inf_{0\neq y\in\mathscr{Y}} \frac{(Cy)(y)}{\|y\|^2_{\mathscr{Y}}} > 0,$$

both C and $\Re(C) := \frac{1}{2}(C + C')$ are in $\mathcal{L}\mathrm{is}(\mathscr{Y}, \mathscr{Y}')$ with

$$\|\Re(C)\|_{\mathcal{L}(\mathscr{Y},\mathscr{Y}')} \leq \|C\|_{\mathcal{L}(\mathscr{Y},\mathscr{Y}')},$$
$$\|C^{-1}\|_{\mathcal{L}(\mathscr{Y}',\mathscr{Y})} \leq \|\Re(C)^{-1}\|_{\mathcal{L}(\mathscr{Y}',\mathscr{Y})} = \Big(\inf_{0\neq y\in\mathscr{Y}} \frac{(Cy)(y)}{\|y\|^2_{\mathscr{Y}}}\Big)^{-1}.$$

The subset of coercive operators in $\mathcal{L}\mathrm{is}(\mathscr{Y}, \mathscr{Y}')$ is denoted as $\mathcal{L}\mathrm{is}_c(\mathscr{Y}, \mathscr{Y}')$. If $C \in \mathcal{L}\mathrm{is}_c(\mathscr{Y}, \mathscr{Y}')$, then $C^{-1} \in \mathcal{L}\mathrm{is}_c(\mathscr{Y}', \mathscr{Y})$ and $\|\Re(C^{-1})^{-1}\|_{\mathcal{L}(\mathscr{Y},\mathscr{Y}')} \leq \|C\|^2_{\mathcal{L}(\mathscr{Y},\mathscr{Y}')}\|\Re(C)^{-1}\|_{\mathcal{L}(\mathscr{Y}',\mathscr{Y})}$.

Two countable collections $\Upsilon = (\upsilon_i)_i$ and $\tilde{\Upsilon} = (\tilde{\upsilon}_i)_i$ in a Hilbert space will be called *biorthogonal* when $\langle \Upsilon, \tilde{\Upsilon}\rangle = [\langle \upsilon_j, \tilde{\upsilon}_i\rangle]_{ij}$ is an *invertible diagonal* matrix, and *biorthonormal* when it is the *identity* matrix.

2.1.3 Organization

In Sect. 2.2 the general principles of operator preconditioning are recalled. In Sect. 2.3, it is applied to operators of negative order discretized with *discontinuous* piecewise constants, first in the domain- and then in the manifold-case. In Sect. 2.4, the same program is followed for trial spaces of *continuous* piecewise linears. In Sect. 2.5 the results from Sect. 2.3-2.4 will be extended to higher order finite element spaces. This will be done by both applying the operator preconditioning framework directly to the higher order spaces, and by using the preconditioner found for the lowest order case in a subspace correction approach. Finally, in Sect. 2.6 we report on some numerical results obtained with the new preconditioners, and compare them with those obtained with the preconditioner from [BC07, HUT16].

2.2 Operator preconditioning

The exposition in this section largely follows [Hip06, Sect. 2] closely. Let $\mathscr{V}$, $\mathscr{W}$ be reflexive Banach spaces. We will search a 'preconditioner' G for an $A \in \mathcal{L}\mathrm{is}(\mathscr{V}, \mathscr{V}')$, i.e. an operator $G \in \mathcal{L}\mathrm{is}(\mathscr{V}', \mathscr{V})$ (whose application is 'easy' compared to that of A^{-1}). It is often useful, e.g. for the application of Conjugate Gradients, when the preconditioner is coercive, i.e., being an operator in $\mathcal{L}\mathrm{is}_c(\mathscr{V}', \mathscr{V})$. The following result is easily verified.

Proposition 2.2.1. *If $B \in \mathcal{L}\mathrm{is}(\mathscr{W}, \mathscr{W}')$ and $D \in \mathcal{L}\mathrm{is}(\mathscr{V}, \mathscr{W}')$, then*

$$G := D^{-1}B(D')^{-1} \in \mathcal{L}\mathrm{is}(\mathscr{V}', \mathscr{V}),$$

and

$$\|G\|_{\mathcal{L}(\mathcal{V}',\mathcal{V})} \leq \|D^{-1}\|^2_{\mathcal{L}(\mathcal{W}',\mathcal{V})}\|B\|_{\mathcal{L}(\mathcal{W},\mathcal{W}')},$$
$$\|G^{-1}\|_{\mathcal{L}(\mathcal{V},\mathcal{V}')} \leq \|D\|^2_{\mathcal{L}(\mathcal{V},\mathcal{W}')}\|B^{-1}\|_{\mathcal{L}(\mathcal{W}',\mathcal{W})}.$$

If additionally $B \in \mathcal{L}\mathrm{is}_c(\mathcal{W},\mathcal{W}')$*, then* $G \in \mathcal{L}\mathrm{is}_c(\mathcal{V}',\mathcal{V})$*, and*

$$\|\Re(G)^{-1}\|_{\mathcal{L}(\mathcal{V},\mathcal{V}')} \leq \|D\|^2_{\mathcal{L}(\mathcal{V},\mathcal{W}')}\|\Re(B)^{-1}\|_{\mathcal{L}(\mathcal{W}',\mathcal{W})}.$$

Remark 2.2.2. We recall that by an application of the *closed range theorem*, $D \in \mathcal{L}(\mathcal{V},\mathcal{W}')$ is in $\mathcal{L}\mathrm{is}(\mathcal{V},\mathcal{W}')$ if and only if for all $w \in \mathcal{W}$ there exists a $v \in \mathcal{V}$ with $(Dv)(w) \neq 0$, and

$$0 < \inf_{0\neq v\in\mathcal{V}} \sup_{0\neq w\in\mathcal{W}} \frac{(Dv)(w)}{\|v\|_{\mathcal{V}}\|w\|_{\mathcal{W}}} \quad \Big(= \|D^{-1}\|^{-1}_{\mathcal{L}(\mathcal{W}',\mathcal{V})}\Big).$$

In particular we are interested in finding a preconditioner for an operator $A_{\mathcal{T}} \in \mathcal{L}\mathrm{is}(\mathcal{V}_{\mathcal{T}},\mathcal{V}'_{\mathcal{T}})$ of the form $G_{\mathcal{T}} = D_{\mathcal{T}}^{-1}B_{\mathcal{T}}(D'_{\mathcal{T}})^{-1}$, where $\mathcal{V}_{\mathcal{T}}$ is some *finite dimensional* (finite- or boundary element) space. In view of Proposition 2.2.1, for that goal we search some finite dimensional space $\mathcal{W}_{\mathcal{T}}$ with

$$\dim \mathcal{W}_{\mathcal{T}} = \dim \mathcal{V}_{\mathcal{T}}, \tag{2.2}$$

and operators $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}(\mathcal{W}_{\mathcal{T}},\mathcal{W}'_{\mathcal{T}})$ and $D_{\mathcal{T}} \in \mathcal{L}\mathrm{is}(\mathcal{V}_{\mathcal{T}},\mathcal{W}'_{\mathcal{T}})$.

A typical setting is that, for some reflexive Banach spaces $\mathcal{V}$ and $\mathcal{W}$, and operators $A \in \mathcal{L}\mathrm{is}_c(\mathcal{V},\mathcal{V}')$ and $B \in \mathcal{L}\mathrm{is}_c(\mathcal{W},\mathcal{W}')$, we have $\mathcal{V}_{\mathcal{T}} \subset \mathcal{V}$ (thus *equipped with* $\|\ \|_{\mathcal{V}}$), $(A_{\mathcal{T}}u)(v) := (Au)(v)$ $(u,v \in \mathcal{V}_{\mathcal{T}})$ and, for a suitable $\mathcal{W}_{\mathcal{T}} \subset \mathcal{W}$ (thus *equipped with* $\|\ \|_{\mathcal{W}}$), take $(B_{\mathcal{T}}w)(z) := (Bw)(z)$ $(w,z \in \mathcal{W}_{\mathcal{T}})$. In this case $A_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathcal{V}_{\mathcal{T}},\mathcal{V}'_{\mathcal{T}})$ and $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathcal{W}_{\mathcal{T}},\mathcal{W}'_{\mathcal{T}})$ with

$$\|A_{\mathcal{T}}\|_{\mathcal{L}(\mathcal{V}_{\mathcal{T}},\mathcal{V}'_{\mathcal{T}})} \leq \|A\|_{\mathcal{L}(\mathcal{V},\mathcal{V}')},\ \|\Re(A_{\mathcal{T}})^{-1}\|_{\mathcal{L}(\mathcal{V}'_{\mathcal{T}},\mathcal{V}_{\mathcal{T}})} \leq \|\Re(A)^{-1}\|_{\mathcal{L}(\mathcal{V}',\mathcal{V})},$$
$$\|B_{\mathcal{T}}\|_{\mathcal{L}(\mathcal{W}_{\mathcal{T}},\mathcal{W}'_{\mathcal{T}})} \leq \|B\|_{\mathcal{L}(\mathcal{W},\mathcal{W}')},\ \|\Re(B_{\mathcal{T}})^{-1}\|_{\mathcal{L}(\mathcal{W}'_{\mathcal{T}},\mathcal{W}_{\mathcal{T}})} \leq \|\Re(B)^{-1}\|_{\mathcal{L}(\mathcal{W}',\mathcal{W})}.$$

An obvious construction of a suitable $D_{\mathcal{T}}$ is discussed in the next proposition.

Proposition 2.2.3 (Fortin projector ([For77]))**.** *For some* $D \in \mathcal{L}\mathrm{is}(\mathcal{V},\mathcal{W}')$*, let* $D_{\mathcal{T}} \in \mathcal{L}(\mathcal{V}_{\mathcal{T}},\mathcal{W}'_{\mathcal{T}})$ *be defined by* $(D_{\mathcal{T}}v)(w) := (Dv)(w)$*. Then*

$$\|D_{\mathcal{T}}\|_{\mathcal{L}(\mathcal{V}_{\mathcal{T}},\mathcal{W}'_{\mathcal{T}})} \leq \|D\|_{\mathcal{L}(\mathcal{V},\mathcal{W}')}.$$

Assuming (2.2)*, additionally one has* $D_{\mathcal{T}} \in \mathcal{L}\mathrm{is}(\mathcal{V}_{\mathcal{T}},\mathcal{W}'_{\mathcal{T}})$ if*, and for* $\mathcal{W}$ *being a Hilbert space,* only if *there exists a projector* $P_{\mathcal{T}} \in \mathcal{L}(\mathcal{W},\mathcal{W})$ *onto* $\mathcal{W}_{\mathcal{T}}$ *with* $(D\mathcal{V}_{\mathcal{T}})((\mathrm{Id} - P_{\mathcal{T}})\mathcal{W}) = 0$*, in which case*

$$\|D_{\mathcal{T}}^{-1}\|_{\mathcal{L}(\mathcal{W}'_{\mathcal{T}},\mathcal{V}_{\mathcal{T}})} \leq \|P_{\mathcal{T}}\|_{\mathcal{L}(\mathcal{W},\mathcal{W})}\|D^{-1}\|_{\mathcal{L}(\mathcal{W}',\mathcal{V})}. \tag{2.3}$$

Proof. The first statement is obvious. Now let us assume existence of a (Fortin) projector $P_{\mathcal{T}}$. Then for $v_{\mathcal{T}} \in \mathscr{V}_{\mathcal{T}}$,

$$\begin{aligned}\|D^{-1}\|^{-1}_{\mathcal{L}(\mathscr{W}',\mathscr{V})}\|v_{\mathcal{T}}\|_{\mathscr{V}} &\leq \sup_{0\neq w\in\mathscr{W}} \frac{(Dv_{\mathcal{T}})(w)}{\|w\|_{\mathscr{W}}} = \sup_{0\neq w\in\mathscr{W}} \frac{(Dv_{\mathcal{T}})(P_{\mathcal{T}}w)}{\|w\|_{\mathscr{W}}} \\ &\leq \|P_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{W},\mathscr{W})} \sup_{0\neq w_{\mathcal{T}}\in\mathscr{W}_{\mathcal{T}}} \frac{(Dv_{\mathcal{T}})(w_{\mathcal{T}})}{\|w_{\mathcal{T}}\|_{\mathscr{W}}},\end{aligned}$$

which together with Remark 2.2.2 and (2.2) shows that $D_{\mathcal{T}} \in \mathcal{L}\mathrm{is}(\mathscr{V}_{\mathcal{T}}, \mathscr{W}'_{\mathcal{T}})$, in particular (2.3).

Conversely (cf. [Bra01, Remark 4.9]), assume $D_{\mathcal{T}} \in \mathcal{L}\mathrm{is}(\mathscr{V}_{\mathcal{T}}, \mathscr{W}'_{\mathcal{T}})$, and let $\mathscr{W}$ be a Hilbert space. Then given $w \in \mathscr{W}$, let $w_{\mathcal{T}}$ be the first component of the solution $(w_{\mathcal{T}}, v_{\mathcal{T}}) \in \mathscr{W}_{\mathcal{T}} \times \mathscr{V}_{\mathcal{T}}$ of the well-posed saddle point problem

$$\begin{aligned}\langle w_{\mathcal{T}}, z_{\mathcal{T}}\rangle_{\mathscr{W}} + (D_{\mathcal{T}}v_{\mathcal{T}})(z_{\mathcal{T}}) &= \langle w, z_{\mathcal{T}}\rangle_{\mathscr{W}} && (z_{\mathcal{T}} \in \mathscr{W}_{\mathcal{T}}),\\ (D_{\mathcal{T}}u_{\mathcal{T}})(w_{\mathcal{T}}) &= (D_{\mathcal{T}}u_{\mathcal{T}})(w) && (u_{\mathcal{T}} \in \mathscr{V}_{\mathcal{T}}).\end{aligned}$$

Then $P_{\mathcal{T}} := w \mapsto w_{\mathcal{T}}$ is a valid Fortin projector. □

In applications, one usually has a *family* of spaces $\mathscr{V}_{\mathcal{T}}$ and aims at a *uniform* preconditioner $G_{\mathcal{T}}$. In the setting of Proposition 2.2.3 it means that one searches a Fortin projector $P_{\mathcal{T}}$ such that $\|P_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{W},\mathscr{W})}$ is *uniformly* bounded.

2.2.1 Implementation

Given a finite collection $\Upsilon = \{\upsilon\}$ in a linear space, we set the *synthesis operator*

$$\mathcal{F}_\Upsilon : \mathbb{R}^{\#\Upsilon} \to \operatorname{span}\Upsilon \colon \mathbf{c} \mapsto \mathbf{c}^\top\Upsilon := \sum_{\upsilon\in\Upsilon} c_\upsilon \upsilon.$$

Equipping $\mathbb{R}^{\#\Upsilon}$ with the Euclidean scalar product $\langle\,,\,\rangle$, and identifying $(\mathbb{R}^{\#\Upsilon})'$ with $\mathbb{R}^{\#\Upsilon}$ using the corresponding Riesz map, we infer that the adjoint of $\mathcal{F}_\Upsilon$, known as the *analysis operator*, satisfies

$$\mathcal{F}'_\Upsilon : (\operatorname{span}\Upsilon)' \to \mathbb{R}^{\#\Upsilon} : f \mapsto f(\Upsilon) := [f(\upsilon)]_{\upsilon\in\Upsilon}.$$

A collection Υ is a *basis* for its span when $\mathcal{F}_\Upsilon \in \mathcal{L}\mathrm{is}(\mathbb{R}^{\#\Upsilon}, \operatorname{span}\Upsilon)$ (and so $\mathcal{F}'_\Upsilon \in \mathcal{L}\mathrm{is}((\operatorname{span}\Upsilon)', \mathbb{R}^{\#\Upsilon})$.)

Now let $\Xi_{\mathcal{T}} = \{\xi\}$ and $\Psi_{\mathcal{T}} = \{\psi\}$ be bases for $\mathscr{V}_{\mathcal{T}}$ and $\mathscr{W}_{\mathcal{T}}$, respectively. Then in coordinates the preconditioned system reads as

$$\mathcal{F}^{-1}_{\Xi_{\mathcal{T}}} G_{\mathcal{T}} A_{\mathcal{T}} \mathcal{F}_{\Xi_{\mathcal{T}}} = \boldsymbol{G}_{\mathcal{T}}\boldsymbol{A}_{\mathcal{T}} := \boldsymbol{D}_{\mathcal{T}}^{-1}\boldsymbol{B}_{\mathcal{T}}\boldsymbol{D}_{\mathcal{T}}^{-\top}\boldsymbol{A}_{\mathcal{T}},$$

where

$$\boldsymbol{A}_{\mathcal{T}} := \mathcal{F}'_{\Xi_{\mathcal{T}}} A_{\mathcal{T}} \mathcal{F}_{\Xi_{\mathcal{T}}}, \quad \boldsymbol{B}_{\mathcal{T}} := \mathcal{F}'_{\Psi_{\mathcal{T}}} B_{\mathcal{T}} \mathcal{F}_{\Psi_{\mathcal{T}}}, \quad \boldsymbol{D}_{\mathcal{T}} := \mathcal{F}'_{\Psi_{\mathcal{T}}} D_{\mathcal{T}} \mathcal{F}_{\Xi_{\mathcal{T}}}.$$

By identifying a map in $\mathcal{L}(\mathbb{R}^{\#\Xi_\mathcal{T}}, \mathbb{R}^{\#\Xi_\mathcal{T}})$ with a $\#\Xi_\mathcal{T} \times \#\Xi_\mathcal{T}$ matrix by equipping $\mathbb{R}^{\#\Xi_\mathcal{T}}$ with the canonical basis $\{\boldsymbol{e}_i\}$, and by enumerating the elements of $\Xi_\mathcal{T}$ one has

$$(\boldsymbol{A}_\mathcal{T})_{ij} = \langle \mathcal{F}'_{\Xi_\mathcal{T}} A_\mathcal{T} \mathcal{F}_{\Xi_\mathcal{T}} \boldsymbol{e}_j, \boldsymbol{e}_i \rangle = (A_\mathcal{T} \mathcal{F}_{\Xi_\mathcal{T}} \boldsymbol{e}_j)(\mathcal{F}_{\Xi_\mathcal{T}} \boldsymbol{e}_i) = (A_\mathcal{T} \xi_j)(\xi_i),$$

and similarly,

$$(\boldsymbol{B}_\mathcal{T})_{ij} = (B_\mathcal{T} \psi_j)(\psi_i), \quad (\boldsymbol{D}_\mathcal{T})_{ij} = (D_\mathcal{T} \xi_j)(\psi_i).$$

Preferably $\boldsymbol{D}_\mathcal{T}$ is such that its inverse can be applied in linear complexity, as is the case when $\boldsymbol{D}_\mathcal{T}$ is *diagonal*.

Remark 2.2.4. Using $\sigma(\,)$ and $\rho(\,)$ to denote the spectrum and spectral radius of an operator, clearly $\sigma(\boldsymbol{G}_\mathcal{T}\boldsymbol{A}_\mathcal{T}) = \sigma(G_\mathcal{T} A_\mathcal{T})$. So for the spectral condition number we have

$$\kappa_S(\boldsymbol{G}_\mathcal{T}\boldsymbol{A}_\mathcal{T}) := \rho(\boldsymbol{G}_\mathcal{T}\boldsymbol{A}_\mathcal{T})\rho((\boldsymbol{G}_\mathcal{T}\boldsymbol{A}_\mathcal{T})^{-1}) \leq \kappa_{\mathscr{V}_\mathcal{T}, \mathscr{V}_\mathcal{T}}(G_\mathcal{T} A_\mathcal{T}),$$

which thus holds true *independently* of the choice of the basis $\Xi_\mathcal{T}$ for $\mathscr{V}_\mathcal{T}$. Furthermore, in view of an application of Conjugate Gradients, if $A_\mathcal{T}$ and $B_\mathcal{T}$ are coercive and *self-adjoint*, then $\boldsymbol{A}_\mathcal{T}$ and $\boldsymbol{G}_\mathcal{T}$ are positive definite and symmetric. Equipping $\mathbb{R}^{\dim \mathscr{V}_\mathcal{T}}$ with $\|\!|\cdot|\!\| := \|(\boldsymbol{G}_\mathcal{T})^{-\frac{1}{2}} \cdot \|$ or $\|\!|\cdot|\!\| := \|(\boldsymbol{A}_\mathcal{T})^{\frac{1}{2}} \cdot \|$, in that case we have

$$\kappa_{(\mathbb{R}^{\dim \mathscr{V}_\mathcal{T}}, \|\!|\cdot|\!\|), (\mathbb{R}^{\dim \mathscr{V}_\mathcal{T}}, \|\!|\cdot|\!\|)}(\boldsymbol{G}_\mathcal{T}\boldsymbol{A}_\mathcal{T}) = \kappa_S(\boldsymbol{G}_\mathcal{T}\boldsymbol{A}_\mathcal{T}).$$

2.3 Piecewise constant discretization space

For a bounded polytopal domain $\Omega \subset \mathbb{R}^d$, a measurable, closed, possibly empty $\gamma \subset \partial\Omega$, and an $s \in [0, 1]$, we take

$$\mathscr{W} := [L_2(\Omega), H^1_{0,\gamma}(\Omega)]_{s,2}, \quad \mathscr{V} := \mathscr{W}',$$

where $H^1_{0,\gamma}(\Omega)$ is the closure in $H^1(\Omega)$ of the $C^\infty(\Omega) \cap H^1(\Omega)$ functions that vanish at γ.[2] The role of $D \in \mathcal{L}\text{is}(\mathscr{V}, \mathscr{W}')$ in Proposition 2.2.3 is going to be played by the unique extension to $\mathscr{V} \times \mathscr{W}$ of the duality pairing

$$(Dv)(w) := \langle v, w \rangle_{L_2(\Omega)},$$

which satisfies $\|D\|_{\mathcal{L}(\mathscr{V}, \mathscr{W}')} = \|D^{-1}\|_{\mathcal{L}(\mathscr{W}', \mathscr{V})} = 1$.

Let $(\mathcal{T})_{\mathcal{T} \in \mathbb{T}}$ be a family of *conforming* partitions of Ω into (open) *uniformly shape regular* d-simplices, where we assume that γ is the (possibly empty) union of $(d-1)$-faces of $T \in \mathcal{T}$. Thanks to the conformity and the uniform shape

[2] In the domain case, it is easy to generalize the results to Sobolev spaces with smoothness index $s \in [0, \frac{3}{2})$, or even to $s \in (-\frac{1}{2}, \frac{3}{2})$.

regularity, for $d > 1$ we know that neighbouring $T, T' \in \mathcal{T}$, i.e. $\overline{T} \cap \overline{T'} \neq \emptyset$, have uniformly comparable sizes. For $d = 1$, we impose this uniform '*K-mesh property*' explicitly.[3]

For $\mathcal{T} \in \mathbb{T}$, we define $N^0_\mathcal{T}$ as the set of vertices of $\mathcal{T}$ that are not on γ, and for $\nu \in N^0_\mathcal{T}$ we set its *valence*

$$d_{\mathcal{T},\nu} := \#\{T \in \mathcal{T} \colon \nu \in \overline{T}\}.$$

For $T \in \mathcal{T}$, and with N_T denoting the set of its vertices, we set $N^0_{\mathcal{T},T} := N^0_\mathcal{T} \cap N_T$, and define $h_T := |T|^{1/d}$.

We take

$$\mathscr{V}_\mathcal{T} = \mathscr{S}^{-1,0}_\mathcal{T} := \{u \in L_2(\Omega) \colon u|_T \in \mathcal{P}_0 \, (T \in \mathcal{T})\} \subset \mathscr{V},$$

and, as a first ingredient in the forthcoming construction of a suitable $\mathscr{W}_\mathcal{T}$, define the space of continuous piecewise linears, zero on γ, by

$$\mathscr{S}^{0,1}_{\mathcal{T},0} := \{u \in H^1_{0,\gamma}(\Omega) \colon u|_T \in \mathcal{P}_1 \, (T \in \mathcal{T})\},$$

equipped with the usual bases

$$\Xi_\mathcal{T} = \{\xi_T \colon T \in \mathcal{T}\}, \quad \Phi_\mathcal{T} = \{\phi_{\mathcal{T},\nu} \colon \nu \in N^0_\mathcal{T}\}, \tag{2.4}$$

respectively, defined by

$$\xi_T := \begin{cases} 1 & \text{on } T, \\ 0 & \text{on } \Omega \setminus T, \end{cases} \qquad \phi_{\mathcal{T},\nu}(\nu') = \delta_{\nu,\nu'} \quad (\nu, \nu' \in N^0_\mathcal{T}). \tag{2.5}$$

2.3.1 Construction of $\mathscr{W}_\mathcal{T}$ and $\boldsymbol{D}_\mathcal{T}$.

Aiming at the construction of a (uniform) preconditioner $G_\mathcal{T} \in \mathcal{L}\mathrm{is}_c(\mathscr{V}'_\mathcal{T}, \mathscr{V}_\mathcal{T})$ using the framework of operator preconditioning, we are going to construct a collection $\Psi_\mathcal{T} \subset H^1_{0,\gamma}(\Omega)$ that is biorthogonal to $\Xi_\mathcal{T}$, whose elements are 'locally supported', and for which

$$\mathscr{W}_\mathcal{T} := \operatorname{span} \Psi_\mathcal{T} \subset \mathscr{W}$$

has an 'approximation property'. These three properties of $\Psi_\mathcal{T}$ will allow us to construct a suitable Fortin projector, and they will give rise to a matrix $\boldsymbol{D}_\mathcal{T}$ that is *diagonal*.

The construction of $\Psi_\mathcal{T}$ builds on two collections $\Theta_\mathcal{T}$ and $\Sigma_\mathcal{T}$ of 'locally supported' functions in $H^1_{0,\gamma}(\Omega)$ whose cardinalities are equal to that of $\Xi_\mathcal{T}$,

[3]For our convenience, throughout this chapter we consider trial spaces w.r.t. conforming partitions into uniformly shape regular d-simplices. It will however become clear that families of non-conforming partitions into uniformly shape regular d-simplices or hyperrectangles that satisfy a uniform K-mesh property can be dealt with as well.

the first being biorthogonal to $\Xi_{\mathcal{T}}$, and the second whose span has an 'approximation property' and is inside $\mathscr{S}^{0,1}_{\mathcal{T},0}$.

Let $\Theta_{\mathcal{T}} = \{\theta_T \colon T \in \mathcal{T}\} \subset H^1_{0,\gamma}(\Omega)$ be such that $\theta_T \geq 0$, $\operatorname{supp} \theta_T \subset \overline{T}$,

$$\langle \theta_T, \xi_{T'} \rangle_{L_2(\Omega)} \eqsim \delta_{TT'} \|\theta_T\|_{L_2(\Omega)} \|\xi_{T'}\|_{L_2(\Omega)}, \quad (T, T' \in \mathcal{T}), \tag{2.6}$$

and, for convenience only, that is scaled such that

$$\langle \theta_T, \xi_T \rangle_{L_2(\Omega)} = |T|. \tag{2.7}$$

One obvious possible construction of such $\Theta_{\mathcal{T}}$ is to take θ_T to be the 'bubble function' defined by $\theta_T(x) = \frac{(2d+1)!}{d} \prod_{i=1}^{d+1} \lambda_i(x)$ for $x \in T$, and zero elsewhere, where $(\lambda_1(x), \cdots, \lambda_{d+1}(x))$ are the barycentric coordinates of x w.r.t. T (see e.g. [VS18] for (2.7)). Two forthcoming (harmless) conditions (2.26) and (2.27) on $\Theta_{\mathcal{T}}$ will be satisfied as well by the above specification of θ_T.

Another, equally suited construction is, after making a uniform barycentric refinement of $\mathcal{T}$, to take θ_T as a the continuous piecewise linear hat function associated to the barycenter of T, multiplied by a factor $d+1$.

We emphasize that the resulting preconditioner will *not* depend on the actual construction of $\Theta_{\mathcal{T}}$, but that only *existence* of a collection with the aforementioned properties is relevant.

Defining $\Sigma_{\mathcal{T}} = \{\sigma_{\mathcal{T},T} \colon T \in \mathcal{T}\} \subset \mathscr{S}^{0,1}_{\mathcal{T},0}$ by

$$\sigma_{\mathcal{T},T} := \sum_{\nu \in N^0_{\mathcal{T},T}} d^{-1}_{\mathcal{T},\nu} \phi_{\mathcal{T},\nu},$$

we have

$$\sum_{T \in \mathcal{T}} \sigma_{\mathcal{T},T} = \sum_{\nu \in N^0_{\mathcal{T}}} \phi_{\mathcal{T},\nu},$$

being equal to the constant function $\mathbb{1}$ on $\Omega \setminus \cup_{\{T \in \mathcal{T} \colon \overline{T} \cap \gamma \neq \emptyset\}} \overline{T}$, which yields the aforementioned 'approximation property' (cf. footnote 4).

We now define

$$\Psi_{\mathcal{T}} := \{\psi_{\mathcal{T},T} \colon T \in \mathcal{T}\} \subset \mathscr{S}^{0,1}_{\mathcal{T},0} \oplus \operatorname{span} \Theta_{\mathcal{T}},$$

by

$$\psi_{\mathcal{T},T} := \sigma_{\mathcal{T},T} + \frac{\langle \mathbb{1} - \sigma_{\mathcal{T},T}, \xi_T \rangle_{L_2(\Omega)}}{\langle \theta_T, \xi_T \rangle_{L_2(\Omega)}} \theta_T - \sum_{T' \in \mathcal{T} \setminus \{T\}} \frac{\langle \sigma_{\mathcal{T},T}, \xi_{T'} \rangle_{L_2(\Omega)}}{\langle \theta_{T'}, \xi_{T'} \rangle_{L_2(\Omega)}} \theta_{T'}, \tag{2.8}$$

The third term at the right-hand side corrects $\sigma_{\mathcal{T},T}$ such that it becomes orthogonal to $\xi_{T'}$ for $T' \neq T$, whereas the second term ensures that the $\psi_{\mathcal{T},T}$ sum up to $\mathbb{1}$, possibly except on a strip along the Dirichlet boundary:

Lemma 2.3.1. *It holds that*

$$\sum_{T\in\mathcal{T}} \psi_{\mathcal{T},T} = \sum_{T\in\mathcal{T}} \sigma_{\mathcal{T},T} + \sum_{T\in\mathcal{T}} \frac{\langle \mathbb{1} - \sum_{T'\in\mathcal{T}} \sigma_{\mathcal{T},T'}, \xi_T\rangle_{L_2(\Omega)}}{\langle \theta_T, \xi_T\rangle_{L_2(\Omega)}}\theta_T, \tag{2.9}$$

and

$$\langle \Xi_{\mathcal{T}}, \Psi_{\mathcal{T}}\rangle_{L_2(\Omega)} = \operatorname{diag}\{\langle \mathbb{1}, \xi_T\rangle_{L_2(\Omega)} : T\in\mathcal{T}\}. \tag{2.10}$$

Proof. Writing $\mathbb{1} - \sigma_{\mathcal{T},T} = \sum_{T'\in\mathcal{T}\setminus\{T\}} \sigma_{\mathcal{T},T'} + (\mathbb{1} - \sum_{T'\in\mathcal{T}} \sigma_{\mathcal{T},T'})$, (2.9) follows from (2.8) by using that

$$\sum_{T\in\mathcal{T}} \sum_{T'\in\mathcal{T}\setminus\{T\}} \frac{\langle \sigma_{\mathcal{T},T'}, \xi_T\rangle_{L_2(\Omega)}}{\langle \theta_T, \xi_T\rangle_{L_2(\Omega)}}\theta_T - \sum_{T\in\mathcal{T}} \sum_{T'\in\mathcal{T}\setminus\{T\}} \frac{\langle \sigma_{\mathcal{T},T}, \xi_{T'}\rangle_{L_2(\Omega)}}{\langle \theta_{T'}, \xi_{T'}\rangle_{L_2(\Omega)}}\theta_{T'} = 0.$$

The biorthonormality of $\Xi_{\mathcal{T}}$ and $\{\theta_T/\langle \theta_T, \xi_T\rangle_{L_2(\Omega)} : T\in\mathcal{T}\}$ shows (2.10). □

By expanding $\sigma_{\mathcal{T},T}$ in terms of the nodal basis, and by using $\int_T \phi_{\mathcal{T},\nu}\,dx = \frac{|T|}{d+1}$ and the normalization (2.7), we arrive at the explicit expression

$$\psi_{\mathcal{T},T} := \sum_{\nu\in N^0_{\mathcal{T},T}} d^{-1}_{\mathcal{T},\nu}\phi_{\mathcal{T},\nu} + \big(1 - \tfrac{1}{d+1}\sum_{\nu\in N^0_{\mathcal{T},T}} d^{-1}_{\mathcal{T},\nu}\big)\theta_T - \sum_{T'\in\mathcal{T}\setminus\{T\}} \big(\tfrac{1}{d+1}\sum_{\nu\in N^0_{\mathcal{T},T}\cap N^0_{\mathcal{T},T'}} d^{-1}_{\mathcal{T},\nu}\big)\theta_{T'}, \tag{2.11}$$

see Figure 2.1 for an illustration.

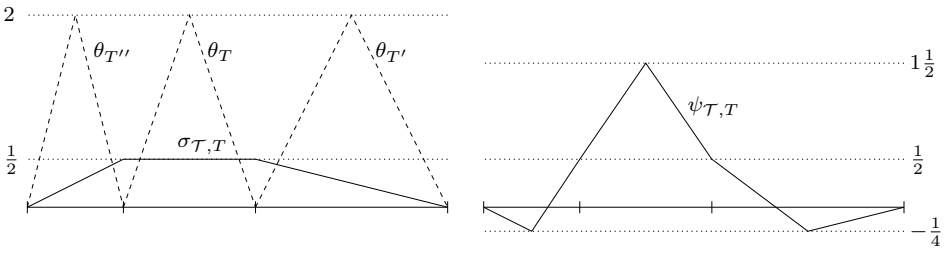


FIGURE 2.1. $\psi_{\mathcal{T},T}$ in one dimension (with bubbles constructed using a barycentric refinement).

As a consequence of $\langle \Xi_{\mathcal{T}}, \Psi_{\mathcal{T}}\rangle_{L_2(\Omega)}$ being invertible, the biorthogonal 'Fortin' projector $P_{\mathcal{T}} : L_2(\Omega) \to H^1_{0,\gamma}(\Omega)$ with $\operatorname{ran} P_{\mathcal{T}} = \mathscr{W}_{\mathcal{T}}$ and $\operatorname{ran}(\mathrm{Id} - P_{\mathcal{T}}) = \mathscr{V}_{\mathcal{T}}^{\perp_{L_2(\Omega)}}$ exists, and is, thanks to (2.10), given by

$$P_{\mathcal{T}}u = \sum_{T\in\mathcal{T}} \frac{\langle u, \xi_T\rangle_{L_2(\Omega)}}{\langle \mathbb{1}, \xi_T\rangle_{L_2(\Omega)}}\psi_{\mathcal{T},T}.$$

To prepare for the proof of (uniform) boundedness of $P_{\mathcal{T}}$, we list a few properties of the collections $\Xi_{\mathcal{T}}$, $\Theta_{\mathcal{T}}$ and $\Sigma_{\mathcal{T}}$. For $T \in \mathcal{T}$, we set $\omega_{\mathcal{T}}^{(0)}(T) := \overline{T}$, and for $i = 0, 1, \ldots$, define the 'rings'

$$R_{\mathcal{T}}^{(i+1)}(T) := \{T' \in \mathcal{T} : \overline{T'} \cap \omega_{\mathcal{T}}^{(i)}(T) \neq \emptyset\}, \quad \omega_{\mathcal{T}}^{(i+1)}(T) := \cup_{T' \in R_{\mathcal{T}}^{(i+1)}(T)} \overline{T'}.$$

It holds that

$$\text{(a) } \operatorname{supp} \xi_T \subset \omega_{\mathcal{T}}^{(0)}(T), \text{ (b) } \operatorname{supp} \sigma_{\mathcal{T},T} \subset \omega_{\mathcal{T}}^{(1)}(T), \text{ (c) } \operatorname{supp} \theta_T \subset \omega_{\mathcal{T}}^{(0)}(T), \tag{2.12}$$

$$\|\sigma_T\|_{H^k(\Omega)} \lesssim h_T^{d/2-k} \quad (k \in \{0,1\}), \tag{2.13}$$

$$\|\theta_T\|_{H^1(\Omega)} \lesssim h_T^{-1} \|\theta_T\|_{L_2(\Omega)}, \tag{2.14}$$

whilst moreover $\Xi_{\mathcal{T}}$ is such that

$$\langle \mathbb{1}, \xi_T \rangle_{L_2(\Omega)} \eqsim h_T^{d/2} \|\xi_T\|_{L_2(\Omega)}. \tag{2.15}$$

From (2.12)(b,c), we obtain that

$$\operatorname{supp} \psi_{\mathcal{T},T} \subset \omega_{\mathcal{T}}^{(1)}(T). \tag{2.16}$$

By using (2.6), (2.14), (2.13) and (2.12)(a) we infer that for $k \in \{0,1\}$

$$\Big\| \frac{\langle \sigma_{\mathcal{T},T} - \delta_{TT'} \mathbb{1}, \xi_{T'} \rangle_{L_2(\Omega)}}{\langle \theta_{T'}, \xi_{T'} \rangle_{L_2(\Omega)}} \theta_{T'} \Big\|_{H^k(\Omega)} \lesssim h_T^{-k} \|\sigma_{\mathcal{T},T} - \delta_{TT'} \mathbb{1}\|_{L_2(\operatorname{supp} \xi_{T'})} \lesssim h_T^{d/2-k},$$

which, by again using (2.13) and (2.14), shows that

$$\|\psi_{\mathcal{T},T}\|_{H^k(\Omega)} \lesssim h_T^{d/2-k} \quad (k \in \{0,1\}). \tag{2.17}$$

Theorem 2.3.2. *It holds that* $\sup_{\mathcal{T} \in \mathbb{T}} \|P_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{W},\mathscr{W})} < \infty$.

Proof. From (2.16), (2.17), and (2.15), we have

$$\begin{aligned} \|P_{\mathcal{T}} u\|_{H^k(T)} &\leq \sum_{T' \in R_{\mathcal{T}}^{(1)}(T)} \|\psi_{\mathcal{T},T'}\|_{H^k(\Omega)} \frac{\|u\|_{L_2(T')} \|\xi_{T'}\|_{L_2(\Omega)}}{|\langle \mathbb{1}, \xi_{T'} \rangle_{L_2(\Omega)}|} \\ &\lesssim h_T^{-k} \|u\|_{L_2(\omega_{\mathcal{T}}^{(1)}(T))} \quad (k \in \{0,1\}), \end{aligned} \tag{2.18}$$

which in particular shows that

$$\sup_{\mathcal{T} \in \mathbb{T}} \|P_{\mathcal{T}}\|_{\mathcal{L}(L_2(\Omega),L_2(\Omega))} < \infty. \tag{2.19}$$

To continue, we revisit the construction of $\mathscr{W}_{\mathcal{T}}$ and its basis $\Psi_{\mathcal{T}}$ by temporarily including in $N_{\mathcal{T}}^0$ also vertices of $\mathcal{T}$ that lie on the Dirichlet boundary

γ. Denoting the extended set of vertices by $N_{\mathcal{T}}$, consequently for the 'new' $\psi_{\mathcal{T},T}$, (2.9) shows that

$$\sum_{T\in\mathcal{T}} \psi_{\mathcal{T},T} = \sum_{\nu\in N_{\mathcal{T}}} \phi_{\mathcal{T},\nu} = \mathbb{1} \text{ on } \Omega. \tag{2.20}$$

For any $\nu \in N_{\mathcal{T}}$, we select a $(d-1)$-face e of a $T \in \mathcal{T}$ with $\nu \in e$ and $e \subset \gamma$ if $\nu \in \gamma$, and define the functional

$$g_{\mathcal{T},\nu}(u) := \fint_e u\, ds.$$

By the trace theorem and homogeneity arguments (see e.g [SZ90, (3.6)]), one infers that

$$|g_{\mathcal{T},\nu}(u)| \leq |e|^{-1}\|u\|_{L_1(e)} \lesssim h_T^{-\frac{d}{2}}\|u\|_{L_2(T)} + h_T^{-\frac{d}{2}+1}|u|_{H^1(T)}.$$

For $T \in \mathcal{T}$, we select a $\nu \in N_T$ with $\nu \in \gamma$ if $\overline{T} \cap \gamma \neq \emptyset$, and define

$$g_{\mathcal{T},T} := g_{\mathcal{T},\nu},$$

and a Scott-Zhang ([SZ90]) type quasi-interpolator $\Pi_{\mathcal{T}} : H^1(\Omega) \to \mathscr{W}_{\mathcal{T}}$[4] by

$$\Pi_{\mathcal{T}} u = \sum_{T\in\mathcal{T}} g_{\mathcal{T},T}(u)\psi_{\mathcal{T},T}.$$

It satisfies

$$\|\Pi_{\mathcal{T}} u\|_{H^k(T)} \lesssim h_T^{-k}\|u\|_{L_2(\omega_{\mathcal{T}}^{(2)}(T))} + h_T^{1-k}|u|_{H^1(\omega_{\mathcal{T}}^{(2)}(T))} \quad (k \in \{0,1\}).$$

Invoking (2.20) and using that $g_{\mathcal{T},T}(\mathbb{1}) = 1$, we infer that for $k \in \{0,1\}$

$$\begin{aligned}
\|(\mathrm{Id}-\Pi_{\mathcal{T}})u\|_{H^k(T)} &= \inf_{p\in\mathcal{P}_0} \|(\mathrm{Id}-\Pi_{\mathcal{T}})(u-p)\|_{H^k(T)} \\
&\leq \inf_{p\in\mathcal{P}_0} \|u-p\|_{H^k(T)} + h_T^{-k}\|u-p\|_{L_2(\omega_{\mathcal{T}}^{(2)}(T))} + h_T^{1-k}|u|_{H^1(\omega_{\mathcal{T}}^{(2)}(T))} \\
&\eqsim \inf_{p\in\mathcal{P}_0} h_T^{-k}\|u-p\|_{L_2(\omega_{\mathcal{T}}^{(2)}(T))} + h_T^{1-k}|u|_{H^1(\omega_{\mathcal{T}}^{(2)}(T))} \qquad (2.21)\\
&\eqsim h_T^{1-k}|u|_{H^1(\omega_{\mathcal{T}}^{(2)}(T))}
\end{aligned}$$

by an application of the Bramble-Hilbert lemma (cf. [SZ90, (4.2)]).

Noting that the 'new' $\psi_{\mathcal{T},T}$ differs only from the 'old', original one when $\overline{T} \cap \gamma \neq \emptyset$, and that for those T and $u \in H^1_{0,\gamma}(\Omega)$ it holds that $g_{\mathcal{T},T}(u) = 0$, we conclude that $\operatorname{ran}\Pi_{\mathcal{T}}|_{H^1_{0,\gamma}(\Omega)}$ is included in the original space $\mathscr{W}_{\mathcal{T}}$, which we

[4] The existence of such a $\Pi_{\mathcal{T}}$ which satisfies an estimate of type (2.21) for $k = 0$ can be used as a definition of a (lowest order) approximation property of $\mathscr{W}_{\mathcal{T}}$.

consider again from here on. Using that $P_{\mathcal{T}}$ is a projector onto this $\mathscr{W}_{\mathcal{T}}$, for $u \in H^1_{0,\gamma}(\Omega)$ writing $P_{\mathcal{T}}u = \Pi_{\mathcal{T}}u + P_{\mathcal{T}}(\mathrm{Id} - \Pi_{\mathcal{T}})u$, using (2.18) and (2.21) for $k \in \{0,1\}$ we arrive at

$$\begin{aligned}\|P_{\mathcal{T}}u\|_{H^1(T)} &\lesssim \|\Pi_{\mathcal{T}}u\|_{H^1(T)} + h_T^{-1}\|(\mathrm{Id} - \Pi_{\mathcal{T}})u\|_{L_2(\omega_{\mathcal{T}}^{(1)}(T))} \\ &\lesssim \|u\|_{H^1(\omega_{\mathcal{T}}^{(2)}(T))} + h_T^{-1}\|(\mathrm{Id} - \Pi_{\mathcal{T}})u\|_{L_2(\omega_{\mathcal{T}}^{(1)}(T))} \\ &\lesssim \|u\|_{H^1(\omega_{\mathcal{T}}^{(3)}(T))},\end{aligned}$$

and consequently,

$$\sup_{\mathcal{T}\in\mathbb{T}} \|P_{\mathcal{T}}\|_{\mathcal{L}(H^1_{0,\gamma}(\Omega),H^1_{0,\gamma}(\Omega)))} < \infty.$$

In combination with (2.19), the proof is completed by an application of the Riesz-Thorin interpolation theorem. □

From Proposition 2.2.3 and Theorem 2.3.2 we conclude the following:

Corollary 2.3.3. *For $D_{\mathcal{T}}\colon \mathscr{V}_{\mathcal{T}} \to \mathscr{W}'_{\mathcal{T}}$ defined by $(D_{\mathcal{T}}v)(w) := (Dv)(w) = \langle v, w\rangle_{L_2(\Omega)}$, it holds that $D_{\mathcal{T}} \in \mathcal{L}\mathrm{is}(\mathscr{V}_{\mathcal{T}}, \mathscr{W}'_{\mathcal{T}})$ with $\|D_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{V}_{\mathcal{T}},\mathscr{W}'_{\mathcal{T}})} \le 1$ and $\sup_{\mathcal{T}\in\mathbb{T}} \|D_{\mathcal{T}}^{-1}\|_{\mathcal{L}(\mathscr{W}'_{\mathcal{T}},\mathscr{V}_{\mathcal{T}})} \le \sup_{\mathcal{T}\in\mathbb{T}} \|P_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{W}_{\mathcal{T}},\mathscr{W}_{\mathcal{T}})} < \infty$.*

This result is thus valid *without any additional assumptions on the mesh grading*. The latter is a consequence of the fact that we were able to equip $\mathscr{V}_{\mathcal{T}}$ and $\mathscr{W}_{\mathcal{T}}$ with local biorthogonal bases. (Compare [Ste03a, eq. (2.30)] for conditions on the mesh grading without having local biorthogonal bases). Additionally, the biorthogonality has the important advantage of the matrix

$$\boldsymbol{D}_{\mathcal{T}} = \langle \Xi_{\mathcal{T}}, \Psi_{\mathcal{T}}\rangle_{L_2(\Omega)} = \mathrm{diag}\{|T|\colon T \in \mathcal{T}\}$$

being *diagonal*.

Before we discuss in §2.3.3 the construction of $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{W}_{\mathcal{T}}, \mathscr{W}'_{\mathcal{T}})$, being the last ingredient of our preconditioner, in the following subsection §2.3.2 we revisit the construction of $\mathscr{W}_{\mathcal{T}}$ and $\boldsymbol{D}_{\mathcal{T}}$ in the manifold case.

2.3.2 Construction of $\mathscr{W}_{\mathcal{T}}$ and $\boldsymbol{D}_{\mathcal{T}}$ in the manifold case

Let Γ be a compact d-dimensional Lipschitz, piecewise smooth manifold in $\mathbb{R}^{d'}$ for some $d' \ge d$ with or without boundary $\partial\Gamma$. For some closed measurable $\gamma \subset \partial\Gamma$ and $s \in [0,1]$, let

$$\mathscr{W} := [L_2(\Gamma), H^1_{0,\gamma}(\Gamma)]_{s,2}, \quad \mathscr{V} := \mathscr{W}'.$$

We assume that Γ is given as the essentially disjoint union of $\cup_{i=1}^p \overline{\chi_i(\Omega_i)}$, with, for $1 \le i \le p$, $\chi_i\colon \mathbb{R}^d \to \mathbb{R}^{d'}$ being some smooth regular parametrization, and

$\Omega_i \subset \mathbb{R}^d$ an open polytope. W.l.o.g. assuming that for $i \neq j$, $\overline{\Omega}_i \cap \overline{\Omega}_j = \emptyset$, we define

$$\chi\colon \Omega := \cup_{i=1}^p \Omega_i \to \cup_{i=1}^p \chi_i(\Omega_i) \text{ by } \chi|_{\Omega_i} = \chi_i.$$

Let $\mathbb{T}$ be a family of conforming partitions $\mathcal{T}$ of Γ into 'panels' such that, for $1 \leq i \leq p$, $\chi^{-1}(\mathcal{T}) \cap \Omega_i$ is a uniformly shape regular conforming partition of Ω_i into d-simplices (that for $d = 1$ satisfies a uniform K-mesh property). We assume that γ is a (possibly empty) union of 'faces' of $T \in \mathcal{T}$ (i.e., sets of type $\chi_i(e)$, where e is a $(d-1)$-dimensional face of $\chi_i^{-1}(T)$).

As in Sect. 2.3, for $\mathcal{T} \in \mathbb{T}$, we define $N_{\mathcal{T}}^0$ as the set of vertices of $\mathcal{T}$ that are not on γ, set $d_{\mathcal{T},\nu} := \#\{T \in \mathcal{T}\colon \nu \in \overline{T}\}$, and for $T \in \mathcal{T}$, define $h_T := |T|^{1/d}$ and $N_{\mathcal{T},T}^0 := N_{\mathcal{T}}^0 \cap N_T$, with N_T being the set of the vertices of T.

We set

$$\mathscr{V}_{\mathcal{T}} = \mathscr{S}_{\mathcal{T}}^{-1,0} := \{u \in L_2(\Gamma)\colon u \circ \chi|_{\chi^{-1}(T)} \in \mathcal{P}_0 \ (T \in \mathcal{T})\} \subset \mathscr{V},$$
$$\mathscr{S}_{\mathcal{T},0}^{0,1} := \{u \in H_{0,\gamma}^1(\Gamma)\colon u \circ \chi|_{\chi^{-1}(T)} \in \mathcal{P}_1 \ (T \in \mathcal{T})\},$$

equipped with $\Xi_{\mathcal{T}} = \{\xi_T\colon T \in \mathcal{T}\}$ and $\Phi_{\mathcal{T}} = \{\phi_{\mathcal{T},\nu}\colon \nu \in N_{\mathcal{T}}^0\}$, respectively, defined by $\xi_T := 1$ on T, $\xi_T := 0$ elsewhere, and $\phi_{\mathcal{T},\nu}(\nu') = \delta_{\nu,\nu'}$ $(\nu, \nu' \in N_{\mathcal{T}}^0)$. Furthermore, we define $\Sigma_{\mathcal{T}} = \{\sigma_{\mathcal{T},T}\colon T \in \mathcal{T}\} \subset \mathscr{S}_{\mathcal{T},0}^{0,1}$ and $\Theta_{\mathcal{T}} = \{\theta_T\colon T \in \mathcal{T}\} \subset H_{0,\gamma}^1(\Gamma)$ by $\sigma_{\mathcal{T},T} := \sum_{\nu \in N_{\mathcal{T},T}^0} d_{\mathcal{T},\nu}^{-1} \phi_{\mathcal{T},\nu}$, $\theta_T := \theta_{\chi^{-1}(T)} \circ \chi^{-1}$ on T and $\theta_T := 0$ elsewhere. Thanks to our assumption of $\theta_{\chi^{-1}(T)} \geq 0$, it holds that $\langle \theta_T, \xi_T \rangle_{L_2(\Gamma)} \eqsim \langle \theta_{\chi^{-1}(T)}, \xi_{\chi^{-1}(T)} \rangle_{L_2(\chi^{-1}(T))} \eqsim \|\theta_T\|_{L_2(\Gamma)} \|\xi_T\|_{L_2(\Gamma)}$ (cf. (2.6)).

Now defining $\Psi_{\mathcal{T}} := \{\psi_{\mathcal{T},T}\colon T \in \mathcal{T}\}$ and $\mathscr{W}_{\mathcal{T}} := \operatorname{span} \Psi_{\mathcal{T}} \subset \mathscr{W}$ by

$$\psi_{\mathcal{T},T} := \sigma_{\mathcal{T},T} + \frac{\langle \mathbb{1} - \sigma_{\mathcal{T},T}, \xi_T \rangle_{L_2(\Gamma)}}{\langle \theta_T, \xi_T \rangle_{L_2(\Gamma)}} \theta_T - \sum_{T' \in \mathcal{T} \setminus \{T\}} \frac{\langle \sigma_{\mathcal{T},T}, \xi_{T'} \rangle_{L_2(\Gamma)}}{\langle \theta_{T'}, \xi_{T'} \rangle_{L_2(\Gamma)}} \theta_{T'}, \tag{2.22}$$

and $D_{\mathcal{T}}\colon \mathscr{V}_{\mathcal{T}} \to \mathscr{W}_{\mathcal{T}}'$ by $(D_{\mathcal{T}} v)(w) := (Dv)(w) = \langle v, w \rangle_{L_2(\Gamma)}$, the analysis from Sect. 2.3 applies verbatim by only changing $\langle\, , \rangle_{L_2(\Omega)}$ into $\langle\, , \rangle_{L_2(\Gamma)}$. It yields that $\|D_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{V}_{\mathcal{T}}, \mathscr{W}_{\mathcal{T}}')} \leq 1$, $\sup_{\mathcal{T} \in \mathbb{T}} \|D_{\mathcal{T}}^{-1}\|_{\mathcal{L}(\mathscr{W}_{\mathcal{T}}', \mathscr{V}_{\mathcal{T}})} < \infty$, and $\boldsymbol{D}_{\mathcal{T}} = \operatorname{diag}\{\langle \mathbb{1}, \xi_T \rangle_{L_2(\Gamma)}\colon T \in \mathcal{T}\}$.

A *hidden problem*, however, is that the computation of $\boldsymbol{D}_{\mathcal{T}}$, and that of the scalar products in (2.22) involve integrals over Γ that generally have to be approximated using numerical quadrature. Recalling that, for $s > 0$, the preconditioner $\boldsymbol{G}_{\mathcal{T}} = \boldsymbol{D}_{\mathcal{T}}^{-1} \boldsymbol{B}_{\mathcal{T}} \boldsymbol{D}_{\mathcal{T}}^{-\top}$ is *not* a uniformly well-conditioned matrix, it is a priorily not clear which quadrature errors are allowable, in particular when $\mathcal{T}$ is far from being quasi-uniform. For this reason, in the following §2.3.2 we propose a slightly modified construction of $\mathscr{W}_{\mathcal{T}}$ and $\boldsymbol{D}_{\mathcal{T}}$ that does not require the evaluation of integrals over Γ.

As a preparation, in the following lemma we present a non-standard inverse inequality on the family $(\mathscr{V}_{\mathcal{T}})_{\mathcal{T} \in \mathbb{T}}$. Proofs of this inequality for $d \leq 3$ can be found in [DFG+04, GHS05]. It turns out that our construction of a 'local' collection $\Psi_{\mathcal{T}} \subset H_{0,\gamma}^1(\Omega)$ that is biorthogonal to $\Xi_{\mathcal{T}}$ allows for a very simple proof.

Lemma 2.3.4 (inverse inequality). *With $h_{\mathcal{T}}|_T := h_T$, it holds that*

$$\|h_{\mathcal{T}} v_{\mathcal{T}}\|_{L_2(\Gamma)} \lesssim \|v_{\mathcal{T}}\|_{H^1_{0,\gamma}(\Gamma)'} \quad (v_{\mathcal{T}} \in \mathscr{V}_{\mathcal{T}}).$$

Proof. For $P_{\mathcal{T}} : L_2(\Gamma) \to H^1_{0,\gamma}(\Gamma)$ defined by

$$P_{\mathcal{T}} u = \sum_{T\in\mathcal{T}} \frac{\langle u, \xi_T\rangle_{L_2(\Gamma)}}{\langle \mathbb{1}, \xi_T\rangle_{L_2(\Gamma)}} \psi_{\mathcal{T},T}.$$

we have $\mathrm{ran}(\mathrm{Id} - P_{\mathcal{T}}) = \mathscr{V}_{\mathcal{T}}^{\perp_{L_2(\Gamma)}}$, and as follows from (2.18),

$$\|P_{\mathcal{T}} u\|_{H^1(\Gamma)} \lesssim \|h_{\mathcal{T}}^{-1} u\|_{L_2(\Gamma)} \quad (u \in L_2(\Gamma)).$$

The proof is completed by

$$\|v_{\mathcal{T}}\|_{H^1_{0,\gamma}(\Gamma)'} = \sup_{0\neq w\in H^1_{0,\gamma}(\Gamma)} \frac{\langle v_{\mathcal{T}}, w\rangle_{L_2(\Gamma)}}{\|w\|_{H^1(\Gamma)}} \geq \frac{\langle v_{\mathcal{T}}, P_{\mathcal{T}} h_{\mathcal{T}}^2 v_{\mathcal{T}}\rangle_{L_2(\Gamma)}}{\|P_{\mathcal{T}} h_{\mathcal{T}}^2 v_{\mathcal{T}}\|_{H^1(\Gamma)}} \gtrsim \frac{\langle h_{\mathcal{T}} v_{\mathcal{T}}, h_{\mathcal{T}} v_{\mathcal{T}}\rangle_{L_2(\Gamma)}}{\|h_{\mathcal{T}} v_{\mathcal{T}}\|_{L_2(\Gamma)}}.$$

□

Modified construction for manifolds

Given $\mathcal{T} \in \mathbb{T}$, on $L_2(\Gamma)$ we define an additional, 'mesh-dependent' scalar product

$$\langle u, v\rangle_{\mathcal{T}} := \sum_{T\in\mathcal{T}} \frac{|T|}{|\chi^{-1}(T)|} \int_{\chi^{-1}(T)} u(\chi(x))v(\chi(x))dx.$$

It is constructed from

$$\langle u, v\rangle_{L_2(\Gamma)} = \int_\Omega u(\chi(x))v(\chi(x))|\partial\chi(x)|dx$$

by replacing on each $\chi^{-1}(T)$, the Jacobian $|\partial\chi|$ by its average $\frac{|T|}{|\chi^{-1}(T)|}$ over $\chi^{-1}(T)$.[5]

We now *redefine* $\Psi_{\mathcal{T}} := \{\psi_{\mathcal{T},T} : T \in \mathcal{T}\}$, $\mathscr{W}_{\mathcal{T}} := \operatorname{span}\Psi_{\mathcal{T}} \subset \mathscr{W}$ by

$$\psi_{\mathcal{T},T} := \sigma_{\mathcal{T},T} + \frac{\langle \mathbb{1} - \sigma_{\mathcal{T},T}, \xi_T\rangle_{\mathcal{T}}}{\langle \theta_T, \xi_T\rangle_{\mathcal{T}}}\theta_T - \sum_{T'\in\mathcal{T}\setminus\{T\}} \frac{\langle \sigma_{\mathcal{T},T}, \xi_{T'}\rangle_{\mathcal{T}}}{\langle \theta_{T'}, \xi_{T'}\rangle_{\mathcal{T}}}\theta_{T'},$$

and $D_{\mathcal{T}} : \mathscr{V}_{\mathcal{T}} \to \mathscr{W}_{\mathcal{T}}'$ by $(D_{\mathcal{T}} v_{\mathcal{T}})(w_{\mathcal{T}}) := \langle v_{\mathcal{T}}, w_{\mathcal{T}}\rangle_{\mathcal{T}}$. Then, as in the domain case, we get the explicit formulas

$$\boldsymbol{D}_{\mathcal{T}} = \langle \Xi_{\mathcal{T}}, \Psi_{\mathcal{T}}\rangle_{\mathcal{T}} = \mathrm{diag}\{\langle \mathbb{1}, \xi_T\rangle_{\mathcal{T}} : T \in \mathcal{T}\} = \mathrm{diag}\{|T| : T \in \mathcal{T}\},$$

[5] It will be clear from the following that $\frac{|T|}{|\chi^{-1}(T)|}$ can be read as *any* constant approximation to $|\partial\chi|$ on $L_\infty(\chi^{-1}(T))$-distance $\lesssim h_{\chi^{-1}(T)}$, for example $|\partial\chi(z)|$ for some $z \in \chi^{-1}(T)$. Then in the following, the volumes $|T|$ in the expression for $\boldsymbol{D}_{\mathcal{T}}$ should be read as $|\chi^{-1}(T)||\partial\chi(z)|$, with which also the computation of $|T|$ is avoided.

and

(2.23)
$$\psi_{\mathcal{T},T} = \sum_{\nu\in N^0_{\mathcal{T},T}} d^{-1}_{\mathcal{T},\nu}\phi_{\mathcal{T},\nu} + \big(1 - \tfrac{1}{d+1}\sum_{\nu\in N^0_{\mathcal{T},T}} d^{-1}_{\mathcal{T},\nu}\big)\theta_T - \sum_{T'\in\mathcal{T}\setminus\{T\}} \big(\tfrac{1}{d+1}\sum_{\nu\in N^0_{\mathcal{T},T}\cap N^0_{\mathcal{T},T'}} d^{-1}_{\mathcal{T},\nu}\big)\theta_{T'}.$$

thus with coefficients that are *independent* of χ.

What remains is to prove the uniform boundedness of $\|D_\mathcal{T}\|_{\mathcal{L}(\mathscr{V}_\mathcal{T},\mathscr{W}'_\mathcal{T})}$, and that of $\|D^{-1}_\mathcal{T}\|_{\mathcal{L}(\mathscr{W}'_\mathcal{T},\mathscr{V}_\mathcal{T})}$. Because of the definition of $D_\mathcal{T}$ in terms of the mesh-dependent scalar product, for doing so we cannot simply rely on the 'Fortin criterion' from Proposition 2.2.3.

Lemma 2.3.5. *It holds that* $\sup_{\mathcal{T}\in\mathbb{T}} \|D_\mathcal{T}\|_{\mathcal{L}(\mathscr{V}_\mathcal{T},\mathscr{W}'_\mathcal{T})} < \infty$.

Proof. If $s = 0$, i.e., when $\mathscr{W} = L_2(\Gamma) \simeq L_2(\Gamma)' = \mathscr{V}$, then the uniform boundedness of $\|D_\mathcal{T}\|_{\mathcal{L}(\mathscr{V}_\mathcal{T},\mathscr{W}'_\mathcal{T})}$ follows directly from $\langle\cdot,\cdot\rangle_\mathcal{T} \eqsim \|\cdot\|^2_{L_2(\Gamma)}$.

By an interpolation argument, in the following it suffices to consider the case $s = 1$, i.e., $\mathscr{W} = H^1_{0,\gamma}(\Gamma)$ and $\mathscr{V} = H^1_{0,\gamma}(\Gamma)'$. By definition of $\langle\ ,\ \rangle_\mathcal{T}$, it holds that

$$|\langle v,u\rangle_\mathcal{T} - \langle v,u\rangle_{L_2(\Gamma)}| \lesssim \|h_\mathcal{T} v\|_{L_2(\Gamma)}\|u\|_{L_2(\Gamma)} \quad (v,u\in L_2(\Gamma)). \tag{2.24}$$

By writing $(D_\mathcal{T} v_\mathcal{T})(w_\mathcal{T}) = \langle v_\mathcal{T}, w_\mathcal{T}\rangle_{L_2(\Gamma)} + \langle v_\mathcal{T}, w_\mathcal{T}\rangle_\mathcal{T} - \langle v_\mathcal{T}, w_\mathcal{T}\rangle_{L_2(\Gamma)}$, the uniform boundedness of $\|D_\mathcal{T}\|_{\mathcal{L}(\mathscr{V}_\mathcal{T},\mathscr{W}'_\mathcal{T})}$ (for $s = 1$) now follows by combining (2.24) and Lemma 2.3.4. □

The $\langle\ ,\ \rangle_\mathcal{T}$-biorthogonal projector $\check{P}_\mathcal{T} : L_2(\Omega) \to H^1_{0,\gamma}(\Omega)$ with $\operatorname{ran}\check{P}_\mathcal{T} = \mathscr{W}_\mathcal{T}$ and $\operatorname{ran}(\mathrm{Id}-\check{P}_\mathcal{T}) = \mathscr{V}_\mathcal{T}^{\perp_{\langle\ ,\ \rangle_\mathcal{T}}}$ exists and is given by $\check{P}_\mathcal{T} u = \sum_{T\in\mathcal{T}} |T|^{-1}\langle u,\xi_T\rangle_\mathcal{T}\psi_{\mathcal{T},T}$. Since $\langle\ ,\ \rangle_\mathcal{T}$ gives rise to a norm that is uniformly equivalent to $\|\ \|_{L_2(\Gamma)}$, the proof of Theorem 2.3.2 again applies, and shows that

$$\sup_{\mathcal{T}\in\mathbb{T}} \|\check{P}_\mathcal{T}\|_{\mathcal{L}(L_2(\Gamma),L_2(\Gamma))} < \infty, \quad \sup_{\mathcal{T}\in\mathbb{T}} \|\check{P}_\mathcal{T}\|_{\mathcal{L}(H^1_{0,\gamma}(\Gamma),H^1_{0,\gamma}(\Gamma))} < \infty,$$

as well as

$$\|\check{P}_\mathcal{T} u\|_{H^1(\Gamma)} \lesssim \|h^{-1}_\mathcal{T} u\|_{L_2(\Gamma)} \quad (u\in L_2(\Gamma)). \tag{2.25}$$

These properties of $\check{P}_\mathcal{T}$ will be the key to prove the uniform boundedness of $\|D^{-1}_\mathcal{T}\|_{\mathcal{L}(\mathscr{W}'_\mathcal{T},\mathscr{V}_\mathcal{T})}$. Indeed, for $s = 0$ uniform boundedness of $\|D^{-1}_\mathcal{T}\|_{\mathcal{L}(\mathscr{V}_\mathcal{T},\mathscr{W}'_\mathcal{T})}$ follows from

$$(D_\mathcal{T} v_\mathcal{T})(\check{P}_\mathcal{T} v_\mathcal{T}) = \langle v_\mathcal{T}, v_\mathcal{T}\rangle_\mathcal{T} \eqsim \|v_\mathcal{T}\|^2_{L_2(\Gamma)} \gtrsim \|v_\mathcal{T}\|_{L_2(\Gamma)}\|\check{P}_\mathcal{T} v_\mathcal{T}\|_{L_2(\Gamma)}.$$

To conclude, by an interpolation argument, uniform boundedness of $\|D^{-1}_\mathcal{T}\|_{\mathcal{L}(\mathscr{V}_\mathcal{T},\mathscr{W}'_\mathcal{T})}$ for any $s\in[0,1]$, it is sufficient to verify the case $s = 1$, which will be done using the following modified inverse inequality.

Lemma 2.3.6. *It holds that*

$$\|h_{\mathcal{T}} v_{\mathcal{T}}\|_{L_2(\Gamma)} \lesssim \sup_{0\neq w\in H^1_{0,\gamma}(\Gamma)} \frac{\langle v_{\mathcal{T}}, w\rangle_{\mathcal{T}}}{\|w\|_{H^1(\Gamma)}} \quad (v_{\mathcal{T}} \in \mathscr{V}_{\mathcal{T}}).$$

Proof. Similar to proof of Lemma 2.3.4, using (2.25) for $v_{\mathcal{T}} \in \mathscr{V}_{\mathcal{T}}$ we estimate

$$\sup_{0\neq w\in H^1_{0,\gamma}(\Gamma)} \frac{\langle v_{\mathcal{T}}, w\rangle_{\mathcal{T}}}{\|w\|_{H^1(\Gamma)}} \geq \frac{\langle v_{\mathcal{T}}, \check{P}_{\mathcal{T}} h_{\mathcal{T}}^2 v_{\mathcal{T}}\rangle_{\mathcal{T}}}{\|\check{P}_{\mathcal{T}} h_{\mathcal{T}}^2 v_{\mathcal{T}}\|_{H^1(\Gamma)}} \gtrsim \frac{\langle h_{\mathcal{T}} v_{\mathcal{T}}, h_{\mathcal{T}} v_{\mathcal{T}}\rangle_{\mathcal{T}}}{\|h_{\mathcal{T}} v_{\mathcal{T}}\|_{L_2(\Gamma)}} \eqsim \|h_{\mathcal{T}} v_{\mathcal{T}}\|_{L_2(\Gamma)}. \square$$

Corollary 2.3.7. *It holds that*

$$\|v_{\mathcal{T}}\|_{H^1_{0,\gamma}(\Gamma)'} \eqsim \sup_{0\neq w_{\mathcal{T}}\in \mathscr{W}_{\mathcal{T}}} \frac{\langle v_{\mathcal{T}}, w_{\mathcal{T}}\rangle_{\mathcal{T}}}{\|w_{\mathcal{T}}\|_{H^1(\Gamma)}} \quad (v_{\mathcal{T}} \in \mathscr{V}_{\mathcal{T}}),$$

(*with '$\lesssim$' being the statement* $\sup_{\mathcal{T}\in\mathbb{T}} \|D_{\mathcal{T}}^{-1}\|_{\mathcal{L}(\mathscr{W}'_{\mathcal{T}}, \mathscr{V}_{\mathcal{T}})} < \infty$ *for* $s = 1$).

Proof. The inequality '$\gtrsim$' is the statement of Lemma 2.3.5 for $s = 1$.

To prove the other direction, for $v \in L_2(\Gamma)$, (2.24) shows that

$$\Big| \|v\|_{H^1_{0,\gamma}(\Gamma)'} - \sup_{0\neq w\in H^1_{0,\gamma}(\Gamma)} \frac{\langle v, w\rangle_{\mathcal{T}}}{\|w\|_{H^1(\Gamma)}} \Big| \lesssim \|h_{\mathcal{T}} v\|_{L_2(\Gamma)},$$

Taking $v = v_{\mathcal{T}} \in \mathscr{V}_{\mathcal{T}}$, from Lemma 2.3.6 we conclude that

$$\begin{aligned}\|v_{\mathcal{T}}\|_{H^1_{0,\gamma}(\Gamma)'} &\lesssim \sup_{0\neq w\in H^1_{0,\gamma}(\Gamma)} \frac{\langle v_{\mathcal{T}}, w\rangle_{\mathcal{T}}}{\|w\|_{H^1(\Gamma)}} = \sup_{0\neq w\in H^1_{0,\gamma}(\Gamma)} \frac{\langle v_{\mathcal{T}}, \check{P}_{\mathcal{T}} w\rangle_{\mathcal{T}}}{\|w\|_{H^1(\Gamma)}} \\ &\leq \|\check{P}_{\mathcal{T}}\|_{\mathcal{L}(H^1_{0,\gamma}(\Gamma), H^1_{0,\gamma}(\Gamma))} \sup_{0\neq w_{\mathcal{T}}\in \mathscr{W}_{\mathcal{T}}} \frac{\langle v_{\mathcal{T}}, w_{\mathcal{T}}\rangle_{\mathcal{T}}}{\|w_{\mathcal{T}}\|_{H^1(\Gamma)}} \lesssim \sup_{0\neq w_{\mathcal{T}}\in \mathscr{W}_{\mathcal{T}}} \frac{\langle v_{\mathcal{T}}, w_{\mathcal{T}}\rangle_{\mathcal{T}}}{\|w_{\mathcal{T}}\|_{H^1(\Gamma)}}\end{aligned}$$

by $\sup_{\mathcal{T}\in\mathbb{T}} \|\check{P}_{\mathcal{T}}\|_{\mathcal{L}(H^1_{0,\gamma}(\Gamma), H^1_{0,\gamma}(\Gamma))} < \infty$. $\square$

2.3.3 Construction of $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{W}_{\mathcal{T}}, \mathscr{W}'_{\mathcal{T}})$.

Having established $\sup_{T\in\mathbb{T}} \max\big(\|D_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{V}_{\mathcal{T}}, \mathscr{W}'_{\mathcal{T}})}, \|D_{\mathcal{T}}^{-1}\|_{\mathcal{L}(\mathscr{W}'_{\mathcal{T}}, \mathscr{V}_{\mathcal{T}})}\big) < \infty$, in both domain and manifold case, for the construction of uniform preconditioners it remains to find $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{W}_{\mathcal{T}}, \mathscr{W}'_{\mathcal{T}})$ with $\sup_{T\in\mathbb{T}} \|B_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{W}_{\mathcal{T}}, \mathscr{W}'_{\mathcal{T}})} < \infty$ and $\sup_{T\in\mathbb{T}} \|\Re(B_{\mathcal{T}})^{-1}\|_{\mathcal{L}(\mathscr{W}'_{\mathcal{T}}, \mathscr{W}_{\mathcal{T}})} < \infty$.

We add the following two assumptions on the collection $\Theta_{\mathcal{T}}$ of 'bubbles' and their span $\mathscr{B}_{\mathcal{T}} := \operatorname{span}\Theta_{\mathcal{T}}$. For $k \in \{0, 1\}$ it holds that

$$\|\sum_{T\in\mathcal{T}} c_T\theta_T\|^2_{H^k(\Omega)} \eqsim \sum_{T\in\mathcal{T}} h_T^{-2k}|c_T|^2\|\theta_T\|^2_{L_2(\Omega)}, \quad ((c_T)_{T\in\mathcal{T}} \subset \mathbb{R}), \tag{2.26}$$

$$\|u + v\|^2_{H^k(\Omega)} \gtrsim \|u\|^2_{H^k(\Omega)} + \|v\|^2_{H^k(\Omega)} \quad (u \in \mathscr{S}^{0,1}_{\mathcal{T},0},\ v \in \mathscr{B}_{\mathcal{T}}). \tag{2.27}$$

(Here and in the following, Ω should be read as Γ in the manifold case). Both properties are easily verified by a standard homogeneity argument for both our earlier specifications of possible $\Theta_\mathcal{T}$. From (2.27) it follows that $\mathscr{S}^{0,1}_{\mathcal{T},0} \cap \mathscr{B}_\mathcal{T} = \{0\}$. Let $I^\mathscr{S}_\mathcal{T}$ be the linear projector defined on $\mathscr{S}^{0,1} \oplus \mathscr{B}_\mathcal{T}$ by $\operatorname{ran} I^\mathscr{S}_\mathcal{T} = \mathscr{S}^{0,1}_{\mathcal{T},0}$ and $\operatorname{ran} I^\mathscr{B}_\mathcal{T} = \mathscr{B}_\mathcal{T}$, where $I^\mathscr{B}_\mathcal{T} := \mathrm{Id} - I^\mathscr{S}_\mathcal{T}$.

Below we give a construction of suitable $B_\mathcal{T}$ that is *independent* of the particular bubbles $\Theta_\mathcal{T}$ being chosen. Like $\mathscr{W}_\mathcal{T}$, we equip $\mathscr{S}^{0,1}_{\mathcal{T},0}$, $\mathscr{B}_\mathcal{T}$, and $\mathscr{S}^{0,1}_{\mathcal{T},0} \oplus \mathscr{B}_\mathcal{T}$ with $\|\,\|_\mathscr{W}$.

Proposition 2.3.8. *Given $B^\mathscr{S}_\mathcal{T} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}^{0,1}_{\mathcal{T},0}, (\mathscr{S}^{0,1}_{\mathcal{T},0})')$ and $B^\mathscr{B}_\mathcal{T} \in \mathcal{L}\mathrm{is}_c(\mathscr{B}_\mathcal{T}, \mathscr{B}'_\mathcal{T})$, let $B^{\mathscr{S}\oplus\mathscr{B}}_\mathcal{T} \colon \mathscr{S}^{0,1}_{\mathcal{T},0} \oplus \mathscr{B}_\mathcal{T} \to (\mathscr{S}^{0,1}_{\mathcal{T},0} \oplus \mathscr{B}_\mathcal{T})'$ be defined by*

$$(B^{\mathscr{S}\oplus\mathscr{B}}_\mathcal{T} w)(\tilde w) := (B^\mathscr{S}_\mathcal{T} I^\mathscr{S}_\mathcal{T} w)(I^\mathscr{S}_\mathcal{T} \tilde w) + (B^\mathscr{B}_\mathcal{T} I^\mathscr{B}_\mathcal{T} w)(I^\mathscr{B}_\mathcal{T} \tilde w).$$

Then thanks to (2.27)*, one has $B^{\mathscr{S}\oplus\mathscr{B}}_\mathcal{T} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}^{0,1}_{\mathcal{T},0} \oplus \mathscr{B}_\mathcal{T}, (\mathscr{S}^{0,1}_{\mathcal{T},0} \oplus \mathscr{B}_\mathcal{T})')$, and*

$$\begin{aligned}
&\|\Re(B^{\mathscr{S}\oplus\mathscr{B}}_\mathcal{T})^{-1}\|_{\mathcal{L}((\mathscr{S}^{0,1}_{\mathcal{T},0}\oplus\mathscr{B}_\mathcal{T})', \mathscr{S}^{0,1}_{\mathcal{T},0}\oplus\mathscr{B}_\mathcal{T})} \\
&\qquad \le 2\max(\|\Re(B^\mathscr{S}_\mathcal{T})^{-1}\|_{\mathcal{L}((\mathscr{S}^{0,1}_{\mathcal{T},0})', \mathscr{S}^{0,1}_{\mathcal{T},0})}, \|\Re(B^\mathscr{B}_\mathcal{T})^{-1}\|_{\mathcal{L}(\mathscr{B}'_\mathcal{T}, \mathscr{B}_\mathcal{T})}), \\
&\|B^{\mathscr{S}\oplus\mathscr{B}}_\mathcal{T}\|_{\mathcal{L}(\mathscr{S}^{0,1}_{\mathcal{T},0}\oplus\mathscr{B}_\mathcal{T}, (\mathscr{S}^{0,1}_{\mathcal{T},0}\oplus\mathscr{B}_\mathcal{T})')} \lesssim \max(\|B^\mathscr{S}_\mathcal{T}\|_{\mathcal{L}(\mathscr{S}^{0,1}_{\mathcal{T},0}, (\mathscr{S}^{0,1}_{\mathcal{T},0})')}, \|B^\mathscr{B}_\mathcal{T}\|_{\mathcal{L}(\mathscr{B}_\mathcal{T}, \mathscr{B}'_\mathcal{T})}).
\end{aligned}$$

Proof. One has

$$\begin{aligned}
|(B^{\mathscr{S}\oplus\mathscr{B}}_\mathcal{T} w)(w)| &\ge \min(\|\Re(B^\mathscr{S}_\mathcal{T})^{-1}\|^{-1}_{\mathcal{L}((\mathscr{S}^{0,1}_{\mathcal{T},0})', \mathscr{S}^{0,1}_{\mathcal{T},0})}, \|\Re(B^\mathscr{B}_\mathcal{T})^{-1}\|^{-1}_{\mathcal{L}(\mathscr{B}'_\mathcal{T}, \mathscr{B}_\mathcal{T})}) \\
&\qquad \times (\|I^\mathscr{S}_\mathcal{T} w\|^2_\mathscr{W} + \|I^\mathscr{B}_\mathcal{T} w\|^2_\mathscr{W}),
\end{aligned}$$

and

$$\begin{aligned}
|(B^{\mathscr{S}\oplus\mathscr{B}}_\mathcal{T} w)(\tilde w)| &\le \max(\|B^\mathscr{S}_\mathcal{T}\|_{\mathcal{L}(\mathscr{S}^{0,1}_{\mathcal{T},0}, (\mathscr{S}^{0,1}_{\mathcal{T},0})')}, \|B^\mathscr{B}_\mathcal{T}\|_{\mathcal{L}(\mathscr{B}_\mathcal{T}, \mathscr{B}'_\mathcal{T})}) \\
&\qquad \times \sqrt{\|I^\mathscr{S}_\mathcal{T} w\|^2_\mathscr{W} + \|I^\mathscr{B}_\mathcal{T} w\|^2_\mathscr{W}} \sqrt{\|I^\mathscr{S}_\mathcal{T} \tilde w\|^2_\mathscr{W} + \|I^\mathscr{B}_\mathcal{T} \tilde w\|^2_\mathscr{W}}.
\end{aligned}$$

From the triangle inequality and (2.27), one has $\frac{1}{2}\|w\|^2_\mathscr{W} \le \|I^\mathscr{S}_\mathcal{T} w\|^2_\mathscr{W} + \|I^\mathscr{B}_\mathcal{T} w\|^2_\mathscr{W} \lesssim \|w\|^2_\mathscr{W}$, which completes the proof. □

By equipping $\mathscr{W}_\mathcal{T}$, $\mathscr{S}^{0,1}_{\mathcal{T},0}$ and $\mathscr{B}_\mathcal{T}$ by $\Psi_\mathcal{T}$, $\Phi_\mathcal{T}$, and $\Theta_\mathcal{T}$, respectively, the applications of $I^\mathscr{S}_\mathcal{T}|_{\mathscr{W}_\mathcal{T}}$ and $I^\mathscr{B}_\mathcal{T}|_{\mathscr{W}_\mathcal{T}}$ can easily determined in linear complexity. Therefore a suitable definition of $B_\mathcal{T} \colon \mathscr{W}_\mathcal{T} \to \mathscr{W}'_\mathcal{T}$ is given by $(B_\mathcal{T} w)(\tilde w) := (B^{\mathscr{S}\oplus\mathscr{B}}_\mathcal{T} w)(\tilde w)$. Clearly,

$$\begin{aligned}
\|B_\mathcal{T}\|_{\mathcal{L}(\mathscr{W}_\mathcal{T}, \mathscr{W}'_\mathcal{T})} &\le \|B^{\mathscr{S}\oplus\mathscr{B}}_\mathcal{T}\|_{\mathcal{L}(\mathscr{S}^{0,1}_{\mathcal{T},0}\oplus\mathscr{B}_\mathcal{T}, (\mathscr{S}^{0,1}_{\mathcal{T},0}\oplus\mathscr{B}_\mathcal{T})')}, \\
\|\Re(B_\mathcal{T})^{-1}\|_{\mathcal{L}(\mathscr{W}'_\mathcal{T}, \mathscr{W}_\mathcal{T})} &\le \|\Re(B^{\mathscr{S}\oplus\mathscr{B}}_\mathcal{T})^{-1}\|_{\mathcal{L}((\mathscr{S}^{0,1}_{\mathcal{T},0}\oplus\mathscr{B}_\mathcal{T})', \mathscr{S}^{0,1}_{\mathcal{T},0}\oplus\mathscr{B}_\mathcal{T})}.
\end{aligned}$$

An obvious choice for $B_{\mathcal{T}}^{\mathscr{B}} \in \mathcal{L}\mathrm{is}_c(\mathscr{B}_{\mathcal{T}}, (\mathscr{B}_{\mathcal{T}})')$ such that

$$\max\big(\sup_{\mathcal{T}\in\mathbb{T}} \|B_{\mathcal{T}}^{\mathscr{B}}\|_{\mathcal{L}(\mathscr{B}_{\mathcal{T}},\mathscr{B}'_{\mathcal{T}})}, \|\Re(B_{\mathcal{T}}^{\mathscr{B}})^{-1}\|_{\mathcal{L}(\mathscr{B}'_{\mathcal{T}},\mathscr{B}_{\mathcal{T}})}\big) < \infty,$$

is, in view of (2.26) and $\|\theta_T\|_{L_2(\Omega)} \overset{(2.6)}{\eqsim} \frac{\langle\theta_T,\xi_T\rangle_{L_2(\Omega)}}{\|\xi_T\|_{L_2(\Omega)}} \overset{(2.7)}{=} |T|\|\xi_T\|_{L_2(\Omega)}^{-1} = h^{d/2}$, given by

$$\big(B_{\mathcal{T}}^{\mathscr{B}} \sum_{T\in\mathcal{T}} c_T\theta_T\big)\big(\sum_{T\in\mathcal{T}} d_T\theta_T\big) := \beta \sum_{T\in\mathcal{T}} h_T^{d-2s} c_T d_T. \tag{2.28}$$

for some constant β.

Possible choices for $B_{\mathcal{T}}^{\mathscr{S}} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}_{\mathcal{T},0}^{0,1}, (\mathscr{S}_{\mathcal{T},0}^{0,1})')$ with

$$\sup_{\mathcal{T}\in\mathbb{T}} \big(\|B_{\mathcal{T}}^{\mathscr{S}}\|_{\mathcal{L}(\mathscr{S}_{\mathcal{T},0}^{0,1},(\mathscr{S}_{\mathcal{T},0}^{0,1})')}, \|\Re(B_{\mathcal{T}}^{\mathscr{S}})^{-1}\|_{\mathcal{L}((\mathscr{S}_{\mathcal{T},0}^{0,1})',\mathscr{S}_{\mathcal{T},0}^{0,1})}\big) < \infty$$

include $(B_{\mathcal{T}}^{\mathscr{S}} u)(v) := (Bu)(v)$ $(u, v \in \mathscr{S}_{\mathcal{T},0}^{0,1})$ for some $B \in \mathcal{L}\mathrm{is}_c(\mathscr{W}, \mathscr{W}')$.

For $d \in \{2, 3\}$ and $\mathscr{W} = H_{00}^{\frac{1}{2}}(\Gamma) := [L_2(\Gamma), H_0^1(\Gamma)]_{\frac{1}{2},2}$, for this B one may take the hypersingular integral operator, whereas for $\partial\Gamma \neq \emptyset$, and $\mathscr{W} = H^{\frac{1}{2}}(\Gamma) = [L_2(\Gamma), H^1(\Gamma)]_{\frac{1}{2},2}$ the recently introduced modified hypersingular integral operator can be applied (see [HJHUT18]). (Note that $H_0^1(\Gamma) = H^1(\Gamma)$ when $\partial\Gamma = \emptyset$.)

For a family of quasi-uniform partitions generated by a repeated application of *uniform refinements* starting from some given initial partition, a computationally attractive alternative choice for $B_{\mathcal{T}}^{\mathscr{S}}$ is provided by the multi-level operator from [BPV00], whose application can be performed in *linear complexity*. In Chapter 3, such operators will be discussed that also apply on locally refined meshes.

For $\mathscr{W} = H_{0,\gamma}^1(\Omega)$, i.e., when A is an operator of order -2 (cf. [FH19]), one obviously takes $(B_{\mathcal{T}}^{\mathscr{S}} u)(v) = \int_\Omega \nabla u \cdot \nabla v\, dx$, or $(B_{\mathcal{T}}^{\mathscr{S}} u)(v) = \int_\Omega \nabla u \cdot \nabla v\, dx + \int_\Omega uv\, dx$ when $\mathrm{meas}(\gamma) = 0$, whose application can be performed in linear complexity.

2.3.4 Implementation

For both the domain case and the construction in the manifold case in §2.3.2, the matrix representation $\boldsymbol{G}_{\mathcal{T}} = \mathcal{F}_{\Xi_{\mathcal{T}}}^{-1} G_{\mathcal{T}} (\mathcal{F}'_{\Xi_{\mathcal{T}}})^{-1}$ of our preconditioner $G_{\mathcal{T}}$ reads as $\boldsymbol{G}_{\mathcal{T}} = \boldsymbol{D}_{\mathcal{T}}^{-1} \boldsymbol{B}_{\mathcal{T}} \boldsymbol{D}_{\mathcal{T}}^{-\top}$ with

$$\boldsymbol{D}_{\mathcal{T}} = \mathrm{diag}\{|T| : T \in \mathcal{T}\},$$

and

$$\begin{aligned}
\boldsymbol{B}_{\mathcal{T}} &:= \mathcal{F}'_{\Psi_{\mathcal{T}}} B_{\mathcal{T}} \mathcal{F}_{\Psi_{\mathcal{T}}} \\
&= \mathcal{F}'_{\Psi_{\mathcal{T}}} \big((I_{\mathcal{T}}^{\mathscr{S}}|_{\mathscr{W}_{\mathcal{T}}})' B_{\mathcal{T}}^{\mathscr{S}} I_{\mathcal{T}}^{\mathscr{S}}|_{\mathscr{W}_{\mathcal{T}}} + (I_{\mathcal{T}}^{\mathscr{B}}|_{\mathscr{W}_{\mathcal{T}}})' B_{\mathcal{T}}^{\mathscr{B}} I_{\mathcal{T}}^{\mathscr{B}}|_{\mathscr{W}_{\mathcal{T}}}\big) \mathcal{F}_{\Psi_{\mathcal{T}}} \\
&= \boldsymbol{p}_{\mathcal{T}}^\top \boldsymbol{B}_{\mathcal{T}}^{\mathscr{S}} \boldsymbol{p}_{\mathcal{T}} + \boldsymbol{q}_{\mathcal{T}}^\top \boldsymbol{B}_{\mathcal{T}}^{\mathscr{B}} \boldsymbol{q}_{\mathcal{T}},
\end{aligned}$$

where

$$\boldsymbol{B}_{\mathcal{T}}^{\mathscr{S}} := \mathcal{F}'_{\Phi_{\mathcal{T}}} B_{\mathcal{T}}^{\mathscr{S}} \mathcal{F}_{\Phi_{\mathcal{T}}}, \quad \boldsymbol{p}_{\mathcal{T}} := \mathcal{F}_{\Phi_{\mathcal{T}}}^{-1} I_{\mathcal{T}}^{\mathscr{S}} |_{\mathscr{W}_{\mathcal{T}}} \mathcal{F}_{\Psi_{\mathcal{T}}},$$
$$\boldsymbol{B}_{\mathcal{T}}^{\mathscr{B}} := \mathcal{F}'_{\Theta_{\mathcal{T}}} B_{\mathcal{T}}^{\mathscr{B}} \mathcal{F}_{\Theta_{\mathcal{T}}}, \quad \boldsymbol{q}_{\mathcal{T}} := \mathcal{F}_{\Theta_{\mathcal{T}}}^{-1} I_{\mathcal{T}}^{\mathscr{B}} |_{\mathscr{W}_{\mathcal{T}}} \mathcal{F}_{\Psi_{\mathcal{T}}}.$$

By substituting the definition of $B_{\mathcal{T}}^{\mathscr{B}}$ from (2.28), the definition of the basis $\Psi_{\mathcal{T}} = \{\psi_{\mathcal{T},T}\}_{T\in\mathcal{T}}$ for $\mathscr{W}_{\mathcal{T}}$ from (2.11) and (2.23), and that of the bases $\Phi_{\mathcal{T}} = \{\phi_{\mathcal{T},\nu}\}_{\nu\in N_{\mathcal{T}}^0}$ and $\Theta_{\mathcal{T}} = \{\theta_T\}_{T\in\mathcal{T}}$ for $\mathscr{S}_{\mathcal{T},0}^{0,1}$ and $\mathscr{B}_{\mathcal{T}}$, respectively, we find that

$$\boldsymbol{B}_{\mathcal{T}}^{\mathscr{B}} = \beta \boldsymbol{D}_{\mathcal{T}}^{1-\frac{2s}{d}}, \quad (\boldsymbol{p}_{\mathcal{T}})_{\nu T} = \begin{cases} d_{\mathcal{T},\nu}^{-1} & \text{if } \nu \in N_{\mathcal{T},T}^0, \\ 0 & \text{if } \nu \notin N_{\mathcal{T},T}^0, \end{cases}$$
$$(\boldsymbol{q}_{\mathcal{T}})_{T'T} = \delta_{T'T} - \tfrac{1}{d+1} \sum_{\nu\in N_{\mathcal{T},T}^0 \cap N_{\mathcal{T},T'}^0} d_{\mathcal{T},\nu}^{-1},$$

whereas $\boldsymbol{B}_{\mathcal{T}}^{\mathscr{S}}$ depends on $B_{\mathcal{T}}^{\mathscr{S}} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}_{\mathcal{T},0}^{0,1}, (\mathscr{S}_{\mathcal{T},0}^{0,1})')$ being chosen. The cost of the application of $\boldsymbol{G}_{\mathcal{T}}$ is the cost of the application of $\boldsymbol{B}_{\mathcal{T}}^{\mathscr{S}}$ plus cost that scales linearly in $\#\mathcal{T}$.

2.4 Continuous piecewise linear discretization space

Let a bounded polytopal domain $\Omega \subset \mathbb{R}^d$, $\gamma \subset \partial\Omega$, $s \in [0,1]$, $\mathscr{W} := [L_2(\Omega), H_{0,\gamma}^1(\Omega)]_{s,2}$, $\mathscr{V} := \mathscr{W}'$, $D \in \mathcal{L}\mathrm{is}(\mathscr{V}, \mathscr{W}')$, $(\mathcal{T})_{\mathcal{T}\in\mathbb{T}}$, $N_{\mathcal{T}}^0$, $d_{\mathcal{T},\nu}$, $N_{\mathcal{T}}$, and $N_{\mathcal{T},T}^0$ be all as in Sect. 2.3. In addition, for $\mathcal{T} \in \mathbb{T}$ let $N_{\mathcal{T}}$ be the set of *all* vertices of $\mathcal{T}$, so including those on a possibly non-empty γ, and for $\nu \in N_{\mathcal{T}}$ let $\omega_{\mathcal{T}}(\nu) := \cup_{\{T\in\mathcal{T}\colon \nu\in N_T\}} \overline{T}$.

We take

$$\mathscr{V}_{\mathcal{T}} = \mathscr{S}_{\mathcal{T}}^{0,1} := \{u \in H^1(\Omega)\colon u|_T \in \mathcal{P}_1\,(T \in \mathcal{T})\} \subset \mathscr{V},$$

and, as in Sect. 2.3,

$$\mathscr{S}_{\mathcal{T},0}^{0,1} := \{u \in H_{0,\gamma}^1(\Omega)\colon u|_T \in \mathcal{P}_1\,(T \in \mathcal{T})\},$$

equipped with nodal bases $\Xi_{\mathcal{T}} = \{\xi_{\mathcal{T},\nu}\colon \nu \in N_{\mathcal{T}}\}$ and $\Phi_{\mathcal{T}} = \{\phi_{\mathcal{T},\nu}\colon \nu \in N_{\mathcal{T}}^0\}$, respectively, defined by

$$\xi_{\mathcal{T},\nu}(\nu') = \delta_{\nu,\nu'} \quad (\nu,\nu' \in N_{\mathcal{T}}),$$

and $\phi_{\mathcal{T},\nu} = \xi_{\mathcal{T},\nu}$ for $\nu \in N_{\mathcal{T}}^0$.

Analogously to the case of *discontinuous* piecewise constant trial spaces in $\mathscr{V}$ studied in Sect. 2.3, using the framework of operator preconditioning outlined in Sect. 2.2 we are going to construct a family of preconditioners $G_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{V}_{\mathcal{T}}', \mathscr{V}_{\mathcal{T}})$ of type $D_{\mathcal{T}}^{-1} B_{\mathcal{T}} (D_{\mathcal{T}}')^{-1}$ with uniformly bounded $\|G_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{V}_{\mathcal{T}}',\mathscr{V}_{\mathcal{T}})}$ and $\|\Re(G_{\mathcal{T}})^{-1}\|_{\mathcal{L}(\mathscr{V}_{\mathcal{T}},\mathscr{V}_{\mathcal{T}}')}$.

The roles played in Sect. 2.3 by $|T|$ ($= |\operatorname{supp}\xi_T|$) and $h_T = |T|^{1/d}$, are in this section going to be played by $|\omega_{\mathcal{T}}(\nu)|$ ($= |\operatorname{supp}\xi_{\mathcal{T},\nu}|$) and $h_{\mathcal{T},\nu} := |\omega_{\mathcal{T}}(\nu)|^{1/d}$.

2.4.1 Construction of $\mathscr{W}_{\mathcal{T}}$ and $D_{\mathcal{T}}$

To construct a collection $\Psi_{\mathcal{T}} = \{\psi_{\mathcal{T},\nu} \colon \nu \in N_{\mathcal{T}}\} \subset H^1_{0,\gamma}(\Omega)$ that is biorthogonal to $\Xi_{\mathcal{T}}$, consists of locally supported functions, and for which

$$\mathscr{W}_{\mathcal{T}} := \operatorname{span} \Psi_{\mathcal{T}} \subset \mathscr{W}$$

has an 'approximation property', as in Sect. 2.3 we need two collections $\Sigma_{\mathcal{T}} \subset \mathscr{S}^{0,1}_{\mathcal{T},0}$ and $\Theta_{\mathcal{T}} \subset H^1_{0,\gamma}(\Omega)$ of locally supported functions with $\#\Sigma_{\mathcal{T}} = \#\Theta_{\mathcal{T}} = \#\Xi_{\mathcal{T}}$, where $\Theta_{\mathcal{T}}$ is biorthogonal to $\Xi_{\mathcal{T}}$, and $\Sigma_{\mathcal{T}}$ has an 'approximation property'.

We define $\Sigma_{\mathcal{T}} = \{\sigma_{\mathcal{T},\nu} \colon \nu \in N_{\mathcal{T}}\}$ by $\sigma_{\mathcal{T},\nu} := \phi_{\mathcal{T},\nu}$ when $\nu \in N^0_{\mathcal{T}}$, and $\sigma_{\mathcal{T},\nu} := 0$ when $\nu \in N_{\mathcal{T}} \setminus N^0_{\mathcal{T}}$. Then, obviously, $\sum_{\nu \in N_{\mathcal{T}}} \sigma_{\mathcal{T},\nu}$ equals $\mathbb{1}$ on $\Omega \setminus \cup_{\{T \in \mathcal{T} \colon \overline{T} \cap \gamma \neq \emptyset\}} \overline{T}$.

For constructing $\Theta_{\mathcal{T}}$, on a reference d-simplex $\hat{T}$ for $\varepsilon > 0$ we consider a smooth $\eta_\varepsilon \in [0,1]$, symmetric in the barycentric coordinates, with $\eta_\varepsilon(x) = 0$ when $d(x, \partial\hat{T}) < \varepsilon$, and $\eta_\varepsilon(x) = 1$ when $d(x, \partial\hat{T}) > 2\varepsilon$. Then for some fixed $\varepsilon > 0$ small enough, it holds that

$$\inf_{0 \neq p \in \mathcal{P}_1(\hat{T})} \sup_{0 \neq q \in \mathcal{P}_1(\hat{T})} \frac{\langle p, \eta_\varepsilon q \rangle_{L_2(\hat{T})}}{\|p\|_{L_2(\hat{T})} \|\eta_\varepsilon q\|_{L_2(\hat{T})}} > 0,$$

meaning that the biorthogonal projector $P_\varepsilon \in \mathcal{L}(L_2(\hat{T}), L_2(\hat{T}))$ with $\operatorname{ran} P_\varepsilon = \eta_\varepsilon \mathcal{P}_1(\hat{T})$ and $\operatorname{ran}(\mathrm{Id} - P_\varepsilon) = \mathcal{P}_1(\hat{T})^{\perp_{L_2(\hat{T})}}$ exists. Consequently, with $\Phi_{\hat{T}} = \{\phi_{\hat{T},\nu} \colon \nu \in N_{\hat{T}}\}$ being the nodal basis for $\mathcal{P}_1(\hat{T})$, we have that

$$\{\tilde{\phi}_{\hat{T},\varepsilon,\nu} \colon \nu \in N_{\hat{T}}\} := \langle \Phi_{\hat{T}}, \Phi_{\hat{T}} \rangle^{-1}_{L_2(\hat{T})} P_\varepsilon \Phi_{\hat{T}} \subset H^1_0(\hat{T})$$

is $L_2(\hat{T})$-biorthonormal to $\{\phi_{\hat{T},\nu} \colon \nu \in N_{\hat{T}}\}$.

Now for $T \in \mathcal{T}$, let $F_{\hat{T},T} : T \to \hat{T}$ be an affine bijection. Then $\{\tilde{\phi}_{T,\varepsilon,\nu} \colon \nu \in N_T\}$ defined by

$$\tilde{\phi}_{T,\varepsilon,\nu} := \tfrac{|\hat{T}|}{|T|} \tilde{\phi}_{\hat{T},\varepsilon,F_{\hat{T},T}(\nu)} \tag{2.29}$$

is $L_2(T)$-biorthonormal to the nodal basis for $P_1(T)$.

By selecting for $\nu \in N_{\mathcal{T}}$, a $T(\nu) \in \mathcal{T}$ with $\nu \in N_T$, and defining $\Theta_{\mathcal{T}} = \{\theta_{\mathcal{T},\nu} \colon \nu \in N_{\mathcal{T}}\} \subset H^1_{0,\gamma}(\Omega)$ by

$$\theta_{\mathcal{T},\nu} := |\omega_{\mathcal{T}}(\nu)| \tilde{\phi}_{T(\nu),\varepsilon,\nu},$$

where the specific scaling is chosen for convenience, we have for $\nu, \nu' \in N_{\mathcal{T}}$,

$$\begin{aligned} &\delta_{\nu\nu'} |\omega_{\mathcal{T}}(\nu)| = \langle \theta_{\mathcal{T},\nu}, \xi_{\mathcal{T},\nu'} \rangle_{L_2(\Omega)} \eqsim \delta_{\nu\nu'} \|\theta_{\mathcal{T},\nu}\|_{L_2(\Omega)} \|\xi_{\mathcal{T},\nu'}\|_{L_2(\Omega)}, \\ &\operatorname{supp} \theta_{\mathcal{T},\nu} \subset \overline{T(\nu)}, \ |\theta_{\mathcal{T},\nu}|_{H^1(\Omega)} \lesssim h^{-1}_{\mathcal{T},\nu} \|\theta_{\mathcal{T},\nu}\|_{L_2(\Omega)}, \end{aligned} \tag{2.30}$$

i.e., properties analogous to (2.6), (2.12)(c), and (2.14).

Since furthermore $\Sigma_\mathcal{T}$ and $\Xi_\mathcal{T}$ satisfy properties analogous to (2.12)(a,b), (2.13) and (2.15), defining similarly to (2.8)

(2.31)
$$\psi_{\mathcal{T},\nu} := \sigma_{\mathcal{T},\nu} + \frac{\langle \mathbb{1} - \sigma_{\mathcal{T},\nu}, \xi_{\mathcal{T},\nu}\rangle_{L_2(\Omega)}}{\langle \theta_{\mathcal{T},\nu}, \xi_{\mathcal{T},\nu}\rangle_{L_2(\Omega)}}\theta_{\mathcal{T},\nu} - \sum_{\nu'\in N_\mathcal{T}\setminus\{\nu\}} \frac{\langle \sigma_{\mathcal{T},\nu}, \xi_{\mathcal{T},\nu'}\rangle_{L_2(\Omega)}}{\langle \theta_{\mathcal{T},\nu'}, \xi_{\mathcal{T},\nu'}\rangle_{L_2(\Omega)}}\theta_{\mathcal{T},\nu'}$$
$$= \begin{cases} \frac{\theta_{\mathcal{T},\nu}}{d+1} & \nu \in N_\mathcal{T}\setminus N_\mathcal{T}^0, \\ \phi_{\mathcal{T},\nu} + \frac{d}{(d+2)(d+1)}\theta_{\mathcal{T},\nu} - \sum_{\nu'\in N_\mathcal{T}\setminus\{\nu\}} \frac{|\omega_\mathcal{T}(\nu)\cap\omega_\mathcal{T}(\nu')|}{(d+2)(d+1)|\omega_\mathcal{T}(\nu')|}\theta_{\mathcal{T},\nu'} & \nu \in N_\mathcal{T}^0, \end{cases}$$

we infer that $\sum_{\nu\in N_\mathcal{T}}\psi_{\mathcal{T},\nu}$ equals $\mathbb{1}$ possibly except on a strip along the Dirichlet boundary, and similarly to Theorem 2.3.2, that the biorthogonal projector

$$P_\mathcal{T}\colon u \mapsto \sum_{\nu\in N_\mathcal{T}} \frac{\langle u, \xi_{\mathcal{T},\nu}\rangle_{L_2(\Omega)}}{\langle \mathbb{1}, \xi_{\mathcal{T},\nu}\rangle_{L_2(\Omega)}}\psi_{\mathcal{T},\nu}, \tag{2.32}$$

satisfies $\sup_{\mathcal{T}\in\mathbb{T}}\|P_\mathcal{T}\|_{\mathcal{L}(\mathscr{W},\mathscr{W})} < \infty$. With $(D_\mathcal{T}v)(w) := (Dv)(w)$ $((v,w)\in \mathscr{V}_\mathcal{T}\times\mathscr{W}_\mathcal{T})$, we have $\|D_\mathcal{T}\|_{\mathcal{L}(\mathscr{V}_\mathcal{T},\mathscr{W}_\mathcal{T}')}\le 1$ and $\sup_{\mathcal{T}\in\mathbb{T}}\|D_\mathcal{T}^{-1}\|_{\mathcal{L}(\mathscr{W}_\mathcal{T}',\mathscr{V}_\mathcal{T})}<\infty$, and

$$\boldsymbol{D}_\mathcal{T} = \mathcal{F}'_{\Psi_\mathcal{T}} D_\mathcal{T} \mathcal{F}_{\Xi_\mathcal{T}} = \operatorname{diag}\{\langle \mathbb{1}, \xi_{\mathcal{T},\nu}\rangle_{L_2(\Omega)}\colon \nu\in N_\mathcal{T}\} = \operatorname{diag}\Big\{\tfrac{1}{d+1}|\omega_\mathcal{T}(\nu)|\colon \nu\in N_\mathcal{T}\Big\}.$$

2.4.2 Construction of $B_\mathcal{T}\in\mathcal{L}\mathrm{is}_c(\mathscr{W}_\mathcal{T},\mathscr{W}_\mathcal{T}')$.

Since $\Theta_\mathcal{T}$ additionally satisfies, for $k\in\{0,1\}$,

$$\Big\|\sum_{\nu\in N_\mathcal{T}} c_\nu\theta_{\mathcal{T},\nu}\Big\|^2_{H^k(\Omega)} \eqsim \sum_{\nu\in N_\mathcal{T}} h_{\mathcal{T},\nu}^{-k}\|\theta_{\mathcal{T},\nu}\|^2_{L_2(\Omega)}|c_\nu|^2,$$

where $\|\theta_{\mathcal{T},\nu}\|_{L_2(\Omega)} \overset{(2.30)}{=} |\omega_\mathcal{T}(\nu)|\|\xi_{\mathcal{T},\nu}\|^{-1}_{L_2(\Omega)} \eqsim |\omega_\mathcal{T}(\nu)|^{\frac12}$, and

$$\|u+v\|^2_{H^k(\Omega)} \gtrsim \|u\|^2_{H^k(\Omega)} + \|v\|^2_{H^k(\Omega)} \quad (u\in\mathscr{S}^{0,1}_{\mathcal{T},0},\, v\in\mathscr{B}_\mathcal{T} := \operatorname{span}\Theta_\mathcal{T}).$$

(cf. (2.26)-(2.27)), we construct $B_\mathcal{T}$ analogously as in §2.3.3: Assuming that we have a $B^\mathscr{S}_\mathcal{T}\in\mathcal{L}\mathrm{is}_c(\mathscr{S}^{0,1}_{\mathcal{T},0},(\mathscr{S}^{0,1}_{\mathcal{T},0})')$ available with $\sup_{\mathcal{T}\in\mathbb{T}}\|B^\mathscr{S}_\mathcal{T}\|_{\mathcal{L}(\mathscr{S}^{0,1}_{\mathcal{T},0},(\mathscr{S}^{0,1}_{\mathcal{T},0})')}<\infty$ and $\sup_{\mathcal{T}\in\mathbb{T}}\|\Re(B^\mathscr{S}_\mathcal{T})^{-1}\|_{\mathcal{L}((\mathscr{S}^{0,1}_{\mathcal{T},0})',\mathscr{S}^{0,1}_{\mathcal{T},0})}<\infty$, for some constant $\beta>0$ we take

$$\big(B^\mathscr{B}_\mathcal{T}\sum_{\nu\in N_\mathcal{T}} c_\nu\theta_{\mathcal{T},\nu}\big)\big(\sum_{\nu\in N_\mathcal{T}} d_\nu\theta_{\mathcal{T},\nu}\big) := \beta\sum_{\nu\in N_\mathcal{T}}|\omega_\mathcal{T}(\nu)|^{1-\frac{2s}{d}}c_\nu d_\nu,$$

and

$$B_\mathcal{T} := (I^\mathscr{S}_\mathcal{T}|_{\mathscr{W}_\mathcal{T}})' B^\mathscr{S}_\mathcal{T} I^\mathscr{S}_\mathcal{T}|_{\mathscr{W}_\mathcal{T}} + (I^\mathscr{B}_\mathcal{T}|_{\mathscr{W}_\mathcal{T}})' B^\mathscr{B}_\mathcal{T} I^\mathscr{B}_\mathcal{T}|_{\mathscr{W}_\mathcal{T}},$$

where $I_{\mathcal{T}}^{\mathscr{S}}$ is the linear projector defined on $\mathscr{S}_{\mathcal{T},0}^{0,1} \oplus \mathscr{B}_{\mathcal{T}}$ by $\operatorname{ran} I_{\mathcal{T}}^{\mathscr{S}} = \mathscr{S}_{\mathcal{T},0}^{0,1}$ and $\operatorname{ran} I_{\mathcal{T}}^{\mathscr{B}} = \mathscr{B}_{\mathcal{T}}$, where $I_{\mathcal{T}}^{\mathscr{B}} := \mathrm{Id} - I_{\mathcal{T}}^{\mathscr{S}}$. Then one has $\sup_{\mathcal{T}\in\mathbb{T}} \|B_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{W}_{\mathcal{T}},\mathscr{W}'_{\mathcal{T}})} < \infty$ and $\sup_{\mathcal{T}\in\mathbb{T}} \|\Re(B_{\mathcal{T}})^{-1}\|_{\mathcal{L}(\mathscr{W}'_{\mathcal{T}},\mathscr{W}_{\mathcal{T}})} < \infty$.

Substituting the definition of $\psi_{\mathcal{T},\nu}$, one infers that $\boldsymbol{G}_{\mathcal{T}} = \boldsymbol{D}_{\mathcal{T}}^{-1}\boldsymbol{B}_{\mathcal{T}}\boldsymbol{D}_{\mathcal{T}}^{-\top}$, where

$$\boldsymbol{B}_{\mathcal{T}} = \boldsymbol{p}_{\mathcal{T}}^{\top}\boldsymbol{B}_{\mathcal{T}}^{\mathscr{S}}\boldsymbol{p}_{\mathcal{T}} + \boldsymbol{q}_{\mathcal{T}}^{\top}\boldsymbol{B}_{\mathcal{T}}^{\mathscr{B}}\boldsymbol{q}_{\mathcal{T}},$$

and

$$(\boldsymbol{q}_{\mathcal{T}})_{\nu'\nu} := \begin{cases} \frac{\delta_{\nu'\nu}}{d+1} & \nu \in N_{\mathcal{T}} \setminus N_{\mathcal{T}}^0, \\ \frac{d}{(d+2)(d+1)} & \nu \in N_{\mathcal{T}}^0, \nu' = \nu, \\ -\frac{|\omega_{\mathcal{T}}(\nu)\cap\omega_{\mathcal{T}}(\nu')|}{(d+2)(d+1)|\omega_{\mathcal{T}}(\nu')|} & \nu \in N_{\mathcal{T}}^0, \nu' \neq \nu, \end{cases} \qquad \boldsymbol{B}_{\mathcal{T}}^{\mathscr{S}} := \mathcal{F}'_{\Phi_{\mathcal{T}}} B_{\mathcal{T}}^{\mathscr{S}} \mathcal{F}_{\Phi_{\mathcal{T}}},$$

$$(\boldsymbol{p}_{\mathcal{T}})_{\nu'\nu} := \delta_{\nu'\nu}\ (\nu' \in N_{\mathcal{T}}^0,\ \nu \in N_{\mathcal{T}}), \qquad \boldsymbol{B}_{\mathcal{T}}^{\mathscr{B}} := \operatorname{diag}\{\beta|\omega_{\mathcal{T}}(\nu)|^{1-\frac{2s}{d}} : \nu \in N_{\mathcal{T}}\}.$$

2.4.3 Manifold case.

From Sect. 2.3.2 recall the definitions of Γ, γ, $\mathscr{W}$, $\mathscr{V}$, $\chi\colon \Omega \to \cup_{i=1}^{p}\chi_i(\Omega_i)$, and that of the family of conforming partitions $\mathbb{T}$ of Γ.

As in the domain case discussed in Sect. 2.4.1, for $\mathcal{T} \in \mathbb{T}$ let $N_{\mathcal{T}}$ be the set of vertices of $\mathcal{T}$, and $N_{\mathcal{T}}^0$ its subset of vertices not on γ, for $T \in \mathcal{T}$ let N_T be the vertices of T, $N_{\mathcal{T},T}^0 := N_{\mathcal{T}}^0 \cap N_T$, and for $\nu \in N_{\mathcal{T}}$ let $\omega_{\mathcal{T}}(\nu) := \cup_{\{T\in\mathcal{T}\colon\, \nu\in N_T\}}\overline{T}$.

We take

$$\mathscr{V}_{\mathcal{T}} = \mathscr{S}_{\mathcal{T}}^{0,1} := \{u \in H^1(\Gamma)\colon u\circ\chi|_{\chi^{-1}(T)} \in \mathcal{P}_1\,(T \in \mathcal{T})\} \subset \mathscr{V},$$
$$\mathscr{S}_{\mathcal{T},0}^{0,1} := \{u \in H_{0,\gamma}^1(\Gamma)\colon u\circ\chi|_{\chi^{-1}(T)} \in \mathcal{P}_1\,(T \in \mathcal{T})\},$$

equipped with nodal bases $\Xi_{\mathcal{T}} = \{\xi_{\mathcal{T},\nu}\colon \nu \in N_{\mathcal{T}}\}$ and $\Phi_{\mathcal{T}} = \{\phi_{\mathcal{T},\nu}\colon \nu \in N_{\mathcal{T}}^0\}$, respectively, defined by

$$\xi_{\mathcal{T},\nu}(\nu') = \delta_{\nu,\nu'} \quad (\nu,\nu' \in N_{\mathcal{T}}),$$

and $\phi_{\mathcal{T},\nu} = \xi_{\mathcal{T},\nu}$ for $\nu \in N_{\mathcal{T}}^0$.

Actually exclusively for the deriving an inverse inequality analogous to Lemma 2.3.4, first we construct a collection $\Psi_{\mathcal{T}} = \{\psi_{\mathcal{T},\nu}\colon \nu \in N_{\mathcal{T}}\} \subset H_{0,\gamma}^1(\Gamma)$ that has an 'approximation property' and that is biorthogonal to $\Xi_{\mathcal{T}}$ w.r.t. the *true* $L_2(\Gamma)$-scalar product: We define $\Sigma_{\mathcal{T}} = \{\sigma_{\mathcal{T},\nu}\colon \nu \in N_{\mathcal{T}}\}$ by $\sigma_{\mathcal{T},\nu} := \phi_{\mathcal{T},\nu}$ when $\nu \in N_{\mathcal{T}}^0$, and $\sigma_{\mathcal{T},\nu} := 0$ when $\nu \in N_{\mathcal{T}} \setminus N_{\mathcal{T}}^0$. Then, obviously, $\sum_{\nu\in N_{\mathcal{T}}}\sigma_{\mathcal{T},\nu}$ equals $\mathbb{1}$ on $\Gamma \setminus \cup_{\{T\in\mathcal{T}\colon\, \overline{T}\cap\gamma\neq\emptyset\}}\overline{T}$.

Given a d-simplex $T \subset \mathbb{R}^d$, by means of an affine bijection we transport the function η_ε, defined in Sect. 2.3.2 on a reference d-simplex $\hat{T}$, to a function on T and denote it by $\eta_{T,\varepsilon}$. Then for any panel $T \in \mathcal{T} \in \mathbb{T}$, for some $\varepsilon > 0$ small enough it holds that

$$\inf_{0\neq p\in\mathcal{P}_1(\chi^{-1}(T))} \sup_{0\neq q\in\mathcal{P}_1(\chi^{-1}(T))} \frac{\langle p\circ\chi^{-1}, (\eta_{T,\varepsilon}q)\circ\chi^{-1}\rangle_{L_2(T)}}{\|p\circ\chi^{-1}\|_{L_2(T)}\|(\eta_{T,\varepsilon}q)\circ\chi^{-1}\|_{L_2(T)}} > 0$$

Moreover, since the panels T get increasingly flat when $\operatorname{diam} T \to 0$, there exists an $\varepsilon > 0$ such that above inf-sup condition is satisfied *uniformly* over all $T \in \mathcal{T} \in \mathbb{T}$.

By selecting for each $\nu \in N_{\mathcal{T}}$ a $T(\nu) \in \mathcal{T}$ with $\nu \in N_T$, as in Sect. 2.4.1 we obtain a collection $\Theta_{\mathcal{T}} = \{\theta_{\mathcal{T},\nu} \colon \nu \in N_{\mathcal{T}}\}$ with $\theta_{\mathcal{T},\nu} \subset H^1_0(T(\nu))$ that is biorthogonal to $\Xi_{\mathcal{T}}$, in particular that satisfies (2.30), after which we define the $\psi_{\mathcal{T},\nu}$ by means of formula (2.31). Having constructed the biorthogonal collections $\Xi_{\mathcal{T}}$ and $\Psi_{\mathcal{T}}$, we set the biorthogonal projector $P_{\mathcal{T}} : L_2(\Gamma) \to H^1_{0,\gamma}(\Gamma) \colon u \mapsto \sum_{\nu \in N_{\mathcal{T}}} \frac{\langle u, \xi_{\mathcal{T},\nu}\rangle_{L_2(\Gamma)}}{\langle \mathbb{1}, \xi_{\mathcal{T},\nu}\rangle_{L_2(\Gamma)}} \psi_{\mathcal{T},\nu}$ which satisfies $\|P_{\mathcal{T}} u\|_{H^1(\Gamma)} \lesssim \|h_{\mathcal{T}}^{-1} u\|_{L_2(\Gamma)}$. With the aid of this projector, as in Lemma 2.3.4 one infers that

$$\|h_{\mathcal{T}} v_{\mathcal{T}}\|_{L_2(\Gamma)} \lesssim \|v_{\mathcal{T}}\|_{(H^1_{0,\gamma})'} \quad (v_{\mathcal{T}} \in \mathscr{V}_{\mathcal{T}}). \tag{2.33}$$

Having established this inverse inequality, to arrive at a construction of $\Psi_{\mathcal{T}}$ that does not require the evaluation of integrals over Γ, as in Sect. 2.3.2 we replace $\langle\, ,\, \rangle_{L_2(\Gamma)}$ by $\langle\, ,\, \rangle_{\mathcal{T}}$. We redefine $\Theta_{\mathcal{T}} = \{\theta_{\mathcal{T},\nu} \colon \nu \in N_{\mathcal{T}}\}$ by

$$\theta_{\mathcal{T},\nu} := |\omega_{\mathcal{T}}(\nu)| \tfrac{|\chi^{-1}(T)|}{|T|} \tilde{\phi}_{\chi^{-1}(T),\varepsilon,\chi^{-1}(\nu)} \circ \chi^{-1}$$

with the $\tilde{\phi}$'s defined in (2.29), and following (2.31) set $\Psi_{\mathcal{T}} = \{\psi_{\mathcal{T},\nu} \colon \nu \in N_{\mathcal{T}}\}$ and $\mathscr{W}_{\mathcal{T}} := \operatorname{span} \Psi_{\mathcal{T}}$ by

$$\psi_{\mathcal{T},\nu} = \begin{cases} \frac{\theta_{\mathcal{T},\nu}}{d+1} & \nu \in N_{\mathcal{T}} \setminus N^0_{\mathcal{T}}, \\ \phi_{\mathcal{T},\nu} + \frac{d}{(d+2)(d+1)} \theta_{\mathcal{T},\nu} - \sum_{\nu' \in N_{\mathcal{T}} \setminus \{\nu\}} \frac{|\omega_{\mathcal{T}}(\nu) \cap \omega_{\mathcal{T}}(\nu')|}{(d+2)(d+1)|\omega_{\mathcal{T}}(\nu')|} \theta_{\mathcal{T},\nu'} & \nu \in N^0_{\mathcal{T}}, \end{cases}$$

As in Sect. 2.3.2, we set $(D_{\mathcal{T}} v_{\mathcal{T}})(w_{\mathcal{T}}) := \langle v_{\mathcal{T}}, w_{\mathcal{T}}\rangle_{\mathcal{T}}$ ($v_{\mathcal{T}} \in \mathscr{V}_{\mathcal{T}}$, $w_{\mathcal{T}} \in \mathscr{W}_{\mathcal{T}}$), and as in Sect. 2.3.2, using (2.33) one shows that $\sup_{\mathcal{T} \in \mathbb{T}} \|D_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{V}_{\mathcal{T}}, \mathscr{W}'_{\mathcal{T}})} < \infty$. Similarly as in Lemma 2.3.6, one proves that

$$\|h_{\mathcal{T}} v_{\mathcal{T}}\|_{L_2(\Gamma)} \lesssim \sup_{0 \neq w \in H^1_{0,\gamma}(\Gamma)} \frac{\langle v_{\mathcal{T}}, w_{\mathcal{T}}\rangle_{\mathcal{T}}}{\|w\|_{H^1(\Gamma)}},$$

and with that $\sup_{\mathcal{T} \in \mathbb{T}} \|D_{\mathcal{T}}^{-1}\|_{\mathcal{L}(\mathscr{W}'_{\mathcal{T}}, \mathscr{V}_{\mathcal{T}})} < \infty$.

Constructing $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{W}_{\mathcal{T}}, \mathscr{W}'_{\mathcal{T}})$ as in Sect. 2.4.2, one arrives at the same expressions for $\boldsymbol{D}_{\mathcal{T}}$, $\boldsymbol{G}_{\mathcal{T}}$, $\boldsymbol{B}_{\mathcal{T}}$, $\boldsymbol{q}_{\mathcal{T}}$, $\boldsymbol{B}^{\mathscr{S}}_{\mathcal{T}}$, $\boldsymbol{p}_{\mathcal{T}}$, and $\boldsymbol{B}^{\mathscr{D}}_{\mathcal{T}}$ as in Sect. 2.4.1-2.4.2 in the domain case.

2.5 Higher order case

In this section, we discuss the construction of an uniform preconditioner for $\mathscr{V}_{\mathcal{T}}$ being either the space $\mathscr{S}^{-1,\ell}_{\mathcal{T}}$ of *discontinuous* piecewise polynomials of degree $\ell > 0$ w.r.t. $\mathcal{T}$, or the space $\mathscr{S}^{0,\ell}_{\mathcal{T}}$ of *continuous* piecewise polynomials of degree $\ell > 1$ w.r.t. $\mathcal{T}$.

We write the spaces $\mathscr{V}_\mathcal{T}$, $\mathscr{B}_\mathcal{T}$, $\mathscr{W}_\mathcal{T}$, and their bases $\Xi_\mathcal{T}$, $\Theta_\mathcal{T}$, $\Psi_\mathcal{T}$ from Sect. 2.3 or 2.4 as $\mathscr{V}^0_\mathcal{T}$, $\mathscr{B}^0_\mathcal{T}$, $\mathscr{W}^0_\mathcal{T}$, and $\Xi^0_\mathcal{T}$, $\Theta^0_\mathcal{T}$, $\Psi^0_\mathcal{T}$, respectively. The biorthogonal projector formerly denoted as $P_\mathcal{T}$ will now be denoted as $P^0_\mathcal{T}$, and the matrices $\mathbf{B}_\mathcal{T}$ and $\mathbf{D}_\mathcal{T}$ as $\mathbf{B}^0_\mathcal{T}$ and $\mathbf{D}^0_\mathcal{T}$.

Although we consider the domain case, the results extend to the manifold case following the approach outlined in §2.3.2 or §2.4.3.

In order to construct an uniform preconditioner, obvious possibilities are to apply the framework of operator preconditioning directly to the higher order polynomial space $\mathscr{V}_\mathcal{T}$, or to use the preconditioner developed for the lowest order case within a subspace correction framework. We investigate both possibilities.

2.5.1 Application of the operator preconditioning framework

Discontinuous piecewise polynomials

Given $\ell > 0$, for $\mathcal{T} \in \mathbb{T}$, let $\mathscr{V}_\mathcal{T} = \mathscr{S}^{-1,\ell}_\mathcal{T}$. With $m = m(\ell) := \binom{d+\ell}{\ell} - 1$, we equip $\mathscr{V}_\mathcal{T}$ with $\Xi_\mathcal{T} = \{\xi_{T,i} \colon T \in \mathcal{T},\, 0 \le i \le m\}$, where for each $T \in \mathcal{T}$, $\{\xi_{T,i} \colon 0 \le i \le m\}$ is constructed by the common affine lifting approach from a basis for the polynomials of degree ℓ on a reference d-simplex, such that $\{\xi_{T,0} \colon T \in \mathcal{T}\} = \Xi^0_\mathcal{T}$, $\operatorname{supp} \xi_{T,i} \subset \overline{T}$, and $\|\xi_{T,i}\|_{L_2(\Omega)} \eqsim |T|^{\frac{1}{2}}$.[6]

A straightforward generalization of the construction in the first paragraphs of §2.4.1 of a collection in $H^1_0(T)$ that is biorthogonal to the nodal basis of $P_1(T)$ shows the following: There exists a set of 'bubbles' $\Theta_\mathcal{T} = \{\theta_{T,i} \colon T \in \mathcal{T},\, 0 \le i \le m\} \subset H^1_{0,\gamma}(\Omega)$ such that for $T, T' \in \mathcal{T}$, $0 \le i, i' \le m$, $k \in \{0,1\}$,

$$\delta_{(T,i),(T',i')}|T| = \langle \theta_{T,i}, \xi_{T',i'} \rangle_{L_2(\Omega)} \eqsim \delta_{(T,i),(T',i')} \|\theta_{T,i}\|_{L_2(\Omega)} \|\xi_{T',i'}\|_{L_2(\Omega)}, \tag{2.34}$$

$$\|\theta_{T,i}\|_{H^1(\Omega)} \lesssim h_T^{-1} \|\theta_{T,i}\|_{L_2(\Omega)}, \tag{2.35}$$

$$\operatorname{sup} \theta_{T,i} \subset \overline{T}, \tag{2.36}$$

$$\Big\| \sum_{\{T\in\mathcal{T}, 0\le i\le m\}} c_{T,i}\theta_{T,i} \Big\|^2_{H^k(\Omega)} \eqsim \sum_{\{T\in\mathcal{T}, 0\le i\le m\}} h_T^{-2k} |c_{T,i}|^2 \|\theta_{T,i}\|^2_{L_2(\Omega)}, \tag{2.37}$$

$$\|u+v\|^2_{H^k(\Omega)} \gtrsim \|u\|^2_{H^k(\Omega)} + \|v\|^2_{H^k(\Omega)} \quad (u \in \mathscr{S}^{0,1}_{\mathcal{T},0},\, v \in \mathscr{B}_\mathcal{T} := \operatorname{span} \Theta_\mathcal{T}), \tag{2.38}$$

$$\{\theta_{T,0} \colon T \in \mathcal{T}\} = \Theta^0_\mathcal{T}. \tag{2.39}$$

Writing $\Psi^0_\mathcal{T} = \{\psi^0_{\mathcal{T},T} \colon T \in \mathcal{T}\}$, we define $\Psi_\mathcal{T} := \{\psi_{\mathcal{T},T,i} \colon T \in \mathcal{T},\, 0 \le i \le m\}$, $\mathscr{W}_\mathcal{T} := \operatorname{span} \Psi_\mathcal{T}$ by

$$\psi_{\mathcal{T},T,i} := \begin{cases} \psi^0_{\mathcal{T},T} - \sum_{\{T'\in\mathcal{T}, 1\le i'\le m\}} \frac{\langle \psi^0_{\mathcal{T},T}, \xi_{T',i'}\rangle_{L_2(\Omega)}}{\langle \theta_{T',i'}, \xi_{T',i'}\rangle_{L_2(\Omega)}} \theta_{T',i'} & i = 0, \\ \theta_{T,i} & 1 \le i \le m. \end{cases}$$

[6] For $i > 1$, it is allowed that $\xi_{T,i}$ is (nearly) orthogonal to $\mathbb{1}|_T$, i.e., (2.15) is not required for these $\xi_{T,i}$.

Knowing that $\Psi^0_{\mathcal{T}}$ and $\Xi^0_{\mathcal{T}}$, and $\Theta_{\mathcal{T}}$ and $\Xi_{\mathcal{T}}$ are biorthogonal, the correction made to the $\psi^0_{\mathcal{T},T}$ ensures that $\Psi_{\mathcal{T}}$ and $\Xi_{\mathcal{T}}$ are biorthogonal, in particular that

$$\langle \psi_{\mathcal{T},T,i}, \xi_{T',i'} \rangle = \delta_{(T,i),(T',i')}|T|.$$

For use later, notice that $\mathscr{W}^0_{\mathcal{T}} \subset \mathscr{W}_{\mathcal{T}}$, and that by definition of $\psi^0_{\mathcal{T},T}$ and $\sigma_{\mathcal{T},T}$, for $i' > 0$,

$$\begin{aligned}(\mathbf{R}_{\mathcal{T}})_{(T',i'),T} &:= -\frac{\langle \psi^0_{\mathcal{T},T}, \xi_{T',i'}\rangle_{L_2(\Omega)}}{\langle \theta_{T',i'}, \xi_{T',i'}\rangle_{L_2(\Omega)}} = -|T'|^{-1}\langle \sigma_{\mathcal{T},T}, \xi_{T',i'}\rangle_{L_2(\Omega)} \\ &= -|T'|^{-1} \sum_{\nu \in N^0_{\mathcal{T},T} \cap N^0_{\mathcal{T},T'}} d^{-1}_{\mathcal{T},\nu} \langle \phi_{\mathcal{T},\nu}, \xi_{T',i'}\rangle_{L_2(\Omega)}.\end{aligned} \tag{2.40}$$

The biorthogonal projector $P_{\mathcal{T}}\colon L_2(\Omega) \to H^1_{0,\gamma}(\Omega)$ with $\operatorname{ran} P_{\mathcal{T}} = \mathscr{W}_{\mathcal{T}}$ and $\operatorname{ran}(\mathrm{Id} - P_{\mathcal{T}}) = \mathscr{V}_{\mathcal{T}}^{\perp_{L_2(\Omega)}}$ is given by

$$P_{\mathcal{T}} u = \sum_{\{T\in\mathcal{T}, 0\le i\le m\}} \frac{\langle u, \xi_{T,i}\rangle_{L_2(\Omega)}}{\langle \psi_{\mathcal{T},T,i}, \xi_{T,i}\rangle_{L_2(\Omega)}} \psi_{\mathcal{T},T,i}.$$

Theorem 2.5.1. *It holds that* $\sup_{\mathcal{T}\in\mathbb{T}} \|P_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{W},\mathscr{W})} < \infty$.

Proof. For $T'' \in \mathcal{T}$, $k \in \{0,1\}$, from $\operatorname{supp}\xi_{T,i} \subset \overline{T}$ and $\operatorname{supp}\psi_{\mathcal{T},T'',i} \subset \omega^{(1)}_{\mathcal{T}}(T'')$ we have

$$\|P_{\mathcal{T}} u\|_{H^k(T'')} \le \sum_{\{T\in R^{(1)}_{\mathcal{T}}(T''),\, 0\le i\le m\}} \|u\|_{L_2(T)} \frac{\|\psi_{\mathcal{T},T,i}\|_{H^k(\Omega)}\|\xi_{T,i}\|_{L_2(\Omega)}}{|\langle \psi_{\mathcal{T},T,i}, \xi_{T,i}\rangle_{L_2(\Omega)}|}. \tag{2.41}$$

To bound the right-hand side we distinguish between terms with $i = 0$ and those with $i > 0$.

From $\|\psi^0_{\mathcal{T},T}\|_{H^k(\Omega)} \lesssim h_T^{d/2-k}$ ((2.17)), and for $T' \in \mathcal{T}$, $1 \le i' \le m$,

$$\Big\| \frac{\langle \psi^0_{\mathcal{T},T}, \xi_{T',i'}\rangle_{L_2(\Omega)}}{\langle \theta_{T',i'}, \xi_{T',i'}\rangle_{L_2(\Omega)}} \theta_{T',i'} \Big\|_{H^k(\Omega)} \lesssim \frac{\|\psi^0_{\mathcal{T},T}\|_{L_2(\Omega)}\|\xi_{T',i'}\|_{L_2(\Omega)}}{\|\theta_{T',i'}\|_{L_2(\Omega)}\|\xi_{T',i'}\|_{L_2(\Omega)}} h_{T'}^{-k} \|\theta_{T',i'}\|_{L_2(\Omega)} \lesssim h_T^{d/2-k},$$

we infer that

$$\|\psi_{\mathcal{T},T,0}\|_{H^k(\Omega)} \lesssim h_T^{d/2-k},$$

while, thanks to (2.15),

$$\frac{\|\xi_{T,0}\|_{L_2(\Omega)}}{|\langle \psi_{\mathcal{T},T,0}, \xi_{T,0}\rangle_{L_2(\Omega)}|} = \frac{\|\xi_{T,0}\|_{L_2(\Omega)}}{|\langle \psi^0_{\mathcal{T},T}, \xi_{T,0}\rangle_{L_2(\Omega)}|} = \frac{\|\xi_{T,0}\|_{L_2(\Omega)}}{|\langle \mathbb{1}, \xi_{T,0}\rangle_{L_2(\Omega)}|} \eqsim h_T^{-d/2}.$$

For $1 \le i \le m$,

$$\frac{\|\psi_{\mathcal{T},T,i}\|_{H^k(\Omega)}\|\xi_{T,i}\|_{L_2(\Omega)}}{|\langle \psi_{\mathcal{T},T,i}, \xi_{T,i}\rangle_{L_2(\Omega)}|} = \frac{\|\theta_{T,i}\|_{H^k(\Omega)}\|\xi_{T,i}\|_{L_2(\Omega)}}{|\langle \theta_{T,i}, \xi_{T,i}\rangle_{L_2(\Omega)}|} \eqsim \frac{\|\theta_{T,i}\|_{H^k(\Omega)}}{\|\theta_{T,i}\|_{L_2(\Omega)}} \lesssim h_T^{-k}.$$

Combining the above inequality with (2.41) shows that

$$\|P_{\mathcal{T}}u\|_{H^k(T'')} \lesssim h_T^{-k}\|u\|_{L_2(\omega_{\mathcal{T}}^{(1)}(T''))}, \tag{2.42}$$

which is the analogue of estimate (2.18) that was proven for $P_{\mathcal{T}}^0$. Since $\mathscr{W}_{\mathcal{T}}^0 \subset \mathscr{W}_{\mathcal{T}}$, by making use of the *same* Scott-Zhang type quasi-interpolator $\Pi_{\mathcal{T}}$ as has been used in the proof of Theorem 2.3.2, copying the remainder of that proof yields the claim. □

Thanks this theorem, $D_{\mathcal{T}}\colon \mathscr{V}_{\mathcal{T}} \to \mathscr{W}_{\mathcal{T}}'$ defined by $(D_{\mathcal{T}}v)(w) = \langle v, w\rangle_{L_2(\Omega)}$ satisfies $\sup_{\mathcal{T}\in\mathbb{T}} \max\big(\|D_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{V}_{\mathcal{T}},\mathscr{W}_{\mathcal{T}}')}, \|D_{\mathcal{T}}^{-1}\|_{\mathcal{L}(\mathscr{W}_{\mathcal{T}}',\mathscr{V}_{\mathcal{T}})}\big) < \infty$. By enumerating all $\xi_{T,i}$ and $\psi_{\mathcal{T},T,i}$ with index $i = 0$ before those with index $i > 0$, its matrix representation reads as $\mathbf{D}_{\mathcal{T}} = \begin{bmatrix} \mathbf{D}_{\mathcal{T}}^0 & 0 \\ 0 & \mathbf{D}_{\mathcal{T}}^1 \end{bmatrix}$, where $\mathbf{D}_{\mathcal{T}}^1 := \operatorname{diag}\{|T|\mathrm{Id}_{m\times m}\colon T \in \mathcal{T}\}$.

Thanks to (2.37)-(2.38), a suitable $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{W}_{\mathcal{T}}, \mathscr{W}_{\mathcal{T}}')$ can be defined similarly as in §2.3.3: With $I_{\mathcal{T}}^{\mathscr{S}}$ being the linear projector defined on $\mathscr{S}_{\mathcal{T},0}^{0,1} \oplus \mathscr{B}_{\mathcal{T}}$ by $\operatorname{ran} I_{\mathcal{T}}^{\mathscr{S}} = \mathscr{S}_{\mathcal{T},0}^{0,1}$ and $\operatorname{ran} I_{\mathcal{T}}^{\mathscr{B}} = \mathscr{B}_{\mathcal{T}}$, where $I_{\mathcal{T}}^{\mathscr{B}} := \mathrm{Id} - I_{\mathcal{T}}^{\mathscr{S}}$, we define $B_{\mathcal{T}}^{\mathscr{S}\oplus\mathscr{B}} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}_{\mathcal{T},0}^{0,1} \oplus \mathscr{B}_{\mathcal{T}}, (\mathscr{S}_{\mathcal{T},0}^{0,1} \oplus \mathscr{B}_{\mathcal{T}})')$ by

$$(B_{\mathcal{T}}^{\mathscr{S}\oplus\mathscr{B}}w)(\tilde{w}) = (B_{\mathcal{T}}^{\mathscr{S}} I_{\mathcal{T}}^{\mathscr{S}} w)(I_{\mathcal{T}}^{\mathscr{S}}\tilde{w}) + (B_{\mathcal{T}}^{\mathscr{B}} I_{\mathcal{T}}^{\mathscr{B}} w)(I_{\mathcal{T}}^{\mathscr{B}}\tilde{w}),$$

where

$$(B_{\mathcal{T}}^{\mathscr{B}} \sum_{\{T\in\mathcal{T},\, 0\le i\le m\}} c_{T,i}\theta_{\mathcal{T},i})(\sum_{\{T\in\mathcal{T},\, 0\le i\le m\}} d_{T,i}\theta_{\mathcal{T},i}) := \beta \sum_{\{T\in\mathcal{T},\, 0\le i\le m\}} h_T^{d-2s} c_{T,i}d_{T,i},$$

and $B_{\mathcal{T}}^{\mathscr{S}}$ is as in §2.3.3, and define $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{W}_{\mathcal{T}}, \mathscr{W}_{\mathcal{T}}')$ by $(B_{\mathcal{T}}w)(\tilde{w}) = (B_{\mathcal{T}}^{\mathscr{S}\oplus\mathscr{B}}w)(\tilde{w})$.

Using that with $\mathbf{R}_{\mathcal{T}}$ as defined in (2.40), $\begin{bmatrix} \mathrm{Id} & 0 \\ \mathbf{R}_{\mathcal{T}} & \mathrm{Id} \end{bmatrix}$ is the basis transformation from $\Psi_{\mathcal{T}}$ to $\Psi_{\mathcal{T}}^0 \cup \{\theta_{T,i}\colon T \in \mathcal{T}, 1 \le i \le m\}$, one infers that the representation of the resulting uniform preconditioner reads as

$$\mathbf{G}_{\mathcal{T}} = \begin{bmatrix} \mathbf{D}_{\mathcal{T}}^0 & 0 \\ 0 & \mathbf{D}_{\mathcal{T}}^1 \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{Id} & \mathbf{R}_{\mathcal{T}}^\top \\ 0 & \mathrm{Id} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{\mathcal{T}}^0 & 0 \\ 0 & \beta(\mathbf{D}_{\mathcal{T}}^1)^{1-\frac{2s}{d}} \end{bmatrix} \begin{bmatrix} \mathrm{Id} & 0 \\ \mathbf{R}_{\mathcal{T}} & \mathrm{Id} \end{bmatrix} \begin{bmatrix} \mathbf{D}_{\mathcal{T}}^0 & 0 \\ 0 & \mathbf{D}_{\mathcal{T}}^1 \end{bmatrix}^{-1}, \tag{2.43}$$

which is thus independent of the particular bubbles $\Theta_{\mathcal{T}}$ being chosen.

Continuous piecewise polynomials

Given $\ell > 1$, for $\mathcal{T} \in \mathbb{T}$, let $\mathscr{V}_{\mathcal{T}} = \mathscr{S}_{\mathcal{T}}^{0,\ell}$. We equip $\mathscr{V}_{\mathcal{T}}$ with a basis $\Xi_{\mathcal{T}} = \Xi_{\mathcal{T}}^0 \cup (\Xi_{\mathcal{T}} \setminus \Xi_{\mathcal{T}}^0)$, where for each $T \in \mathcal{T}$, the set of restrictions to T of those basis functions that do not identically vanish on T is a lifted version of a fixed basis for

the polynomials of degree ℓ on a reference d-simplex under an affine bijection; the support of each basis function $\xi \in \Xi_{\mathcal{T}}$ is connected and extends to a uniformly bounded number of $T \in \mathcal{T}$; and finally, $\|\xi\|_{L_2(\Omega)} \eqsim |\operatorname{supp}\xi|^{\frac{1}{2}} \eqsim |T|^{\frac{1}{2}}$ for some $T \in \mathcal{T}$ with $\operatorname{supp}\xi \cap T \neq \emptyset$.

Similar to the previous §2.5.1, a biorthogonal collection $\Theta_{\mathcal{T}} = \{\theta(\xi)\colon \xi \in \Xi_{\mathcal{T}}\}$ of 'bubbles' exists that has properties analogous to (2.34)-(2.39), reading $|T|$ as $|\operatorname{supp}\xi|$, and h_T as $|\operatorname{supp}\xi|^{1/d}$. Writing the collection $\Psi^0_{\mathcal{T}}$ as found in §2.4 as $\{\psi^0(\xi)\colon \xi \in \Xi^0_{\mathcal{T}}\}$, we define $\Psi_{\mathcal{T}} = \{\psi(\xi)\colon \xi \in \Xi_{\mathcal{T}}\}$, biorthogonal to $\Xi_{\mathcal{T}}$, and $\mathscr{W}_{\mathcal{T}} := \operatorname{span}\Psi_{\mathcal{T}}$ by

$$\psi(\xi) := \begin{cases} \psi^0(\xi) - \sum_{\xi' \in \Xi_{\mathcal{T}} \setminus \Xi^0_{\mathcal{T}}} \frac{\langle \psi^0(\xi), \xi' \rangle_{L_2(\Omega)}}{\langle \theta(\xi'), \xi' \rangle_{L_2(\Omega)}} \theta(\xi') & \xi \in \Xi^0_{\mathcal{T}}, \\ \theta(\xi) & \xi \in \Xi_{\mathcal{T}} \setminus \Xi^0_{\mathcal{T}}. \end{cases}$$

Theorem 2.5.1 extends to the current setting and shows that the corresponding biorthogonal projector is uniformly bounded. The representation of resulting uniform preconditioner reads as (2.43), obviously now with $\mathbf{D}^0_{\mathcal{T}}$ and $\mathbf{B}^0_{\mathcal{T}}$ as found in the continuous piecewise linear case, and the matrix $\mathbf{D}^1_{\mathcal{T}} := \operatorname{diag}\{|\operatorname{supp}\xi|\colon \xi \in \Xi_{\mathcal{T}} \setminus \Xi^0_{\mathcal{T}}\}$, and $(\mathbf{R}^1_{\mathcal{T}})_{\xi,\nu} := -|\operatorname{supp}\xi|^{-1}\langle \phi_{\mathcal{T},\nu}, \xi\rangle_{L_2(\Omega)}$ for $(\xi,\nu) \in (\Xi_{\mathcal{T}} \setminus \Xi^0_{\mathcal{T}}) \times N^0_{\mathcal{T}}$, and $(\mathbf{R}^1_{\mathcal{T}})_{\xi,\nu} = 0$ for $\nu \in N_{\mathcal{T}} \setminus N^0_{\mathcal{T}}$ (recall $\#\Xi^0_{\mathcal{T}} = \#N_{\mathcal{T}}$).

2.5.2 Application of a subspace correction framework

For $\mathscr{V}_{\mathcal{T}} = \mathscr{S}^{-1,\ell}_{\mathcal{T}}$, in §2.5.1 we have demonstrated existence of a biorthogonal projector that satisfies (2.42). In §2.5.1 we have shown that a similar result holds true for $\mathscr{V}_{\mathcal{T}} = \mathscr{S}^{0,\ell}_{\mathcal{T}}$. From the proof of Lemma 2.3.4 we learn that this implies that for either choice of $\mathscr{V}_{\mathcal{T}}$,

$$\|h_{\mathcal{T}} v_{\mathcal{T}}\|_{L_2(\Omega)} \lesssim \|v_{\mathcal{T}}\|_{(H^1_{0,\gamma})'} \quad (v_{\mathcal{T}} \in \mathscr{V}_{\mathcal{T}}). \tag{2.44}$$

Using this inverse inequality, we are going to decompose $\mathscr{V}_{\mathcal{T}}$ in a uniformly stable way into $\mathscr{V}^0_{\mathcal{T}}$ and a complement space on which the $\|\ \|_{\mathscr{V}}$-norm is equivalent to a scaled $L_2(\Omega)$-norm.

Proposition 2.5.2. *Let $Q^0_{\mathcal{T}} \in \mathcal{L}(L_2(\Omega), L_2(\Omega))$ be a projector with $\operatorname{ran} Q^0_{\mathcal{T}} = \mathscr{V}^0_{\mathcal{T}}$, $\sup_{\mathcal{T}\in\mathbb{T}} \|Q^0_{\mathcal{T}}\|_{\mathcal{L}(L_2(\Omega),L_2(\Omega))} < \infty$, and*

$$\sup_{\mathcal{T}\in\mathbb{T}} \|h_{\mathcal{T}}^{-1}(\mathrm{Id} - (Q^0_{\mathcal{T}})^*)\|_{\mathcal{L}(H^1_{0,\gamma}(\Omega),L_2(\Omega))} < \infty.$$

Then $\sup_{\mathcal{T}\in\mathbb{T}} \|Q^0_{\mathcal{T}}|_{\mathscr{V}_{\mathcal{T}}}\|_{\mathcal{L}(\mathscr{V}_{\mathcal{T}},\mathscr{V}_{\mathcal{T}})} < \infty$, and

$$\|h^s_{\mathcal{T}} \cdot\|_{L_2(\Omega)} \eqsim \|\cdot\|_{\mathscr{V}} \quad \text{on } \mathscr{V}^1_{\mathcal{T}} := \operatorname{ran}(\mathrm{Id} - Q^0_{\mathcal{T}})|_{\mathscr{V}_{\mathcal{T}}}. \tag{2.45}$$

Proof. For $u \in \mathscr{V}_\mathcal{T}$, thanks to (2.44) we have

$$\|(\mathrm{Id} - Q^0_\mathcal{T})u\|_{H^1_{0,\gamma}(\Omega)'} = \sup_{v \in H^1_{0,\gamma}(\Omega)} \frac{\langle u, (\mathrm{Id} - (Q^0_\mathcal{T})^*)v\rangle_{L_2(\Omega)}}{\|v\|_{H^1(\Omega)}} \lesssim \|h_\mathcal{T} u\|_{L_2(T)} \lesssim \|u\|_{H^1_{0,\gamma}(\Omega)'},$$

so that the first statement follows by interpolation.

Again by interpolation, the second statement needs only to be proven for $s = 1$. For that case it follows from the above equation and (2.44). □

Next we use the decomposition $\mathscr{V}_\mathcal{T} = \mathscr{V}^0_\mathcal{T} \oplus \mathscr{V}^1_\mathcal{T}$ from Proposition 2.5.2 to build a preconditioner on $\mathscr{V}_\mathcal{T}$ from preconditioners on the subspaces.

Proposition 2.5.3. *In the situation of Proposition 2.5.2, for $i = 0, 1$, let $I^i_\mathcal{T}$ denote the embedding of $\mathscr{V}^i_\mathcal{T}$ into $\mathscr{V}_\mathcal{T}$, and let $G^i_\mathcal{T} \in \mathcal{L}\mathrm{is}_c((\mathscr{V}^i_\mathcal{T})', \mathscr{V}^i_\mathcal{T})$. Then $G_\mathcal{T} := \sum_{i=0}^1 I^i_\mathcal{T} G^i_\mathcal{T} (I^i_\mathcal{T})' \in \mathcal{L}\mathrm{is}_c((\mathscr{V}_\mathcal{T})', \mathscr{V}_\mathcal{T})$ with*

$$\|G_\mathcal{T}\|_{\mathcal{L}(\mathscr{V}'_\mathcal{T}, \mathscr{V}_\mathcal{T})} \le 2 \max_i \|G^i_\mathcal{T}\|_{\mathcal{L}((\mathscr{V}^i_\mathcal{T})', \mathscr{V}^i_\mathcal{T})},$$
$$\|\Re(G_\mathcal{T})^{-1}\|_{\mathcal{L}(\mathscr{V}_\mathcal{T}, \mathscr{V}'_\mathcal{T})} \le 2\|Q^0_\mathcal{T}|_{\mathscr{V}_\mathcal{T}}\|^2_{\mathcal{L}(\mathscr{V}_\mathcal{T}, \mathscr{V}_\mathcal{T})} \max_i \|\Re(G^i_\mathcal{T})^{-1}\|_{\mathcal{L}(\mathscr{V}^i_\mathcal{T}, (\mathscr{V}^i_\mathcal{T})')}$$

Proof. The result follows as an easy case from the general theory of (additive) subspace correction methods (e.g. [Osw94] + references cited there), together with the inequalities

$$\frac{1}{2}\|\cdot\|^2_\mathscr{V} \le \|Q^0_\mathcal{T} \cdot\|^2_\mathscr{V} + \|(\mathrm{Id} - Q^0_\mathcal{T})\cdot\|^2_\mathscr{V} \le 2\|Q^0_\mathcal{T}\|_{\mathcal{L}(\mathscr{V}_\mathcal{T}, \mathscr{V}_\mathcal{T})}\|\cdot\|^2_\mathscr{V} \quad \text{on } \mathscr{V}_\mathcal{T},$$

where we used that $\|(\mathrm{Id} - Q^0_\mathcal{T})|_{\mathscr{V}_\mathcal{T}}\|_{\mathcal{L}(\mathscr{V}_\mathcal{T}, \mathscr{V}_\mathcal{T})} = \|Q^0_\mathcal{T}|_{\mathscr{V}_\mathcal{T}}\|_{\mathcal{L}(\mathscr{V}_\mathcal{T}, \mathscr{V}_\mathcal{T})}$, because $Q^0_\mathcal{T}|_{\mathscr{V}_\mathcal{T}}$ is a projector unequal to both the zero map and the identity ([Kat60, XZ03]) □

On $\mathscr{V}^0_\mathcal{T}$ we already have our uniform preconditioner $G^0_\mathcal{T}$ available, so it remains to construct such a preconditioner on the complement space $\mathscr{V}^1_\mathcal{T}$. In the situation of Proposition 2.5.2, let $\Xi_\mathcal{T} = \Xi^0_\mathcal{T} \cup (\Xi_\mathcal{T} \setminus \Xi^0_\mathcal{T})$ be a basis for $\mathscr{V}_\mathcal{T}$ such that $\Xi^1_\mathcal{T} := \Xi_\mathcal{T} \setminus \Xi^0_\mathcal{T}$ is a basis for $\mathscr{V}^1_\mathcal{T}$ for which

$$\|h^s_\mathcal{T} \sum_{\xi \in \Xi^1_\mathcal{T}} c_\xi \xi\|^2_{L_2(\Omega)} \eqsim \sum_{\xi \in \Xi^1_\mathcal{T}} |c_\xi|^2 \|h^s_\mathcal{T} \xi\|^2_{L_2(\Omega)}. \tag{2.46}$$

Then

$$(H^1_\mathcal{T} \sum_{\xi \in \Xi^1_\mathcal{T}} c_\xi \xi)(\sum_{\xi \in \Xi^1_\mathcal{T}} d_\xi \xi) := \sum_{\xi \in \Xi^1_\mathcal{T}} c_\xi d_\xi \|h^s_\mathcal{T} \xi\|^2_{L_2(\Omega)}$$

satisfies $\sup_{\mathcal{T}\in\mathbb{T}} \max(\|H^1_\mathcal{T}\|_{\mathcal{L}(\mathscr{V}^1_\mathcal{T},(\mathscr{V}^1_\mathcal{T})')}, \|\Re(H^1_\mathcal{T})^{-1}\|_{\mathcal{L}((\mathscr{V}^1_\mathcal{T})',\mathscr{V}_\mathcal{T})}) < \infty$ thanks to (2.45), and so $G^1_\mathcal{T} := (H^1_\mathcal{T})^{-1}$ is a suitable choice. The implementation of the resulting uniform preconditioner $G_\mathcal{T}$ reads as

$$\mathbf{G}_\mathcal{T} = \begin{bmatrix} \mathbf{G}^0_\mathcal{T} & 0 \\ 0 & \operatorname{diag}\{\|h^s_\mathcal{T}\xi\|^{-2}_{L_2(\Omega)} : \xi \in \Xi^1_\mathcal{T}\} \end{bmatrix}. \tag{2.47}$$

What remains is, for both options for $\mathscr{V}_\mathcal{T}$, to specify a $Q^0_\mathcal{T}$ that satisfies the conditions from Proposition 2.5.2, and to equip $\mathscr{V}_\mathcal{T}$ with a basis $\Xi_\mathcal{T}$ that is the union of the basis $\Xi^0_\mathcal{T}$ for $\mathscr{V}^0_\mathcal{T}$, and a basis $\Xi^1_\mathcal{T}$ for $\mathscr{V}^1_\mathcal{T} = \operatorname{ran}(\mathrm{Id} - Q^0_\mathcal{T})|_{\mathscr{V}_\mathcal{T}}$, the latter being *uniformly stable* w.r.t. $\|h^s_\mathcal{T} \cdot\|_{L_2(\Omega)}$, i.e., one that satisfies (2.46).

Discontinuous piecewise polynomials

For $\mathscr{V}_\mathcal{T} = \mathscr{S}^{-1,\ell}_\mathcal{T}$, the conditions of Proposition 2.5.2 are fulfilled by taking $Q^0_\mathcal{T}$ to be the $L_2(\Omega)$-orthogonal projector onto $\mathscr{V}^0_\mathcal{T} = \mathscr{S}^{-1,0}_\mathcal{T}$. [7]

By taking $\Xi_\mathcal{T} = \{\xi_{T,i}\}_{T\in\mathcal{T},0\le i\le m}$ to be an $L_2(\Omega)$-orthogonal basis for $\mathscr{V}_\mathcal{T}$ such that $\xi_{\mathcal{T},0} = \xi_T$ and $\operatorname{supp}\xi_{\mathcal{T},i} \subset \overline{T}$, (2.46) is valid.

Remarkably, with these specifications and by scaling $\|\xi_{\mathcal{T},i}\|_{L_2(\Omega)} = |T|^{\frac{1}{2}}$, the resulting $\mathbf{G}_\mathcal{T}$ is given by (2.43) with $\mathbf{R}_\mathcal{T}$ reading as the zero map ($\mathbf{R}_\mathcal{T}$ from (2.40) is non-zero).

Continuous piecewise polynomials

Let $\mathscr{V}_\mathcal{T} = \mathscr{S}^{0,\ell}_\mathcal{T}$ for $\ell > 1$. It is no option to take $Q^0_\mathcal{T}$ to be the $L_2(\Omega)$-orthogonal projector onto $\mathscr{V}^0_\mathcal{T} = \mathscr{S}^{0,1}_\mathcal{T}$ because in that case we will not be able to equip $\operatorname{ran}(\mathrm{Id} - Q^0_\mathcal{T})|_{\mathscr{V}_\mathcal{T}}$ with a locally supported basis.

From (2.32) recall the biorthogonal projector $P^0_\mathcal{T}$ onto $\mathscr{W}^0_\mathcal{T}$ with $\operatorname{ran}(\mathrm{Id} - P^0_\mathcal{T}) = \mathscr{V}^0_\mathcal{T}$. Writing $(\mathrm{Id}-P^0_\mathcal{T}) = (\mathrm{Id}-P^0_\mathcal{T})(\mathrm{Id}-\Pi^0_\mathcal{T})$ with $\Pi^0_\mathcal{T}$ being a Scott-Zhang type interpolator, one shows that $\sup_{\mathcal{T}\in\mathbb{T}} \|h^{-1}_\mathcal{T}(\mathrm{Id} - P^0_\mathcal{T})\|_{\mathcal{L}(H^1_{0,\gamma}(\Omega),L_2(\Omega))} < \infty$. Since furthermore $\sup_{\mathcal{T}\in\mathbb{T}} \|P^0_\mathcal{T}\|_{\mathcal{L}(L_2(\Omega),L_2(\Omega))} < \infty$, the conditions of Proposition 2.5.2 are satisfied by taking $Q^0_\mathcal{T} := (P^0_\mathcal{T})^* = u \mapsto \sum_{\nu\in N_\mathcal{T}} \frac{(d+1)\langle u,\psi^0_{\mathcal{T},\nu}\rangle_{L_2(\Omega)}}{|\omega_\mathcal{T}(\nu)|}\xi^0_{\mathcal{T},\nu}$.

Let us denote the *weighted* $L_2(\Omega)$-norm $\|h^s_\mathcal{T} \cdot\|_{L_2(\Omega)}$ by $|\!|\!| \cdot |\!|\!|$. We need to equip $\mathscr{V}^1_\mathcal{T} = \operatorname{ran}(\mathrm{Id} - Q^0_\mathcal{T})|_{\mathscr{V}_\mathcal{T}}$ with a basis that is uniformly stable w.r.t. $|\!|\!| \cdot |\!|\!|$. Since the supports of the basis functions will extend to multiple $T \in \mathcal{T}$, this task is more complex than for the discontinuous piecewise polynomial case. Let $\Xi_\mathcal{T} = \Xi^0_\mathcal{T} \cup (\Xi_\mathcal{T} \setminus \Xi^0_\mathcal{T})$ be a basis for $\mathscr{V}_\mathcal{T}$ such that for each $T \in \mathcal{T}$, $\{\xi|_T : \xi \in \Xi, \xi|_T \not\equiv 0\}$ is a uniformly $L_2(T)$-stable basis for its span. Common affine equivalent constructions yield such a basis. Then $\Xi_\mathcal{T}$ is stable w.r.t. $|\!|\!| \cdot |\!|\!|$, uniformly in $\mathcal{T} \in \mathbb{T}$, and so in particular, with $\bar{\mathscr{V}}^1_\mathcal{T} := \operatorname{span}\Xi_\mathcal{T} \setminus \Xi^0_\mathcal{T}$, $\mathscr{V}_\mathcal{T} = \mathscr{V}^0_\mathcal{T} \oplus \bar{\mathscr{V}}^1_\mathcal{T}$ is a uniformly stable decomposition w.r.t. $|\!|\!| \cdot |\!|\!|$.

[7] Notice that for $s \ge \frac{1}{2}$, $Q^0_\mathcal{T} \notin \mathcal{L}(\mathscr{V},\mathscr{V})$, so taking its restriction to $\mathscr{V}_\mathcal{T}$ is essential in Proposition 2.5.2.

Corresponding to this decomposition, for $v \in \mathscr{V}_\mathcal{T}$ we write $v = v^0 + \bar{v}^1$. Taking $v \in \mathscr{V}_\mathcal{T}^1$, it holds that $0 = Q_\mathcal{T}^0 v = Q_\mathcal{T}^0 v^0 + Q_\mathcal{T}^0 \bar{v}^1 = v^0 + Q_\mathcal{T}^0 \bar{v}^1$, or $v^0 = -Q_\mathcal{T}^0 \bar{v}^1$ i.e. $v = (\mathrm{Id} - Q_\mathcal{T}^0)\bar{v}^1$, showing that $\mathrm{Id} - Q_\mathcal{T}^0 \colon \bar{\mathscr{V}}_\mathcal{T}^1 \to \mathscr{V}_\mathcal{T}^1$ is surjective. Injectivity of this map follows from $|\!|\!| v |\!|\!| \eqsim |\!|\!| Q_\mathcal{T}^0 \bar{v}^1 |\!|\!| + |\!|\!| \bar{v}^1 |\!|\!| \geq |\!|\!| \bar{v}^1 |\!|\!|$, and bounded invertibility, uniformly in $\mathcal{T}$, will follow from $Q_\mathcal{T}^0$ being uniformly bounded w.r.t. $|\!|\!| \cdot |\!|\!|$. The latter holds true because of $\|\psi_{\mathcal{T},\nu}^0\|_{L_2(\Omega)} \|\xi_{\mathcal{T},\nu}^0\|_{L_2(\Omega)} \eqsim \frac{|\omega_\mathcal{T}(\nu)|}{d+1}$, the local supports of the $\psi_{\mathcal{T},\nu}^0$ and $\xi_{\mathcal{T},\nu}^0$, and the uniform K-mesh property of $\mathcal{T}$. We conclude that $(\mathrm{Id} - Q_\mathcal{T}^0)(\Xi_\mathcal{T} \setminus \Xi_\mathcal{T}^0)$ is a basis for $\mathscr{V}_\mathcal{T}^1$ that is uniformly stable w.r.t. $|\!|\!| \cdot |\!|\!|$.

Since $\Xi_\mathcal{T}^0 \cup (\mathrm{Id} - Q_\mathcal{T}^0)(\Xi_\mathcal{T} \setminus \Xi_\mathcal{T}^0)$ is not the basis of choice to set up the stiffness matrix, we give the implementation $\mathbf{G}_\mathcal{T}$ of the uniform preconditioner $G_\mathcal{T}$ for $\mathscr{V}_\mathcal{T}$ being equipped with $\Xi_\mathcal{T}$, partitioned into $\Xi_\mathcal{T}^0 = \{\xi_{\mathcal{T},\nu} \colon \nu \in N_\mathcal{T}\}$ and $\Xi_\mathcal{T} \setminus \Xi_\mathcal{T}^0$. It reads as

$$\mathbf{G}_\mathcal{T} = \begin{bmatrix} \mathrm{Id} & \mathbf{S}_\mathcal{T} \\ 0 & \mathrm{Id} \end{bmatrix} \begin{bmatrix} \mathbf{G}_\mathcal{T}^0 & 0 \\ 0 & \mathrm{diag}\{\|h_\mathcal{T}^s \xi\|_{L_2(\Omega)}^{-2} \colon \xi \in \Xi_\mathcal{T} \setminus \Xi_\mathcal{T}^0\} \end{bmatrix} \begin{bmatrix} \mathrm{Id} & 0 \\ \mathbf{S}_\mathcal{T}^\top & \mathrm{Id} \end{bmatrix},$$

where for $(\nu, \xi) \in N_\mathcal{T} \times (\Xi_\mathcal{T} \setminus \Xi_\mathcal{T}^0)$, $(\mathbf{S}_\mathcal{T})_{\nu\xi} := -\frac{(d+1)\langle \xi, \psi_{\mathcal{T},\nu}^0 \rangle_{L_2(\Omega)}}{|\omega_\mathcal{T}(\nu)|}$, and where in (2.47) we have replaced $\|h_\mathcal{T}^s (\mathrm{Id} - Q_\mathcal{T}^0)\xi\|_{L_2(\Omega)}^{-2}$ by the equivalent $\|h_\mathcal{T}^s \xi\|_{L_2(\Omega)}^{-2}$.

2.6 Numerical experiments

Let $\Gamma = \partial[0,1]^3 \subset \mathbb{R}^3$ be the two-dimensional manifold without boundary given as the boundary of the unit cube, $\mathscr{W} := H^{1/2}(\Gamma)$, $\mathscr{V} := H^{-1/2}(\Gamma)$, and $\mathscr{V}_\mathcal{T} = \mathscr{S}_\mathcal{T}^{-1,\ell} \subset \mathscr{V}$.

The role of the opposite order operator $B_\mathcal{T}^\mathscr{S} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}_{\mathcal{T},0}^{0,1}, (\mathscr{S}_{\mathcal{T},0}^{0,1})')$ from Section 2.3.3 will be fulfilled by $(B_\mathcal{T}^\mathscr{S} u)(v) := (Bu)(v)$ for an adapted hypersingular operator $B \in \mathcal{L}\mathrm{is}_c(\mathscr{W}, \mathscr{W}')$. The hypersingular operator $\tilde{B} \in \mathcal{L}(\mathscr{W}, \mathscr{W}')$ itself is only semi-coercive, but there are various options to change it into a coercive operator ([SW98]). We consider $B \in \mathcal{L}\mathrm{is}_c(\mathscr{W}, \mathscr{W}')$ given by $(Bu)(v) = (\tilde{B}u)(v) + \alpha \langle u, \mathbb{1} \rangle_{L_2(\Gamma)} \langle v, \mathbb{1} \rangle_{L_2(\Gamma)}$ for some $\alpha > 0$. By comparing different values numerically, we find $\alpha = 0.05$ to give good results in our examples.

With $m := \binom{2+\ell}{\ell} - 1$, as in §2.5.2 we equip $\mathscr{V}_\mathcal{T}$ with a usual $L_2(\Gamma)$-orthogonal basis $\{\mathbb{1}|_T \colon T \in \mathcal{T}\} \cup \{\xi_{T,i} \colon T \in \mathcal{T}, 1 \leq i \leq m\}$ where $\mathrm{supp}\, \xi_{T,i} \subset \overline{T}$, $\|\xi_{T,i}\|_{L_2(\Omega)} = |T|^{\frac{1}{2}}$, and denote the resulting stiffness matrix by $\mathbf{A}_\mathcal{T}$. The lowest order case $\ell = 0$ corresponds to $m = 0$.

Equipping $\mathscr{S}_{\mathcal{T},0}^{0,1}$ with the nodal basis $\Phi_\mathcal{T}$ defined in (2.4)-(2.5), for $\ell = 0$ the matrix representation of the preconditioner reads as

$$\boldsymbol{G}_\mathcal{T} = \boldsymbol{D}_\mathcal{T}^{-1} \big(\boldsymbol{p}_\mathcal{T}^\top \boldsymbol{B}_\mathcal{T}^\mathscr{S} \boldsymbol{p}_\mathcal{T} + \beta \boldsymbol{q}_\mathcal{T}^\top \boldsymbol{D}_\mathcal{T}^{1/2} \boldsymbol{q}_\mathcal{T} \big) \boldsymbol{D}_\mathcal{T}^{-1},$$

with $\boldsymbol{D}_\mathcal{T} = \mathrm{diag}\{|T| \colon T \in \mathcal{T}\}$ and uniformly sparse $\boldsymbol{p}_\mathcal{T}$ and $\boldsymbol{q}_\mathcal{T}$ as given in Sect. 2.3.4.

Denoting above $\boldsymbol{G}_{\mathcal{T}}$ by $\boldsymbol{G}^0_{\mathcal{T}}$, by applying for $\ell > 0$ the subspace correction method from §2.5.2, the matrix representation of the resulting uniform preconditioner is given by

$$\boldsymbol{G}_{\mathcal{T}} = \begin{bmatrix} \boldsymbol{G}^0_{\mathcal{T}} & 0 \\ 0 & \beta \operatorname{diag}\{|T|^{-3/2}\mathrm{Id}_{m\times m} \colon T \in \mathcal{T}\} \end{bmatrix}$$

The (full) matrix representations of the discretized singular integral operators $\boldsymbol{A}_{\mathcal{T}}$ and $\boldsymbol{B}^{\mathscr{S}}_{\mathcal{T}}$ are calculated using the BETL2 software package [HK12] (alternatively, one may apply low rank approximations in a hierarchical format). Condition numbers are determined using Lanczos iteration with respect to $|\!|\!|\cdot|\!|\!| := \|\boldsymbol{A}^{\frac{1}{2}}_{\mathcal{T}} \cdot \|$. The constant β is approximately optimized by comparing different choices numerically.

We will compare our preconditioner to the diagonal preconditioner $\operatorname{diag}(\boldsymbol{A}_{\mathcal{T}})^{-1}$, and in the piecewise constant case, also to the related preconditioner $\hat{\boldsymbol{G}}_{\mathcal{T}}$ from [HUT16], where $\hat{\boldsymbol{G}}_{\mathcal{T}} = \hat{\boldsymbol{D}}^{-1}_{\mathcal{T}} \boldsymbol{E}^{\top}_{\mathcal{T}} \boldsymbol{B}^{\mathscr{S}}_{\hat{\mathcal{T}}} \boldsymbol{E}_{\mathcal{T}} \hat{\boldsymbol{D}}^{-\top}_{\mathcal{T}}$ is defined as follows. With $\hat{\mathcal{T}}$ being the barycentric refinement of $\mathcal{T}$, a collection $\hat{\Psi}_{\mathcal{T}} \subset \mathscr{S}^{0,1}_{\hat{\mathcal{T}},0}$ is constructed in [BC07] such that the Fortin projector $\hat{P}_{\mathcal{T}}$ with $\operatorname{ran}\hat{P}_{\mathcal{T}} = \hat{\mathscr{W}}_{\mathcal{T}} := \operatorname{span}\hat{\Psi}_{\mathcal{T}}$ and $\operatorname{ran}(\mathrm{Id} - \hat{P}_{\mathcal{T}}) = \mathscr{V}^{\perp_{L_2(\Gamma)}}_{\mathcal{T}}$ exists, and, under an additional sufficiently mildly-grading condition on the partition, has a uniformly bounded norm $\|\hat{P}_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{W},\mathscr{W})}$ (cf. Theorem 2.3.2); $\hat{\boldsymbol{D}}_{\mathcal{T}} := \langle \Xi_{\mathcal{T}}, \hat{\Psi}_{\mathcal{T}} \rangle_{L_2(\Gamma)}$; $\boldsymbol{E}_{\mathcal{T}}$ is the representation of the embedding $\hat{\mathscr{W}}_{\mathcal{T}} \hookrightarrow \mathscr{S}^{0,1}_{\hat{\mathcal{T}},0}$ equipped with $\hat{\Psi}_{\mathcal{T}}$ and the nodal basis of $\mathscr{S}^{0,1}_{\hat{\mathcal{T}},0}$, respectively; and $B^{\mathscr{S}}_{\hat{\mathcal{T}}} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}^{0,1}_{\hat{\mathcal{T}}}, (\mathscr{S}^{0,1}_{\hat{\mathcal{T}}})')$ is an opposite order operator that we take as $(B^{\mathscr{S}}_{\hat{\mathcal{T}}} u)(v) := (Bu)(v)$, with B the adapted hypersingular operator.

Compared to our $\boldsymbol{G}_{\mathcal{T}} = \boldsymbol{G}^0_{\mathcal{T}}$, the preconditioner $\hat{\boldsymbol{G}}_{\mathcal{T}}$ has the disadvantages that, besides the aforementioned mildly grading condition, the matrix $\hat{\boldsymbol{D}}_{\mathcal{T}}$, although uniformly sparse, is not diagonal, so that the (sufficiently accurate) application of its inverse cannot be performed in linear complexity; furthermore that it requires evaluating the adapted hypersingular operator on the larger space $\mathscr{S}^{0,1}_{\hat{\mathcal{T}},0} \supset \mathscr{S}^{0,1}_{\mathcal{T},0}$ ($\#\hat{\mathcal{T}} = 6\#\mathcal{T}$); and finally that the non-standard barycentric refinement $\hat{\mathcal{T}}$ has to be generated.

2.6.1 Uniform refinements

Consider a conforming triangulation $\mathcal{T}_1$ of Γ consisting of 2 triangles per side, so 12 triangles in total. We let $\mathbb{T}$ be the sequence $\{\mathcal{T}_k\}_{k\geq 1}$ of uniform red-refinements, where $\mathcal{T}_k \succ \mathcal{T}_{k-1}$ is found by subdividing each triangle from $\mathcal{T}_{k-1}$ into 4 congruent subtriangles.

For $\mathscr{V}_{\mathcal{T}} = \mathscr{S}^{-1,\ell}_{\mathcal{T}}$, Tables 2.1 and 2.2 show the condition numbers of the preconditioned system for $\ell = 0$ and $\ell = 2$, respectively. Aside from being uni-

TABLE 2.1. Spectral condition numbers of the preconditioned single layer system, using uniform refinements, discretized by piecewise constants $\mathscr{S}_{\mathcal{T}}^{-1,0}$. Both matrices $\boldsymbol{G}_{\mathcal{T}}$ and $\hat{\boldsymbol{G}}_{\mathcal{T}}$ are constructed using the adapted hypersingular operator with $\alpha = 0.05$; and $\beta = 1.25$ in $\boldsymbol{G}_{\mathcal{T}}$.

dofs	$\kappa_S(\mathrm{diag}(\boldsymbol{A}_{\mathcal{T}})^{-1}\boldsymbol{A}_{\mathcal{T}})$	$\kappa_S(\boldsymbol{G}_{\mathcal{T}}\boldsymbol{A}_{\mathcal{T}})$	$\kappa_S(\hat{\boldsymbol{G}}_{\mathcal{T}}\boldsymbol{A}_{\mathcal{T}})$
12	14.56	2.51	1.29
48	29.30	2.52	1.58
192	58.25	2.66	1.77
768	116.3	2.71	1.89
3072	230.0	2.74	1.94
12288	444.8	2.79	

TABLE 2.2. Spectral condition numbers of the preconditioned single layer system, using uniform refinements, discretized by discontinuous piecewise quadratics $\mathscr{S}_{\mathcal{T}}^{-1,2}$. The matrix $\boldsymbol{G}_{\mathcal{T}}$ is constructed using the adapted hypersingular operator, with $\alpha = 0.05$, and $\beta = 1.25$.

dofs	$\kappa_S(\mathrm{diag}(\boldsymbol{A}_{\mathcal{T}})^{-1}\boldsymbol{A}_{\mathcal{T}})$	$\kappa_S(\boldsymbol{G}_{\mathcal{T}}\boldsymbol{A}_{\mathcal{T}})$
72	167.16	9.58
288	309.12	10.4
1152	616.03	11.1
4608	1211.3	11.3
18432	2337.2	11.4

formly bounded, the condition numbers of our preconditioner $\boldsymbol{G}_{\mathcal{T}}$ are of modest size. In the constant case, $\ell = 0$, Table 2.1 reveals that the preconditioner $\hat{\boldsymbol{G}}_{\mathcal{T}}$ from [BC07, HUT16] gives better condition numbers. As described above, this quantitative gain comes at a price. In the result of $\dim \mathscr{S}_{\mathcal{T}}^{-1,0} = 3072$, using full matrices for the discretized adapted hypersingular operator, we found a setup and application time of 1816s and 0.0971s for $\hat{\boldsymbol{G}}_{\mathcal{T}}$, compared to 385s and 0.00284s for $\boldsymbol{G}_{\mathcal{T}}$. These differences are due to numerical inversion of $\hat{\boldsymbol{D}}_{\mathcal{T}}$ by LU factorization with partial pivoting, and the enlargement $\mathscr{S}^{0,1}_{\hat{\mathcal{T}},0} \supset \mathscr{S}^{0,1}_{\mathcal{T},0}$, also causing our test machine to go out of memory in calculating $\hat{\boldsymbol{G}}_{\mathcal{T}}$ for the last refinement. Although we expect them to be in any case significant, these differences can be made smaller when the exact inversion of $\hat{\boldsymbol{D}}_{\mathcal{T}}$ is avoided, and $\boldsymbol{B}^{\mathscr{S}}_{\hat{\mathcal{T}}}$ and $\boldsymbol{B}^{\mathscr{S}}_{\mathcal{T}}$ are replaced by suitable low rank approximations.

Table 2.3. Spectral condition numbers of the preconditioned single layer system discretized by piecewise constants $\mathscr{S}_{\mathcal{T}}^{-1,0}$ using local refinements at each of the eight cube corners. Both matrices $\boldsymbol{G}_{\mathcal{T}}$ and $\hat{\boldsymbol{G}}_{\mathcal{T}}$ are constructed using the adapted hypersingular operator with $\alpha = 0.05$; and $\beta = 1.2$ in $\boldsymbol{G}_{\mathcal{T}}$. The second column is defined by $h_{\mathcal{T},min} := \min_{T\in\mathcal{T}} h_T$.

dofs	$h_{\mathcal{T},min}$	$\kappa_S(\mathrm{diag}(\boldsymbol{A}_{\mathcal{T}})^{-1}\boldsymbol{A}_{\mathcal{T}})$	$\kappa_S(\boldsymbol{G}_{\mathcal{T}}\boldsymbol{A}_{\mathcal{T}})$	$\kappa_S(\hat{\boldsymbol{G}}_{\mathcal{T}}\boldsymbol{A}_{\mathcal{T}})$
12	$7.0 \cdot 10^{-1}$	14.56	2.61	1.29
432	$2.2 \cdot 10^{-2}$	68.66	2.64	2.91
912	$6.9 \cdot 10^{-4}$	73.15	2.64	3.14
1872	$6.7 \cdot 10^{-7}$	73.70	2.64	3.25
2352	$2.1 \cdot 10^{-8}$	73.80	2.64	3.26
2976	$2.3 \cdot 10^{-10}$	73.66	2.64	

2.6.2 Local refinements

Here we take $\mathbb{T}$ to be the sequence $\{\mathcal{T}_k\}_{k\geq 1}$ of locally refined triangulations, where $\mathcal{T}_k \succ \mathcal{T}_{k-1}$ is constructed using conforming newest vertex bisection to refine all triangles in $\mathcal{T}_{k-1}$ that touch a corner of the cube.

As noted before, the preconditioner $\hat{\boldsymbol{G}}_{\mathcal{T}}$ provides uniformly bounded condition numbers if the family $\mathbb{T}$ satisfies some sufficiently mildly-grading condition on the partition [Ste03a, HUT16]. It is not directly clear whether $\mathbb{T}$ satisfies this condition, but we included the results nonetheless.

Table 2.3 gives the results for the preconditioned single layer operator discretized by piecewise constants $\mathscr{S}_{\mathcal{T}}^{-1,0}$. The condition numbers $\kappa_S(\boldsymbol{G}_{\mathcal{T}}\boldsymbol{A}_{\mathcal{T}})$ are nicely bounded under local refinements. In this case our preconditioner gives condition numbers slightly smaller than the ones found with $\hat{\boldsymbol{G}}_{\mathcal{T}}$. The calculation of the LU decomposition with partial pivoting of $\hat{\boldsymbol{D}}_{\mathcal{T}}$ turns out to break down in the last result ($\dim \mathscr{S}_{\mathcal{T}}^{-1,0} = 2976$).

2.7 Conclusion

In this chapter, we have seen how a uniformly boundedly invertible operator $B_{\mathcal{T}}^{\mathscr{S}}$ from the space of continuous piecewise linears w.r.t. any conforming shape regular partition $\mathcal{T}$, equipped with the norm of $H^s(\Omega)$ (or $H^s(\Gamma)$) for some $s \in [0,1]$, to its dual can be used to uniformly precondition a boundedly invertible operator of opposite order discretized by discontinuous or continuous polynomials of any fixed degree w.r.t. $\mathcal{T}$. The cost of the resulting preconditioner is the sum of a cost that scales linearly in $\#\mathcal{T}$ and the cost of the application of $B_{\mathcal{T}}^{\mathscr{S}}$. For $\mathcal{T}$ being member of a nested sequence of quasi-uniform partitions, $B_{\mathcal{T}}^{\mathscr{S}}$ has been constructed so that it requires linear cost. In the following chapter, we will realize this also for locally refined partitions.

3 Problems of negative order: preconditioning at linear cost

3.1 Introduction

In this chapter, we construct a multi-level type preconditioner for operators of negative orders $-2s \in [-2, 0]$ that can be applied in linear time and yields uniformly bounded condition numbers. The preconditioner will be constructing using the framework of 'operator preconditioning' discussed in Chapter 2. The role of the 'opposite order operator' will be fulfilled by a multi-level type operator, based on the work of Wu and Zheng in [WZ17].

For some d-dimensional domain (or manifold) Ω, a measurable, closed, possibly empty $\gamma \subset \partial\Omega$, and an $s \in [0, 1]$, we consider the Sobolev spaces

$$\mathscr{W} := [L_2(\Omega), H^1_{0,\gamma}(\Omega)]_{s,2}, \quad \mathscr{V} := \mathscr{W}'.$$

with $H^1_{0,\gamma}(\Omega)$ being the closure in $H^1(\Omega)$ of the smooth functions on Ω that vanish at γ. Let $(\mathscr{V}_\mathcal{T})_{\mathcal{T}\in\mathbb{T}} \subset \mathscr{V}$ be a family of *piecewise* or *continuous piecewise* polynomials of some fixed degree w.r.t. uniformly shape regular, possibly locally refined partitions. With, for $\mathcal{T} \in \mathbb{T}$, $A_\mathcal{T} \colon \mathscr{V}_\mathcal{T} \to \mathscr{V}_\mathcal{T}'$ being some boundedly invertible linear operator, we are interested in constructing a *preconditioner* $G_\mathcal{T} \colon \mathscr{V}_\mathcal{T}' \to \mathscr{V}_\mathcal{T}$ such that the preconditioned operator $G_\mathcal{T} A_\mathcal{T} \colon \mathscr{V}_\mathcal{T} \to \mathscr{V}_\mathcal{T}$ is uniformly boundedly invertible, and an application of $G_\mathcal{T}$ can be evaluated in $\mathcal{O}(\dim \mathscr{V}_\mathcal{T})$ arithmetic operations.

In order to create such a preconditioner, we will use the framework described in Chapter 2. Given $\mathscr{V}_\mathcal{T}$, we constructed an auxiliary space $\mathscr{W}_\mathcal{T} \subset \mathscr{W}$ with $\dim \mathscr{W}_\mathcal{T} = \dim \mathscr{V}_\mathcal{T}$, such that for $D_\mathcal{T}$ defined by $(D_\mathcal{T} v)(w) := \langle v, w\rangle_{L_2(\Omega)}$ ($v \in \mathscr{V}_\mathcal{T}, w \in \mathscr{W}_\mathcal{T}$) and some suitable 'opposite order' operator $B^{\mathscr{W}}_\mathcal{T} : \mathscr{W}_\mathcal{T} \to \mathscr{W}_\mathcal{T}'$, a preconditioner $G_\mathcal{T}$ of the form $G_\mathcal{T} := D_\mathcal{T}^{-1} B^{\mathscr{W}}_\mathcal{T} (D'_\mathcal{T})^{-1}$ is found. The space $\mathscr{W}_\mathcal{T}$ is equipped with a basis that, modulo a scaling, is biorthogonal to the canonical basis for $\mathscr{V}_\mathcal{T}$, so that the representation of $D_\mathcal{T}$ is an invertible diagonal matrix.

With $\mathscr{S}^{0,1}_{\mathcal{T},0} \subset \mathscr{W}$ being the space of continuous piecewise linears w.r.t. $\mathcal{T}$, zero on γ, the above preconditioning approach hinges on the availability of

a uniformly boundedly invertible operator $B_{\mathcal{T}}^{\mathscr{S}}\colon \mathscr{S}^{0,1}_{\mathcal{T},0} \to (\mathscr{S}^{0,1}_{\mathcal{T},0})'$, which is generally the most demanding requirement. For example, if $s = \frac{1}{2}$ and $\gamma = \emptyset$, a viable option is to take $B_{\mathcal{T}}^{\mathscr{S}}$ as the discretized hypersingular operator. While this induces a uniform preconditioner, the application of $B_{\mathcal{T}}^{\mathscr{S}}$ cannot be evaluated in linear complexity.

In this chapter we construct a suitable multi-level type operator $B_{\mathcal{T}}^{\mathscr{S}}$ that *can* be applied in linear complexity. For this construction we require $\mathbb{T}$ to be a family of conforming partitions created by Newest Vertex Bisection ([Mau95, Tra97]). In the aforementioned setting of having an arbitrary $s \in [0,1]$, this multi-level operator $B_{\mathcal{T}}^{\mathscr{S}}$ induces a uniform preconditioner $G_{\mathcal{T}}$, i.e., $G_{\mathcal{T}} A_{\mathcal{T}}$ is uniformly well-conditioned, where the cost of applying $G_{\mathcal{T}}$ scales linearly in $\dim \mathscr{V}_{\mathcal{T}}$. We also show that the preconditioner extends to the more general *manifold* case, where Ω is a d-dimensional (piecewise) smooth Lipschitz manifold, and the trial space $\mathscr{V}_{\mathcal{T}}$ is the parametric lift of a space of piecewise or continuous piecewise polynomials.

Finally, we remark that common multi-level preconditioners based on overlapping subspace decompositions are known not to work well for operators of negative order. A solution is provided by resorting to direct sum multi-level subspace decompositions. Examples are given by wavelet preconditioners, or closely related, the preconditioners from [BPV00], for the latter assuming quasi-uniform partitions.

For $-s = -\frac{1}{2}$, an optimal multi-level preconditioner based on a *non*-overlapping subspace decomposition for operators defined on the boundary of a 2- or 3-dimensional Lipschitz polyhedron was recently introduced in [FHPS19].

3.1.1 Outline

In Sect. 3.2 we summarize the (operator) preconditioning framework from Chapter 2. In Sect. 3.3 we provide the multi-level type operator that can be used as the 'opposite order' operator inside the preconditioner framework. In Sect. 3.4 we comment on how to generalize the results to the case of piecewise smooth manifolds. In Sect. 3.5 we conclude with numerical results.

3.1.2 Notation

In this work, by $\lambda \lesssim \mu$ we mean that λ can be bounded by a multiple of μ, independently of parameters which λ and μ may depend on, with the sole exception of the space dimension d, or in the manifold case, on the parametrization of the manifold that is used to define the finite element spaces on it. Obviously, $\lambda \gtrsim \mu$ is defined as $\mu \lesssim \lambda$, and $\lambda \eqsim \mu$ as $\lambda \lesssim \mu$ and $\lambda \gtrsim \mu$.

For normed linear spaces $\mathscr{Y}$ and $\mathscr{Z}$, in this work for convenience over $\mathbb{R}$, $\mathcal{L}(\mathscr{Y}, \mathscr{Z})$ will denote the space of bounded linear mappings $\mathscr{Y} \to \mathscr{Z}$ endowed with the operator norm $\|\cdot\|_{\mathcal{L}(\mathscr{Y},\mathscr{Z})}$. The subset of invertible operators in

$\mathcal{L}(\mathscr{Y},\mathscr{Z})$ with inverses in $\mathcal{L}(\mathscr{Z},\mathscr{Y})$ will be denoted as $\mathcal{L}\mathrm{is}(\mathscr{Y},\mathscr{Z})$. The *condition number* of a $C\in\mathcal{L}\mathrm{is}(\mathscr{Y},\mathscr{Z})$ is defined as $\kappa_{\mathscr{Y},\mathscr{Z}}(C):=\|C\|_{\mathcal{L}(\mathscr{Y},\mathscr{Z})}\|C^{-1}\|_{\mathcal{L}(\mathscr{Z},\mathscr{Y})}$.

For $\mathscr{Y}$ a reflexive Banach space and $C\in\mathcal{L}(\mathscr{Y},\mathscr{Y}')$ being *coercive*, i.e.,

$$\inf_{0\neq y\in\mathscr{Y}}\frac{(Cy)(y)}{\|y\|^2_{\mathscr{Y}}}>0,$$

both C and $\Re(C):=\frac{1}{2}(C+C')$ are in $\mathcal{L}\mathrm{is}(\mathscr{Y},\mathscr{Y}')$ with

$$\begin{aligned}\|\Re(C)\|_{\mathcal{L}(\mathscr{Y},\mathscr{Y}')}&\leq\|C\|_{\mathcal{L}(\mathscr{Y},\mathscr{Y}')},\\ \|C^{-1}\|_{\mathcal{L}(\mathscr{Y}',\mathscr{Y})}&\leq\|\Re(C)^{-1}\|_{\mathcal{L}(\mathscr{Y}',\mathscr{Y})}=\Big(\inf_{0\neq y\in\mathscr{Y}}\frac{(Cy)(y)}{\|y\|^2_{\mathscr{Y}}}\Big)^{-1}.\end{aligned}$$

The subset of coercive operators in $\mathcal{L}\mathrm{is}(\mathscr{Y},\mathscr{Y}')$ is denoted as $\mathcal{L}\mathrm{is}_c(\mathscr{Y},\mathscr{Y}')$. If $C\in\mathcal{L}\mathrm{is}_c(\mathscr{Y},\mathscr{Y}')$, then $C^{-1}\in\mathcal{L}\mathrm{is}_c(\mathscr{Y}',\mathscr{Y})$ and $\|\Re(C^{-1})^{-1}\|_{\mathcal{L}(\mathscr{Y},\mathscr{Y}')}\leq\|C\|^2_{\mathcal{L}(\mathscr{Y},\mathscr{Y}')}\|\Re(C)^{-1}\|_{\mathcal{L}(\mathscr{Y}',\mathscr{Y})}$.

Given a family of operators $C_i\in\mathcal{L}\mathrm{is}(\mathscr{Y}_i,\mathscr{Z}_i)$ ($\mathcal{L}\mathrm{is}_c(\mathscr{Y}_i,\mathscr{Z}_i)$), we will write $C_i\in\mathcal{L}\mathrm{is}(\mathscr{Y}_i,\mathscr{Z}_i)$ ($\mathcal{L}\mathrm{is}_c(\mathscr{Y}_i,\mathscr{Z}_i)$) uniformly in i, or simply 'uniform', when

$$\sup_i\max(\|C_i\|_{\mathcal{L}(\mathscr{Y}_i,\mathscr{Z}_i)},\|C_i^{-1}\|_{\mathcal{L}(\mathscr{Z}_i,\mathscr{Y}_i)})<\infty,$$

or

$$\sup_i\max(\|C_i\|_{\mathcal{L}(\mathscr{Y}_i,\mathscr{Z}_i)},\|\Re(C_i)^{-1}\|_{\mathcal{L}(\mathscr{Z}_i,\mathscr{Y}_i)})<\infty.$$

3.2 Operator preconditioning

Let $(\mathcal{T})_{\mathcal{T}\in\mathbb{T}}$ be a family of *conforming* partitions of a domain $\Omega\subset\mathbb{R}^d$ into (open) *uniformly shape regular* d-simplices, where we assume that γ is the (possibly empty) union of $(d-1)$-faces of $T\in\mathcal{T}$. For $d\geq 2$, such partitions automatically satisfy a uniform K-mesh property, and for $d=1$ we impose this as an additional condition. The discussion of the manifold case is postponed to Sect. 3.4.

Recalling that $\mathscr{V}_\mathcal{T}\subset\mathscr{V}$ is a family of piecewise or continuous piecewise polynomials of some fixed degree w.r.t. $\mathcal{T}$, let $A_\mathcal{T}\in\mathcal{L}\mathrm{is}(\mathscr{V}_\mathcal{T},\mathscr{V}_\mathcal{T}')$ uniformly in $\mathcal{T}\in\mathbb{T}$. A common setting is that $(A_\mathcal{T}v)(\tilde v):=(Av)(\tilde v)$ ($v,\tilde v\in\mathscr{V}_\mathcal{T}$) for some $A\in\mathcal{L}\mathrm{is}_c(\mathscr{V},\mathscr{V}')$. We are interested in finding optimal *preconditioners* $G_\mathcal{T}$ for $A_\mathcal{T}$, i.e., $G_\mathcal{T}\in\mathcal{L}\mathrm{is}(\mathscr{V}_\mathcal{T}',\mathscr{V}_\mathcal{T})$ uniformly in $\mathcal{T}\in\mathbb{T}$, whose application moreover requires $\mathcal{O}(\dim\mathscr{V}_\mathcal{T})$ arithmetic operations.

Recall the space

$$\mathscr{S}^{0,1}_{\mathcal{T},0}:=\{u\in H^1_{0,\gamma}(\Omega)\colon u|_T\in\mathcal{P}_1\,(T\in\mathcal{T})\}\subset\mathscr{W}$$

(thus equipped with $\|\cdot\|_\mathscr{W}$). In Chapter 2, using operator preconditioning, we reduced the issue of constructing such preconditioners $G_\mathcal{T}$ to the issue of constructing $B^\mathscr{S}_\mathcal{T}\in\mathcal{L}\mathrm{is}_c(\mathscr{S}^{0,1}_{\mathcal{T},0},(\mathscr{S}^{0,1}_{\mathcal{T},0})')$ uniformly. In the next section we summarize this reduction.

3.2.1 Construction of optimal preconditioners

For the moment, consider the lowest order case of $\mathscr{V}_\mathcal{T}$ being either the space of piecewise constants or continuous piecewise linears. In Chapter 2 a space $\mathscr{W}_\mathcal{T} \subset \mathscr{W}$ was constructed with $\dim \mathscr{W}_\mathcal{T} = \dim \mathscr{V}_\mathcal{T}$ and

$$\inf_{\mathcal{T}\in\mathbb{T}} \inf_{0\neq v\in\mathscr{V}_\mathcal{T}} \sup_{0\neq w\in\mathscr{W}_\mathcal{T}} \frac{\langle v, w\rangle_{L_2(\Omega)}}{\|v\|_\mathscr{V}\|w\|_\mathscr{W}} > 0. \tag{3.1}$$

Moreover, $\mathscr{W}_\mathcal{T} \subset \mathscr{W}$ was equipped with a locally supported basis $\Psi_\mathcal{T}$ that, modulo a scaling, is $L_2(\Omega)$-*biorthogonal* to the canonical basis $\Xi_\mathcal{T}$ of $\mathscr{V}_\mathcal{T}$.

As a consequence of (3.1), $D_\mathcal{T}$ defined by $(D_\mathcal{T} v)(w) := \langle v, w\rangle_{L_2(\Omega)}$ ($v \in \mathscr{V}_\mathcal{T}$, $w \in \mathscr{W}_\mathcal{T}$) is in $\mathcal{L}\mathrm{is}(\mathscr{V}_\mathcal{T}, \mathscr{W}_\mathcal{T}')$ uniformly. We infer that once we have constructed $B_\mathcal{T}^\mathscr{W} \in \mathcal{L}\mathrm{is}(\mathscr{W}_\mathcal{T}, \mathscr{W}_\mathcal{T}')$ uniformly, then by taking

$$G_\mathcal{T} := D_\mathcal{T}^{-1} B_\mathcal{T}^\mathscr{W} (D_\mathcal{T}')^{-1}, \tag{3.2}$$

we have $G_\mathcal{T} \in \mathcal{L}\mathrm{is}(\mathscr{V}_\mathcal{T}', \mathscr{V}_\mathcal{T})$ uniformly. Biorthogonality, modulo a scaling, of the bases $\Psi_\mathcal{T}$ and $\Xi_\mathcal{T}$ implies that the matrix representation of $D_\mathcal{T}$ is diagonal, so that $D_\mathcal{T}^{-1}$ and its adjoint can be applied in linear complexity.

The aforementioned space $\mathscr{W}_\mathcal{T}$ is a subspace of $\mathscr{S}^{0,1}_{\mathcal{T},0} \oplus \mathscr{B}_\mathcal{T} \subset \mathscr{W}$, where $\mathscr{B}_\mathcal{T}$ is a 'bubble space' with $\dim \mathscr{B}_\mathcal{T} = \mathcal{O}(\#\mathcal{T})$, such that the projector $I_\mathcal{T}$ on $\mathscr{S}^{0,1}_{\mathcal{T},0} \oplus \mathscr{B}_\mathcal{T}$, defined by $\operatorname{ran} I_\mathcal{T} = \mathscr{S}^{0,1}_{\mathcal{T},0}$ and $\operatorname{ran}(\mathrm{Id} - I_\mathcal{T}) = \mathscr{B}_\mathcal{T}$, is 'local' and uniformly bounded, and the canonical basis $\Theta_\mathcal{T}$ of 'bubbles' for $\mathscr{B}_\mathcal{T}$ is, when normalized in $\|\cdot\|_\mathscr{W}$, a uniformly Riesz basis for $\mathscr{B}_\mathcal{T}$. Because of the latter, $B_\mathcal{T}^\mathscr{B}$ defined by

$$(B_\mathcal{T}^\mathscr{B} \mathbf{c}^\top \Theta_\mathcal{T})(\mathbf{d}^\top \Theta_\mathcal{T}) := \beta(\boldsymbol{\Delta}_\mathcal{T}\mathbf{c})^\top \mathbf{d}$$

for some diagonal $\boldsymbol{\Delta}_\mathcal{T} \eqsim \operatorname{diag}(\langle \Theta_\mathcal{T}, \Theta_\mathcal{T}\rangle_\mathscr{W})$ and constant $\beta > 0$ is in $\mathcal{L}\mathrm{is}_c(\mathscr{B}_\mathcal{T}, \mathscr{B}_\mathcal{T}')$ uniformly.

Given some 'opposite order' operator $B_\mathcal{T}^\mathscr{S} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}^{0,1}_{\mathcal{T},0}, (\mathscr{S}^{0,1}_{\mathcal{T},0})')$, by taking

$$B_\mathcal{T}^\mathscr{W} := I_\mathcal{T}' B_\mathcal{T}^\mathscr{S} I_\mathcal{T} + (\mathrm{Id} - I_\mathcal{T})' B_\mathcal{T}^\mathscr{B} (\mathrm{Id} - I_\mathcal{T}), \tag{3.3}$$

it holds that $B_\mathcal{T}^\mathscr{W} \in \mathcal{L}\mathrm{is}_c(\mathscr{W}_\mathcal{T}, \mathscr{W}_\mathcal{T}')$ uniformly ([SvV20b, Prop. 5.1]), which makes $G_\mathcal{T}$ a uniform preconditioner.

3.2.2 Implementation of $G_\mathcal{T}$

Recalling the aforementioned bases $\Xi_\mathcal{T}$, $\Psi_\mathcal{T}$, and $\Theta_\mathcal{T}$ for $\mathscr{V}_\mathcal{T}$, $\mathscr{W}_\mathcal{T}$ and $\mathscr{B}_\mathcal{T}$, respectively, equipping $\mathscr{S}^{0,1}_{\mathcal{T},0}$ with the nodal basis $\Phi_\mathcal{T}$, and equipping $\mathscr{V}_\mathcal{T}'$, $\mathscr{W}_\mathcal{T}'$, $\mathscr{B}_\mathcal{T}'$, and $(\mathscr{S}^{0,1}_{\mathcal{T},0})'$ with the dual bases $\Xi_\mathcal{T}'$, $\Psi_\mathcal{T}'$, $\Theta_\mathcal{T}'$, and $\Phi_\mathcal{T}'$, respectively, the representation of $A_\mathcal{T} \in \mathcal{L}(\mathscr{V}_\mathcal{T}, \mathscr{V}_\mathcal{T}')$ is the stiffness matrix $\boldsymbol{A}_\mathcal{T} := (A_\mathcal{T}\Xi_\mathcal{T})(\Xi_\mathcal{T}) := [(A_\mathcal{T}\eta)(\xi)]_{(\xi,\eta)\in\Xi_\mathcal{T}}$, and the representation of $G_\mathcal{T} \in \mathcal{L}(\mathscr{V}_\mathcal{T}', \mathscr{V}_\mathcal{T})$ is the matrix $\boldsymbol{G}_\mathcal{T} := (G\Xi_\mathcal{T}')(\Xi_\mathcal{T}')$. It is given by

$$\boldsymbol{G}_\mathcal{T} = \boldsymbol{D}_\mathcal{T}^{-1}\big(\boldsymbol{p}_\mathcal{T}^\top \boldsymbol{B}_\mathcal{T}^\mathscr{S} \boldsymbol{p}_\mathcal{T} + \boldsymbol{q}_\mathcal{T}^\top \boldsymbol{B}_\mathcal{T}^\mathscr{B} \boldsymbol{q}_\mathcal{T}\big) \boldsymbol{D}_\mathcal{T}^{-\top}, \tag{3.4}$$

where both

$$\boldsymbol{D}_{\mathcal{T}} := (D_{\mathcal{T}}\Xi_{\mathcal{T}})(\Psi_{\mathcal{T}}), \quad \boldsymbol{B}^{\mathscr{B}}_{\mathcal{T}} := (B^{\mathscr{B}}_{\mathcal{T}}\Theta_{\mathcal{T}})(\Theta_{\mathcal{T}})$$

are *diagonal*, both

$$\boldsymbol{p}_{\mathcal{T}} := (I_{\mathcal{T}}\Psi_{\mathcal{T}})(\Phi'_{\mathcal{T}}), \qquad \boldsymbol{q}_{\mathcal{T}} := ((\mathrm{Id} - I_{\mathcal{T}})\Psi_{\mathcal{T}})(\Theta'_{\mathcal{T}})$$

are *uniformly sparse*, and

$$\boldsymbol{B}^{\mathscr{S}}_{\mathcal{T}} := (B^{\mathscr{S}}_{\mathcal{T}}\Phi_{\mathcal{T}})(\Phi_{\mathcal{T}}). \tag{3.5}$$

Note that the cost of the application of $\boldsymbol{G}_{\mathcal{T}}$ scales *linearly* in $\#\mathcal{T}$ as soon as this holds true for the application of $\boldsymbol{B}^{\mathscr{S}}_{\mathcal{T}}$.

The above preconditioning approach is summarized in the following theorem.

Theorem 3.2.1 (Sect. 2.3)**.** *Given a family $B^{\mathscr{S}}_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}^{0,1}_{\mathcal{T},0}, (\mathscr{S}^{0,1}_{\mathcal{T},0})')$ uniformly in $\mathcal{T} \in \mathbb{T}$. Then for $B^{\mathscr{W}}_{\mathcal{T}}$ as described in* (3.3)*, the operator $G_{\mathcal{T}}$ from* (3.2) *is a uniform preconditioner. Furthermore, if the matrix representation $\boldsymbol{B}^{\mathscr{S}}_{\mathcal{T}}$, cf.* (3.5)*, can be applied in $\mathcal{O}(\#\mathcal{T})$ operations, then the matrix representation of the preconditioner $\boldsymbol{G}_{\mathcal{T}}$, cf.* (3.4)*, can be applied in $\mathcal{O}(\#\mathcal{T})$ operations.*

Because $B^{\mathscr{W}}_{\mathcal{T}}$ in (3.3) is given as the sum of two operators that 'act' on different subspaces of $\mathscr{W}_{\mathcal{T}}$, the condition number of the preconditioned system depends on the relative scaling of both these operators which can be steered by selecting the parameter β. A suitable β will be selected experimentally.

Alternatively, Proposition 4.4.1 shows that a value of β is reasonable if it is chosen such that the interval bounded by the coercivity and boundedness constants of $B^{\mathscr{S}}_{\mathcal{T}}$ is included in that interval corresponding to $B^{\mathscr{B}}_{\mathcal{T}}$ or vice versa. Also these coercivity and boundedness constants can be approximated experimentally or by making some theoretical estimates.

Constructions of $\Psi_{\mathcal{T}}$, $\Theta_{\mathcal{T}}$, and $\boldsymbol{\Delta}_{\mathcal{T}}$, and resulting explicit formulas for matrices $\boldsymbol{D}_{\mathcal{T}}$, $\boldsymbol{B}^{\mathscr{B}}_{\mathcal{T}}$, $\boldsymbol{p}_{\mathcal{T}}$, $\boldsymbol{q}_{\mathcal{T}}$ are derived in Chapter 2. For ease of reading we recall these formulas below for the case that $\mathscr{V}_{\mathcal{T}}$ is the space of *piecewise constants*. For the continuous piecewise linear case we refer to Sect. 2.4.2.

Piecewise constant trial space $\mathscr{V}_{\mathcal{T}}$

For $\mathcal{T} \in \mathbb{T}$, we define $N_{\mathcal{T}}$ as the set of vertices of $\mathcal{T}$, and $N^0_{\mathcal{T}}$ as the set of vertices of $\mathcal{T}$ that are not on γ. For $\nu \in N_{\mathcal{T}}$ we set its *valence*

$$d_{\mathcal{T},\nu} := \#\{T \in \mathcal{T} \colon \nu \in \overline{T}\}.$$

For $T \in \mathcal{T}$, and with N_T denoting the set of its vertices, we set $N^0_{\mathcal{T},T} := N^0_{\mathcal{T}} \cap N_T$.

If one considers $\mathscr{V}_\mathcal{T}$ as the space of discontinuous piecewise constants, i.e.

$$\mathscr{V}_\mathcal{T} = \mathscr{S}_\mathcal{T}^{-1,0} := \{u \in L_2(\Omega)\colon u|_T \in P_0(T \in \mathcal{T})\} \subset \mathscr{V},$$

equipped with the canonical basis $\Xi_\mathcal{T} := \{\mathbb{1}_T\colon T \in \mathcal{T}\}$, then we find, for arbitrary constant $\beta > 0$,

$$\boldsymbol{D}_\mathcal{T} = \operatorname{diag}\{|T|\colon T \in \mathcal{T}\}, \qquad (\boldsymbol{p}_\mathcal{T})_{\nu T} = \begin{cases} d_{\mathcal{T},\nu}^{-1} & \text{if } \nu \in N_{\mathcal{T},T}^0, \\ 0 & \text{if } \nu \notin N_{\mathcal{T},T}^0, \end{cases}$$

$$\boldsymbol{B}_\mathcal{T}^{\mathscr{B}} = \beta \boldsymbol{D}_\mathcal{T}^{1-\frac{2s}{d}}, \qquad (\boldsymbol{q}_\mathcal{T})_{T'T} = \delta_{T'T} - \tfrac{1}{d+1} \sum_{\nu \in N_{\mathcal{T},T}^0 \cap N_{\mathcal{T},T'}^0} d_{\mathcal{T},\nu}^{-1}.$$

3.2.3 Higher order case

For higher order discontinuous or continuous finite element spaces $\mathscr{V}_\mathcal{T}$, suitable preconditioners $G_\mathcal{T}$ can be built either from the current preconditioner $G_\mathcal{T}$ for the lowest order case by application of a subspace correction method (most conveniently in the discontinuous case where on each element the space of polynomials of some fixed degree is split into the space of constants and its orthogonal complement), or by expanding $\mathscr{W}_\mathcal{T}$ by enlarging the bubble space $\mathscr{B}_\mathcal{T}$. While referring to Chapter 2 for details, we recall that with either option the construction of an optimal preconditioner $G_\mathcal{T}$ that can be applied in linear complexity hinges on the availability of an operator $B_\mathcal{T}^\mathscr{S} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}_{\mathcal{T},0}^{0,1}, (\mathscr{S}_{\mathcal{T},0}^{0,1})')$ uniformly in $\mathcal{T} \in \mathbb{T}$, that can be applied in linear complexity.

3.3 An operator $B_\mathcal{T}^\mathscr{S}$ of multi-level type

In this section we will introduce an operator $B_\mathcal{T}^\mathscr{S} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}_{\mathcal{T},0}^{0,1}, (\mathscr{S}_{\mathcal{T},0}^{0,1})')$ of multi-level type. The operator $B_\mathcal{T}^\mathscr{S}$ is based on a stable multi-level decomposition of $\mathscr{S}_{\mathcal{T},0}^{0,1}$ given by Wu and Zheng [WZ17]. Usually such a stable multi-level decomposition is used as a theoretical tool for proving optimality of an additive (or multiplicative) Schwarz type preconditioner for an operator in $\mathcal{L}\mathrm{is}_c(\mathscr{S}_{\mathcal{T},0}^{0,1}, (\mathscr{S}_{\mathcal{T},0}^{0,1})')$. In this work, however, we are going to use their results for the construction of the operator $B_\mathcal{T}^\mathscr{S} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}_{\mathcal{T},0}^{0,1}, (\mathscr{S}_{\mathcal{T},0}^{0,1})')$ for which it is crucial that its application can be implemented in linear complexity.

3.3.1 Definition and analysis of $B_\mathcal{T}^\mathscr{S}$

For $d \geq 2$, let $\mathbb{T}$ be the family of all conforming partitions of Ω into d-simplices that can be created by Newest Vertex Bisection starting from some given conforming initial partition $\mathcal{T}_\perp$ that satisfies a *matching condition* ([Ste08b]).

With $\mathfrak{T} := \cup_{\mathcal{T}\in\mathbb{T}}\{T\colon T \in \mathcal{T}\}$ and $\mathfrak{N} := \cup_{\mathcal{T}\in\mathbb{T}} N_\mathcal{T}$, for $T \in \mathfrak{T}$ let $\operatorname{gen}(T)$ be the number of bisections needed to create T from its ancestor $T' \in \mathcal{T}_\perp$, and for

$\nu \in \mathfrak{N}$ let $\operatorname{gen}(\nu) := \min\{\operatorname{gen}(T)\colon T \in \mathfrak{T},\, \nu \in N_T\}$. Notice that $|T| \eqsim 2^{-\operatorname{gen}(T)}$. For $T \in \mathfrak{T}$, let Q_T denote the $L_2(T)$-orthogonal projector onto $\mathcal{P}_1(T)$.

The case $d = 1$ can be included by letting $\mathbb{T}$ be the family of a partitions of Ω that can be constructed by bisections from $\mathcal{T}_\perp = \{\Omega\}$ such that the generations of any two neighbouring subintervals in any $\mathcal{T} \in \mathbb{T}$ differ by not more than one.

For $\mathcal{T} \in \mathbb{T}$, set $L = L(\mathcal{T}) := \max_{T\in\mathcal{T}} \operatorname{gen}(T)$ and define

$$\mathcal{T}_\perp = \mathcal{T}_0 \prec \mathcal{T}_1 \prec \cdots \prec \mathcal{T}_L = \mathcal{T} \subset \mathbb{T}$$

by constructing $\mathcal{T}_{j-1}$ from $\mathcal{T}_j$ by removing all $\nu \in N_j := N_{\mathcal{T}_j}$ from the latter for which $\operatorname{gen}(\nu) = j$. For $\nu \in N_j^0 := N^0_{\mathcal{T}_j}$, we define $\omega_j(\nu) = \cup\{T \in \mathcal{T}_j\colon \nu \in N_T\}$.

With this hierarchy of partitions, we define an averaging quasi-interpolator $\Pi_j \in \mathcal{L}(\mathscr{S}^{0,1}_{\mathcal{T},0}, \mathscr{S}^{0,1}_{\mathcal{T}_j,0})$ by

$$(\Pi_j u)(\nu) := \frac{\sum_{\{T\in\mathcal{T}_j\colon \nu\in N_T\}} |T|(Q_T u)(\nu)}{\sum_{\{T\in\mathcal{T}_j\colon \nu\in N_T\}} |T|} \quad (u \in \mathscr{S}^{0,1}_{\mathcal{T},0},\, \nu \in N_j^0). \tag{3.6}$$

Since $\mathscr{S}^{0,1}_{\mathcal{T}_j,0}$ is a space of continuous piecewise linears, it indeed suffices to define $\Pi_j u$ at the vertices N_j^0. Recall that $\mathscr{S}^{0,1}_{\mathcal{T},0} \subset \mathscr{W} := [L_2(\Omega), H^1_{0,\gamma}(\Omega)]_{s,2}$ for some $s \in [0,1]$. The next theorem shows that Π_j induces a stable multi-level decomposition of $\mathscr{S}^{0,1}_{\mathcal{T},0}$.

Theorem 3.3.1 ([WZ17, Lemma 3.7])**.** *For the averaging quasi-interpolator* Π_j *from* (3.6)*, and* $\Pi_{-1} := 0$*, it holds that*

$$\|u\|^2_{\mathscr{W}} \eqsim \sum_{j=0}^{L} 4^{js/d} \|(\Pi_j - \Pi_{j-1})u\|^2_{L_2(\Omega)} \quad (u \in \mathscr{S}^{0,1}_{\mathcal{T},0}).$$

Proof. In [WZ17], the inequality '$\gtrsim$' was proven for the case $s = 1$, $d \in \{2,3\}$, and $\gamma = \partial\Omega$. The arguments, however, immediately extend to $s \in [0,1]$, $d \geq 1$, and $\gamma \subsetneq \partial\Omega$.

The proof of the other inequality '$\lesssim$' follows from well-known arguments: For some $t \in (1, \frac{3}{2})$, let $\mathscr{H}^{rt} := [L_2(\Omega), H^1_{0,\gamma}(\Omega) \cap H^t(\Omega)]_{r,2}$ for $r \in [0,1]$. Then $\mathscr{H}^s \simeq \mathscr{W}$ by the *reiteration theorem*, and for $r \in [0,1]$, $\|\cdot\|_{\mathscr{H}^{rt}} \lesssim 2^{jrt/d}\|\cdot\|_{L_2(\Omega)}$ on $\mathscr{S}^{0,1}_{\mathcal{T}_j,0}$.

Let $u \in \mathscr{S}^{0,1}_{\mathcal{T},0}$ be written as $\sum_{j=0}^L u_j$ with $u_j \in \mathscr{S}^{0,1}_{\mathcal{T}_j,0}$. Then for $\varepsilon \in (0,s)$, $\varepsilon \leq t - s$, we have

$$\begin{aligned}\|u\|^2_{\mathscr{H}^s(\Omega)} &\lesssim \sum_{j=0}^{L}\sum_{i=j}^{L} \|u_j\|_{\mathscr{H}^{s+\varepsilon}(\Omega)}\|u_i\|_{\mathscr{H}^{s-\varepsilon}(\Omega)} \\ &\lesssim \sum_{j=0}^{L}\sum_{i=j}^{L} 2^{j(s+\varepsilon)/d}2^{i(s-\varepsilon)/d}\|u_j\|_{L_2(\Omega)}\|u_i\|_{L_2(\Omega)} \lesssim \sum_{j=0}^{L} 4^{js/d}\|u_j\|^2_{L_2(\Omega)}. \qquad \square\end{aligned}$$

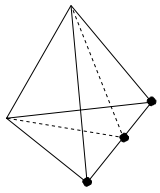


Figure 3.1. For $d = 3$, a tetrahedron $T \in \mathcal{T}_{j-1}$ and its bisection. The dots indicate all vertices in $N_j^0 \setminus M_j^0$.

The relevance of the multi-level decomposition from Theorem 3.3.1 by Wu and Zheng lies in the fact that $(\Pi_j u)(\nu)$ can only differ from $(\Pi_{j-1} u)(\nu)$ in any $\nu \in N_j^0 \setminus N_{j-1}^0$ as well as in only *two*[1] of its neighbours in N_{j-1}^0 (the endpoints of the edge on which ν was inserted):

Proposition 3.3.2 ([WZ17, Lemma 3.1])**.** *With for $j \geq 1$,*

$$M_j^0 := \{\nu \in N_{j-1}^0 \colon \omega_j(\nu) = \omega_{j-1}(\nu)\},$$

it holds that for $\nu \in M_j^0$, $((\Pi_j - \Pi_{j-1})u)(\nu) = 0$, see Figure 3.1.

Remark 3.3.3. The proof from [WZ17] given for $d \in \{2, 3\}$ generalizes to $d \geq 1$. Indeed the arguments that are used are based on the fact that the basis for $S_1(T)$ that is dual to the nodal basis takes equal values in all but one nodal point. This is a consequence of the fact that the mass matrix of the nodal basis for $S_1(T)$, and so its inverse, is invariant under permutations of the barycentric coordinates, which holds true in any dimension.

As a consequence of Proposition 3.3.2, we have

$$\|(\Pi_j - \Pi_{j-1})u\|_{L_2(\Omega)}^2 \eqsim 2^{-j} \sum_{\nu \in N_j^0 \setminus M_j^0} |((\Pi_j - \Pi_{j-1})u)(\nu)|^2.$$

From Theorem 3.3.1, we conclude that $B_{\mathcal{T}}^{\mathscr{S}} = (B_{\mathcal{T}}^{\mathscr{S}})' \in \mathcal{L}\mathrm{is}_c(\mathscr{S}_{\mathcal{T},0}^{0,1}, (\mathscr{S}_{\mathcal{T},0}^{0,1})')$ defined by

$$(B_{\mathcal{T}}^{\mathscr{S}} u)(v) := \sum_{j=0}^{L} 2^{j(\frac{2s}{d}-1)} \sum_{\nu \in N_j^0 \setminus M_j^0} ((\Pi_j - \Pi_{j-1})u)(\nu)((\Pi_j - \Pi_{j-1})v)(\nu) \tag{3.7}$$

is uniform, i.e.

$$\sup_{\mathcal{T} \in \mathbb{T}} \max\left(\|B_{\mathcal{T}}^{\mathscr{S}}\|_{\mathcal{L}(\mathscr{S}_{\mathcal{T},0}^{0,1}, (\mathscr{S}_{\mathcal{T},0}^{0,1})')}, \|(B_{\mathcal{T}}^{\mathscr{S}})^{-1}\|_{\mathcal{L}((\mathscr{S}_{\mathcal{T},0}^{0,1})', \mathscr{S}_{\mathcal{T},0}^{0,1})} \right) < \infty.$$

[1] As pointed out in [WZ17], for $d \geq 3$ the number of such neighbours will be larger when employing the Scott-Zhang quasi-interpolator. Moreover, this interpolator is not suited for $s \leq \frac{1}{2}$.

3.3.2 Implementation of $B^{\mathscr{S}}_{\mathcal{T}}$

Since the operator Π_j is a weighted local $L_2(\Omega)$ projection, it allows for a natural implementation by considering $\mathscr{S}^{-1,1}_{\mathcal{T}}$, the space of discontinuous piecewise linears w.r.t. $\mathcal{T}$. Recall the nodal basis $\Phi_{\mathcal{T}}$ for $\mathscr{S}^{0,1}_{\mathcal{T},0}$, and equip $\mathscr{S}^{-1,1}_{\mathcal{T}}$ with the element-wise nodal basis.

Denote $\boldsymbol{E}_{\mathcal{T}}$ for the representation of the embedding $\mathscr{S}^{0,1}_{\mathcal{T},0}$ into $\mathscr{S}^{-1,1}_{\mathcal{T}}$.

For $0 \leq j \leq L$, let $\boldsymbol{R}_j$ be the representation of the $L_2(\Omega)$-orthogonal projector of $\mathscr{S}^{-1,1}_{\mathcal{T}}$ onto $\mathscr{S}^{-1,1}_{\mathcal{T}_j}$, and let $\boldsymbol{R}_{-1} := 0$.

For $0 \leq j \leq L$, let $\boldsymbol{H}_j$ be the representation of the averaging operator $H_j \colon \mathscr{S}^{-1,1}_{\mathcal{T}_j} \to \mathscr{S}^{0,1}_{\mathcal{T}_j,0}$ defined by

$$(H_j u)(\nu) = \frac{\sum_{\{T\in\mathcal{T}_j \,:\, \nu\in N_T\}} |T|\, u|_T(\nu)}{\sum_{\{T\in\mathcal{T}_j \,:\, \nu\in N_T\}} |T|}, \quad (\nu \in N^0_{\mathcal{T}}), \tag{3.8}$$

and let $\boldsymbol{H}_{-1} := 0$.

For $1 \leq j \leq L$, let $\boldsymbol{P}_j$ be the representation of the embedding $\mathscr{S}^{0,1}_{\mathcal{T}_{j-1},0} \to \mathscr{S}^{0,1}_{\mathcal{T}_j,0}$ (often called *prolongation*), and let $\boldsymbol{P}_0 := 0$.

Then the representation $\boldsymbol{B}^{\mathscr{S}}_{\mathcal{T}}$ of $B^{\mathscr{S}}_{\mathcal{T}}$ from (3.7) is given by

$$\boldsymbol{B}^{\mathscr{S}}_{\mathcal{T}} = \boldsymbol{E}_{\mathcal{T}}^{\top}\Big(\sum_{j=0}^{L}(\boldsymbol{H}_j\boldsymbol{R}_j - \boldsymbol{P}_j\boldsymbol{H}_{j-1}\boldsymbol{R}_{j-1})^{\top} 2^{j(\frac{2s}{d}-1)}(\boldsymbol{H}_j\boldsymbol{R}_j - \boldsymbol{P}_j\boldsymbol{H}_{j-1}\boldsymbol{R}_{j-1})\Big)\boldsymbol{E}_{\mathcal{T}}.$$

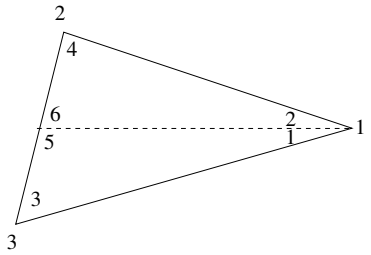
Applying $\boldsymbol{E}_{\mathcal{T}}$ amounts to duplicating values at any internal node with a number equal to the valence of that node.

By representing $\mathcal{T}$ as the leaves of a binary tree with roots being the simplices of $\mathcal{T}_\perp$, computing for $\mathbf{x} \in \operatorname{ran} \boldsymbol{E}_{\mathcal{T}}$ the sequence $(\boldsymbol{R}_j\mathbf{x})_{0\leq j\leq L}$ amounts to computing, while traversing from the leaves to the root, for any parent and both its children the orthogonal projection of a piecewise linear function on the children to the space of linears on the parent. For $d = 2$, the matrix representation of the latter projection is given in Figure 3.2.

Proposition 3.3.4. *The application of $\boldsymbol{B}^{\mathscr{S}}_{\mathcal{T}}$ can be computed in $\mathcal{O}(\#\mathcal{T})$ operations.*

Proof. Because the number of nodes in a binary tree is less than 2 times the number of its leaves, for $\mathbf{x} \in \mathbb{R}^{\dim \mathscr{S}^{0,1}_{\mathcal{T},0}}$ the computation of the sequence $(\boldsymbol{R}_j\boldsymbol{E}_{\mathcal{T}}\mathbf{x})_{0\leq j\leq L}$ takes $\mathcal{O}(\#\mathcal{T})$ operations. From Proposition 3.3.2 recall that any vector in $\operatorname{ran} \boldsymbol{H}_j\boldsymbol{R}_j - \boldsymbol{P}_j\boldsymbol{H}_{j-1}\boldsymbol{R}_{j-1}$ vanishes at M^0_j, so that the number of its non-zero entries is bounded by $\#(N^0_j \setminus M^0_j) \leq 3\#(N^0_j \setminus N^0_{j-1})$. Knowing already $\boldsymbol{R}_j\boldsymbol{E}_{\mathcal{T}}\mathbf{x}$ and $\boldsymbol{R}_{j-1}\boldsymbol{E}_{\mathcal{T}}\mathbf{x}$, computing any non-zero entry of $(\boldsymbol{H}_j\boldsymbol{R}_j - \boldsymbol{P}_j\boldsymbol{H}_{j-1}\boldsymbol{R}_{j-1})\boldsymbol{E}_{\mathcal{T}}\mathbf{x}$ requires $\mathcal{O}(1)$ operations. □

We conclude that the operator $B^{\mathscr{S}}_{\mathcal{T}}$, with above matrix representation $\boldsymbol{B}^{\mathscr{S}}_{\mathcal{T}}$, satisfies the requirements of Theorem 3.2.1.



$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & 0 & \frac{1}{2} \\ \frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} & \frac{1}{2} & 0 \end{bmatrix}$$

FIGURE 3.2. Numbering of the vertices of the parent and that of both children for $d = 2$, and the resulting matrix representation of the orthogonal projection of the space of piecewise linears on the children to the space of linears on the parent.

3.4 Manifold case

Let Γ be a compact d-dimensional Lipschitz, piecewise smooth manifold in $\mathbb{R}^{d'}$ for some $d' \geq d$ with or without boundary $\partial\Gamma$. For some closed measurable $\gamma \subset \partial\Gamma$ and $s \in [0, 1]$, let

$$\mathscr{W} := [L_2(\Gamma), H^1_{0,\gamma}(\Gamma)]_{s,2}, \quad \mathscr{V} := \mathscr{W}'.$$

We assume that Γ is given as the essentially disjoint union of $\cup_{i=1}^p \overline{\chi_i(\Omega_i)}$, with, for $1 \leq i \leq p$, $\chi_i \colon \mathbb{R}^d \to \mathbb{R}^{d'}$ being some smooth regular parametrization, and $\Omega_i \subset \mathbb{R}^d$ an open polytope. W.l.o.g. assuming that for $i \neq j$, $\overline{\Omega}_i \cap \overline{\Omega}_j = \emptyset$, we define

$$\chi \colon \Omega := \cup_{i=1}^p \Omega_i \to \cup_{i=1}^p \chi_i(\Omega_i) \text{ by } \chi|_{\Omega_i} = \chi_i.$$

Let $\mathbb{T}$ be a family of conforming partitions $\mathcal{T}$ of Γ into 'panels' such that, for $1 \leq i \leq p$, $\chi^{-1}(\mathcal{T}) \cap \Omega_i$ is a uniformly shape regular conforming partition of Ω_i into d-simplices (that for $d = 1$ satisfies a uniform K-mesh property). We assume that γ is a (possibly empty) union of 'faces' of $T \in \mathcal{T}$ (i.e., sets of type $\chi_i(e)$, where e is a $(d-1)$-dimensional face of $\chi_i^{-1}(T)$).

We set

$$\mathscr{V}_\mathcal{T} := \{u \in L_2(\Gamma) \colon u \circ \chi|_{\chi^{-1}(T)} \in \mathcal{P}_0 \ (T \in \mathcal{T})\} \subset \mathscr{V},$$

or

$$\mathscr{V}_\mathcal{T} := \{u \in C(\Gamma) \colon u \circ \chi|_{\chi^{-1}(T)} \in \mathcal{P}_1 \ (T \in \mathcal{T})\} \subset \mathscr{V},$$

equipped with canonical basis $\Xi_\mathcal{T}$, and, for the construction of a preconditioner,

$$\mathscr{S}^{0,1}_{\mathcal{T},0} := \{u \in H^1_{0,\gamma}(\Gamma) \colon u \circ \chi|_{\chi^{-1}(T)} \in \mathcal{P}_1 \ (T \in \mathcal{T})\} \subset \mathscr{W},$$

equipped with canonical basis $\Phi_\mathcal{T}$.

As in the domain case, a space $\mathscr{W}_\mathcal{T} \subset \mathscr{W}$ can be constructed with $\dim \mathscr{W}_\mathcal{T} = \dim V_\mathcal{T}$ and $\inf_{\mathcal{T}\in\mathbb{T}} \inf_{0\neq v\in\mathscr{V}_\mathcal{T}} \sup_{0\neq w\in\mathscr{W}_\mathcal{T}} \frac{\langle v,w\rangle_{L_2(\Gamma)}}{\|v\|_\mathscr{V}\|w\|_\mathscr{W}} > 0$, which can be equipped with a locally supported basis $\Psi_\mathcal{T}$ that, modulo a scaling, is $L_2(\Gamma)$-biorthogonal to $\Xi_\mathcal{T}$. Now assuming that a family of $B^\mathscr{S}_\mathcal{T} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}^{0,1}_{\mathcal{T},0}, (\mathscr{S}^{0,1}_{\mathcal{T},0})')$ uniformly is available, the construction of an optimal preconditioner $G_\mathcal{T}$ follows exactly the same lines as outlined in Sect. 3.2 for the domain case.

For the case that Γ is not piecewise polytopal, a hidden problem is, however, that above construction of $\Psi_\mathcal{T}$ requires exact integration of lifted polynomials over the manifold. To circumvent this problem, in Sect. 2.3.2 we have relaxed the condition of $L_2(\Gamma)$-biorthogonality of $\Xi_\mathcal{T}$ and $\Psi_\mathcal{T}$ to biorthogonality w.r.t. to a mesh-dependent scalar product obtained from the $L_2(\Gamma)$-scalar product by replacing the Jacobian on the pull back of each panel by its mean. It was shown that the resulting preconditioner is still optimal, and that the expression for its matrix representation (for the moment without the representation of $B^\mathscr{S}_\mathcal{T}$), that was recalled in Sect. 3.2.2 for the piecewise constant case, applies verbatim by only reading $|T|$ as the volume of the panel.[2]

It remains to discuss the construction of an operator $B^\mathscr{S}_\mathcal{T}$ of multi-level type, where it is now assumed that $\mathbb{T}$ is a family corresponding to newest vertex bisection. An exact copy of the construction of $B^\mathscr{S}_\mathcal{T}$ given in the domain case would require the application of the panel-wise $L_2(T)$-orthogonal projector Q_T, cf. (3.6), which generally poses a quadrature problem. Reconsidering the domain case, the proof of [WZ17, Lemma 3.7] (which provides the proof of the inequality '$\gtrsim$' in our Theorem 3.3.1) builds on the fact that for $\mathcal{T}_0 \prec \mathcal{T}_1 \prec \cdots$ being a sequence of uniformly refined partitions, the decomposition $\mathscr{S}^{0,1}_{\mathcal{T}_L,0} = \sum_{j=0}^L \mathscr{S}^{0,1}_{\mathcal{T}_j,0} \cap (\mathscr{S}^{0,1}_{\mathcal{T}_{j-1},0})^{\perp_{L_2(\Omega)}}$, where $\mathscr{S}^{0,1}_{\mathcal{T}_{-1},0} := \{0\}$, is stable, uniformly in L, w.r.t. the norm on $\mathscr{W}$. This stability holds also true when the orthogonal complements are taken w.r.t. a weighted $L_2(\Omega)$-scalar product, for any weight w with $w,\, 1/w \in L_\infty(\Omega)$.

This has the consequence that for the construction of the multi-level operator $B^\mathscr{S}_\mathcal{T}$ in the manifold case, we may equip $L_2(\Gamma)$ with scalar product

$$\sum_{i=1}^p \int_{\Omega_i} u(\chi_i(x))v(\chi_i(x))\,dx,$$

which is constructed from the canonical $L_2(\Gamma)$-scalar product by simply omitting the Jacobians $|\partial\chi_i(x)|$. With this modified scalar product, the panel-wise orthogonal projector Q_T is the same as in the domain case. We conclude that the resulting $B^\mathscr{S}_\mathcal{T}$ as in (3.7) is in $\mathcal{L}\mathrm{is}_c(\mathscr{S}^{0,1}_{\mathcal{T},0}, (\mathscr{S}^{0,1}_{\mathcal{T},0})')$ uniformly, and that its application can be performed in linear complexity. Indeed, its implementation is equal as in the domain case as described in Sect. 3.3.2 when $|T|$ in (3.8) is read as $|\chi^{-1}(T)|$.

[2] In order to avoid the exact computation of this volume, actually it may read as $|\chi^{-1}(T)||\partial\chi(z)|$ for arbitrary $z \in \chi^{-1}(T)$.

3.5 Numerical experiments

Let $\Gamma = \partial[0,1]^3 \subset \mathbb{R}^3$ be the two-dimensional manifold without boundary given as the boundary of the unit cube, $\mathscr{W} := H^{1/2}(\Gamma)$, $\mathscr{V} := H^{-1/2}(\Gamma)$. We consider the trial space $\mathscr{V}_\mathcal{T} = \mathscr{S}^{-1,0}_\mathcal{T} \subset \mathscr{V}$ of discontinuous piecewise constants. We will evaluate preconditioning of the discretized single layer operator $A_\mathcal{T} \in \mathcal{L}\mathrm{is}_c(\mathscr{V}_\mathcal{T}, \mathscr{V}_\mathcal{T}')$.

The role of the opposite order operator in $\mathcal{L}\mathrm{is}_c(\mathscr{S}^{0,1}_{\mathcal{T},0}, (\mathscr{S}^{0,1}_{\mathcal{T},0})')$ from Sect. 3.2.1 will be fulfilled by the multi-level operator $B^\mathscr{S}_\mathcal{T}$ from (3.7). Equipping $\mathscr{S}^{0,1}_{\mathcal{T},0}$ with the nodal basis $\Phi_\mathcal{T}$, the matrix representation of the preconditioner $G_\mathcal{T}$ from Sect. 3.2.1 reads as

$$\boldsymbol{G}_\mathcal{T} = \boldsymbol{D}_\mathcal{T}^{-1}\big(\boldsymbol{p}_\mathcal{T}^\top \boldsymbol{B}^\mathscr{S}_\mathcal{T} \boldsymbol{p}_\mathcal{T} + \beta \boldsymbol{q}_\mathcal{T}^\top \boldsymbol{D}_\mathcal{T}^{1/2} \boldsymbol{q}_\mathcal{T}\big)\boldsymbol{D}_\mathcal{T}^{-1},$$

for $\boldsymbol{D}_\mathcal{T} = \mathrm{diag}\{|T| : T \in \mathcal{T}\}$, uniformly sparse $\boldsymbol{p}_\mathcal{T}$ and $\boldsymbol{q}_\mathcal{T}$ as given in Sect. 3.2.1, and with the representation of the multi-level operator $\boldsymbol{B}^\mathscr{S}_\mathcal{T}$ given by

$$\boldsymbol{B}^\mathscr{S}_\mathcal{T} = \boldsymbol{E}_\mathcal{T}^\top \Big(\sum_{j=0}^{L} (\boldsymbol{H}_j \boldsymbol{R}_j - \boldsymbol{P}_j \boldsymbol{H}_{j-1} \boldsymbol{R}_{j-1})^\top 2^{-j/2} (\boldsymbol{H}_j \boldsymbol{R}_j - \boldsymbol{P}_j \boldsymbol{H}_{j-1} \boldsymbol{R}_{j-1}) \Big) \boldsymbol{E}_\mathcal{T},$$

for the representations $\boldsymbol{E}_\mathcal{T}, \boldsymbol{H}_j, \boldsymbol{R}_j$ and $\boldsymbol{P}_j$ as provided in Sect. 3.3.2 (the minor adaptations in the manifold case described in Sect. 3.4 to the matrix representations from Sections 3.2.1 and 3.3.2 vanish in the current simple case).

The BEM++ software package [ŚBA+15] is used to approximate the matrix representation of the discretized single layer operator $\boldsymbol{A}_\mathcal{T}$ by hierarchical matrices based on adaptive cross approximation [Hac99, Beb00].

Equipping $\mathscr{V}_\mathcal{T}$ and $\mathbb{R}^{\dim \mathscr{V}_\mathcal{T}}$ with 'energy-norms' $\sqrt{(A_\mathcal{T}\cdot)(\cdot)}$ or $\|\boldsymbol{A}_\mathcal{T}^{\frac{1}{2}} \cdot\|$, respectively, we calculated the (spectral) condition numbers $\kappa_{\mathcal{L}(\mathscr{V}_\mathcal{T}, \mathscr{V}_\mathcal{T})}(G_\mathcal{T} A_\mathcal{T}) = \kappa_{\mathcal{L}(\mathbb{R}^{\dim \mathscr{V}_\mathcal{T}}, \mathbb{R}^{\dim \mathscr{V}_\mathcal{T}})}(\boldsymbol{G}_\mathcal{T} \boldsymbol{A}_\mathcal{T}) = \rho(\boldsymbol{G}_\mathcal{T} \boldsymbol{A}_\mathcal{T}) \rho((\boldsymbol{G}_\mathcal{T} \boldsymbol{A}_\mathcal{T})^{-1})$, where $\rho(\cdot)$ is the spectral radius, using the Lanczos method.

As initial partition $\mathcal{T}_\perp = \mathcal{T}_1$ of Γ we take a conforming partition consisting of 2 triangles per side, so 12 triangles in total, with an assignment of the newest vertices that satisfies the matching condition. We fixed $\beta = 5.3$, being the value for which, for a relative small uniform refinement $\mathcal{T}$ of $\mathcal{T}_\perp$, we found $\rho(\boldsymbol{D}_\mathcal{T}^{-1} \boldsymbol{p}_\mathcal{T}^\top \boldsymbol{B}^\mathscr{S}_\mathcal{T} \boldsymbol{p}_\mathcal{T} \boldsymbol{D}_\mathcal{T}^{-1} \boldsymbol{A}_\mathcal{T}) = \rho(\boldsymbol{D}_\mathcal{T}^{-1} \beta \boldsymbol{q}_\mathcal{T}^\top \boldsymbol{D}_\mathcal{T}^{1/2} \boldsymbol{q}_\mathcal{T} \boldsymbol{D}_\mathcal{T}^{-1} \boldsymbol{A}_\mathcal{T})$.

3.5.1 Uniform refinements

Here we let $\mathbb{T}$ be the sequence $\{\mathcal{T}_k\}_{k\geq 1}$ of (conforming) uniform refinements, that is, $\mathcal{T}_k \succ \mathcal{T}_{k-1}$ is found by bisecting each triangle from $\mathcal{T}_{k-1}$ into 2 subtriangles using Newest Vertex Bisection.

Table 3.1 shows the condition numbers of the preconditioned system in this situation. The condition numbers are relatively small, and the timing results show that the implementation of the preconditioner is indeed linear.

TABLE 3.1. Spectral condition numbers of the preconditioned single layer system discretized by piecewise constants $\mathscr{S}_{\mathcal{T}}^{-1,0}$, using uniform refinements. Preconditioner $\boldsymbol{G}_{\mathcal{T}}$ is constructed using the multi-level operator with $\beta = 5.3$. The last column indicates the number of seconds per degree of freedom per application of $\boldsymbol{G}_{\mathcal{T}}$.

dofs	$\kappa_S(\boldsymbol{A}_{\mathcal{T}})$	$\kappa_S(\boldsymbol{G}_{\mathcal{T}}\boldsymbol{A}_{\mathcal{T}})$	sec / dof
12	14.5	2.6	$2.6 \cdot 10^{-5}$
48	31.0	2.7	$1.4 \cdot 10^{-5}$
192	59.9	2.8	$4.9 \cdot 10^{-6}$
768	118.7	3.3	$1.4 \cdot 10^{-6}$
3072	234.6	3.8	$6.3 \cdot 10^{-7}$
12288	450.4	4.1	$3.3 \cdot 10^{-6}$
49152	852.5	4.3	$6.5 \cdot 10^{-7}$
196608	1566.4	4.5	$7.3 \cdot 10^{-7}$
786432	2730.5	4.6	$7.8 \cdot 10^{-7}$

3.5.2 Local refinements

Here we take $\mathbb{T}$ as a sequence $\{\mathcal{T}_k\}_{k\geq 1}$ of (conforming) locally refined partitions, where $\mathcal{T}_k \succ \mathcal{T}_{k-1}$ is constructed by applying Newest Vertex Bisection to all triangles in $\mathcal{T}_{k-1}$ that touch a corner of the cube.

Table 3.2 contains results for the preconditioned single layer operator discretized by piecewise constants $\mathscr{S}_{\mathcal{T}}^{-1,0}$. The preconditioned condition numbers are nicely bounded, and the timing results confirm that our implementation of the preconditioner is of linear complexity, also in the case of locally refined partitions.

Table 3.2. Spectral condition numbers of the preconditioned single layer system discretized by piecewise constants $\mathscr{S}_{\mathcal{T}}^{-1,0}$, using local refinements at each of the eight cube corners. Operator $\boldsymbol{G}_{\mathcal{T}}$ is applied using the multi-level operator with $\beta = 5.3$. The second column is defined by $h_{\mathcal{T},min} := \min_{T\in\mathcal{T}} \sqrt{|T|}$. The last column indicates the number of seconds per degree of freedom per application of $\boldsymbol{G}_{\mathcal{T}}$.

dofs	$h_{\mathcal{T},min}$	$\kappa_S(\boldsymbol{G}_{\mathcal{T}}\boldsymbol{A}_{\mathcal{T}})$	sec / dof
12	$1.4 \cdot 10^{0}$	2.63	$2.5 \cdot 10^{-5}$
336	$8.8 \cdot 10^{-2}$	2.73	$2.4 \cdot 10^{-6}$
720	$5.5 \cdot 10^{-3}$	2.91	$1.8 \cdot 10^{-6}$
1104	$3.4 \cdot 10^{-4}$	2.96	$1.8 \cdot 10^{-6}$
1488	$2.1 \cdot 10^{-5}$	2.99	$2.2 \cdot 10^{-6}$
1872	$1.3 \cdot 10^{-6}$	2.98	$2.0 \cdot 10^{-6}$
2256	$8.4 \cdot 10^{-8}$	3.00	$2.3 \cdot 10^{-6}$
2640	$5.2 \cdot 10^{-9}$	3.00	$2.0 \cdot 10^{-6}$
3024	$3.2 \cdot 10^{-10}$	3.01	$2.3 \cdot 10^{-6}$
3408	$2.0 \cdot 10^{-11}$	3.01	$2.5 \cdot 10^{-6}$
3696	$2.5 \cdot 10^{-12}$	3.01	$2.6 \cdot 10^{-6}$

4 Problems of positive order

4.1 Introduction

This chapter deals with the construction of *uniform* preconditioners for operators of *positive* order, using the framework of 'operator preconditioning' as described in [Hip06], see e.g. [SW98, CN00] for earlier work. It will build on our experiences with this approach for problems of negative order developed in Chapter 2.

For some d-dimensional domain (or manifold) Ω, a measurable, closed, possibly empty $\gamma \subset \partial\Omega$, and an $s \in [0,1]$, we consider the Sobolev space

$$\mathscr{V} := [L_2(\Omega), H^1_{0,\gamma}(\Omega)]_{s,2},$$

with $H^1_{0,\gamma}(\Omega)$ being the closure in $H^1(\Omega)$ of the smooth functions on Ω that vanish at γ. For $\mathscr{V}_\mathcal{T} \subset \mathscr{V}$ a closed, e.g. finite dimensional subspace, and $A_\mathcal{T}\colon \mathscr{V}_\mathcal{T} \to \mathscr{V}_\mathcal{T}'$ some boundedly invertible linear operator, we are interested in constructing a *preconditioner* $G_\mathcal{T}\colon \mathscr{V}_\mathcal{T}' \to \mathscr{V}_\mathcal{T}$. More specifically, thinking of a *family* of spaces $\mathscr{V}_\mathcal{T}$ and operators $A_\mathcal{T}\colon \mathscr{V}_\mathcal{T} \to \mathscr{V}_\mathcal{T}'$, our aim is to construct preconditioners $G_\mathcal{T}$ such that $G_\mathcal{T}A_\mathcal{T}\colon \mathscr{V}_\mathcal{T} \to \mathscr{V}_\mathcal{T}$ is uniformly boundedly invertible.

It is well-known that such preconditioners of multi-level type are available. The advantage of operator preconditioning is, however, that it does not require a hierarchy of trial spaces.

In order to apply the operator preconditioning framework, one needs to construct families of closed subspaces $\mathscr{W}_\mathcal{T} \subset \mathscr{W} := \mathscr{V}'$, uniformly boundedly invertible $B_\mathcal{T}\colon \mathscr{W}_\mathcal{T} \to \mathscr{W}_\mathcal{T}'$, and uniformly boundedly invertible $D_\mathcal{T}\colon \mathscr{V}_\mathcal{T} \to \mathscr{W}_\mathcal{T}'$. Then the resulting preconditioners $G_\mathcal{T}$ are of the form

$$G_\mathcal{T} := D_\mathcal{T}^{-1} B_\mathcal{T} (D_\mathcal{T}')^{-1}.$$

The canonical setting is that for $A\colon \mathscr{V} \to \mathscr{V}'$, i.e., an operator of order $2s$, and an *opposite order* operator $B\colon \mathscr{W} \to \mathscr{W}'$, both boundedly invertible and coercive, it holds that $(A_\mathcal{T}u)(v) := (Au)(v)$ $(u, v \in \mathscr{V}_\mathcal{T})$, $(B_\mathcal{T}u)(v) := (Bu)(v)$ $(u, v \in \mathscr{W}_\mathcal{T})$, and $(D_\mathcal{T}u)(v) := \langle u, v\rangle_{L_2(\Omega)}$ $(u \in \mathscr{V}_\mathcal{T}, v \in \mathscr{W}_\mathcal{T})$. A typical example

for $s = 1/2$ is that A is the *Hypersingular Integral* operator, and B is the *Weakly Singular Integral* operator, see [SW98].

A careful selection of $\mathscr{W}_\mathcal{T}$ has to be made to ensure that $D_\mathcal{T}\colon \mathscr{V}_\mathcal{T} \to \mathscr{W}_\mathcal{T}'$ is uniformly boundedly invertible. A suitable family of $(\mathscr{V}_\mathcal{T}, \mathscr{W}_\mathcal{T})$ pairs has been introduced in [Ste02, BC07]. Here $\mathcal{T}$ is a triangular partition of a *two-dimensional* domain or manifold, $\mathscr{V}_\mathcal{T}$ is the space of continuous piecewise linears w.r.t. $\mathcal{T}$, and $\mathscr{W}_\mathcal{T}$ is a subspace of the space of piecewise constants w.r.t. a barycentric refinement of $\mathcal{T}$, constructed by subdividing each triangle into 6 subtriangles by connecting its vertices and midpoints with its barycenter. It has been shown in [Ste02, HUT16] that the preconditioner arising from these pairs $(\mathscr{V}_\mathcal{T}, \mathscr{W}_\mathcal{T})$ is a uniform preconditioner for families of partitions that satisfy a certain mildly-grading condition.

A problem with the constructions from [Ste02, BC07] appears when one considers the matrix representation $G_\mathcal{T}$ in the standard bases, i.e. $\boldsymbol{G}_\mathcal{T} = \boldsymbol{D}_\mathcal{T}^{-1}\boldsymbol{B}_\mathcal{T}\boldsymbol{D}_\mathcal{T}^{-\top}$. Indeed, this matrix $\boldsymbol{D}_\mathcal{T}$ is *not* diagonal, and its inverse is densely populated so that it has to be approximated. Moreover, in order to get a uniform preconditioner, since $G_\mathcal{T}$, being spectrally equivalent with $A_\mathcal{T}^{-1}$, gets increasingly ill-conditioned with a decreasing minimal mesh-size, the accuracy with which $\boldsymbol{D}_\mathcal{T}^{-1}$ has to be approximated increases with a decreasing minimal mesh-size. As a result, an application of $\boldsymbol{D}_\mathcal{T}^{-1}$ cannot be expected to execute in linear time.

Another (practical) issue with these constructions is the need for the construction of the non-standard *barycentrical* refinement of $\mathcal{T}$. This refinement increases the number of elements by a factor 6, and therefore also increases the cost of evaluating $B_\mathcal{T} : \mathscr{W}_\mathcal{T} \to \mathscr{W}_\mathcal{T}'$.

4.1.1 Contributions

With $\mathscr{V}_\mathcal{T}$ being the space of continuous piecewise linears, the construction of $\mathscr{W}_\mathcal{T}$ presented in this chapter improves on the existing approach from [Ste02, BC07] concerning the following aspects:

- The matrix representation $\boldsymbol{D}_\mathcal{T}$ of $D_\mathcal{T}$ will be diagonal, allowing one to (exactly) evaluate $\boldsymbol{D}_\mathcal{T}^{-1}$ in linear time;
- The operator $G_\mathcal{T}$ will be a uniformly well-conditioned preconditioner for families of uniformly shape regular partitions, without requiring a mildly-grading assumption on the partitions;
- By using a stable decomposition of an enclosing space of $\mathscr{W}_\mathcal{T}$ into a standard finite element space $\mathscr{U}_\mathcal{T}$ w.r.t. $\mathcal{T}$ (either being the space of piecewise constants or $\mathscr{U}_\mathcal{T} = \mathscr{V}_\mathcal{T}$) and some bubble space, our $B_\mathcal{T}$ will be the *sum* of the corresponding Galerkin discretization operator of the opposite order operator B, and an operator whose representation is a diagonal, with which the undesired barycentrical refinement is avoided;

- The construction of $\mathscr{W}_{\mathcal{T}}$ applies in any space dimension, and extends to non piecewise planar manifolds.

We will extend the preconditioners to higher order finite element spaces by applying a subspace correction framework.

Due to the interchanged roles of primal and dual spaces, compared to our work in Chapter 2 on preconditioning operators of negative order, here the stable construction of $\mathscr{W}_{\mathcal{T}}$ is simpler, but, on the other hand, the stable decomposition of an enclosing space of $\mathscr{W}_{\mathcal{T}}$ is more delicate.

4.1.2 Outline

Sect. 4.1.3 recalls some notation that will be used throughout the article. In Sect. 4.2 the general theory of operator preconditioning is summarized. In Sect. 4.3, the framework is specialized to operators of positive order discretized with *continuous* piecewise linears. Sect. 4.4 give two constructions of $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{W}_{\mathcal{T}}, \mathscr{W}_{\mathcal{T}}')$ that avoid any refinement of the partition $\mathcal{T}$ that underlies the trial space $\mathscr{V}_{\mathcal{T}}$. In Sect. 4.5 the preconditioners are generalized to higher order finite element spaces, and to spaces defined on manifolds. Finally, in Sect. 4.6 we report some numerical results obtained with the new preconditioners.

4.1.3 Notations

By $\lambda \lesssim \mu$ we will mean that λ can be bounded by a multiple of μ, independently of parameters which λ and μ may depend on, with the sole exception of the space dimension d, or in the manifold case, on the parametrization of the manifold that is used to define the finite element spaces on it. Obviously, $\lambda \gtrsim \mu$ is defined as $\mu \lesssim \lambda$, and $\lambda \eqsim \mu$ as $\lambda \lesssim \mu$ and $\lambda \gtrsim \mu$.

For normed linear spaces $\mathscr{Y}$ and $\mathscr{Z}$, in this work for convenience over $\mathbb{R}$, $\mathcal{L}(\mathscr{Y}, \mathscr{Z})$ will denote the space of bounded linear mappings $\mathscr{Y} \to \mathscr{Z}$ endowed with the operator norm $\|\cdot\|_{\mathcal{L}(\mathscr{Y},\mathscr{Z})}$. The subset of invertible operators in $\mathcal{L}(\mathscr{Y}, \mathscr{Z})$ with inverses in $\mathcal{L}(\mathscr{Z}, \mathscr{Y})$ will be denoted as $\mathcal{L}\mathrm{is}(\mathscr{Y}, \mathscr{Z})$. The *condition number* of a $C \in \mathcal{L}\mathrm{is}(\mathscr{Y}, \mathscr{Z})$ is defined as $\kappa_{\mathscr{Y},\mathscr{Z}}(C) := \|C\|_{\mathcal{L}(\mathscr{Y},\mathscr{Z})}\|C^{-1}\|_{\mathcal{L}(\mathscr{Z},\mathscr{Y})}$.

For $\mathscr{Y}$ a reflexive Banach space and $C \in \mathcal{L}(\mathscr{Y}, \mathscr{Y}')$ being *coercive*, i.e.,

$$\inf_{0\neq y\in\mathscr{Y}} \frac{(Cy)(y)}{\|y\|^2_{\mathscr{Y}}} > 0,$$

both C and $\Re(C) := \frac{1}{2}(C + C')$ are in $\mathcal{L}\mathrm{is}(\mathscr{Y}, \mathscr{Y}')$ with

$$\begin{aligned}\|\Re(C)\|_{\mathcal{L}(\mathscr{Y},\mathscr{Y}')} &\leq \|C\|_{\mathcal{L}(\mathscr{Y},\mathscr{Y}')},\\ \|C^{-1}\|_{\mathcal{L}(\mathscr{Y}',\mathscr{Y})} \leq \|\Re(C)^{-1}\|_{\mathcal{L}(\mathscr{Y}',\mathscr{Y})} &= \Big(\inf_{0\neq y\in\mathscr{Y}} \frac{(Cy)(y)}{\|y\|^2_{\mathscr{Y}}}\Big)^{-1}.\end{aligned}$$

The subset of coercive operators in $\mathcal{L}\mathrm{is}(\mathscr{Y},\mathscr{Y}')$ is denoted as $\mathcal{L}\mathrm{is}_c(\mathscr{Y},\mathscr{Y}')$. If $C\in\mathcal{L}\mathrm{is}_c(\mathscr{Y},\mathscr{Y}')$, then $C^{-1}\in\mathcal{L}\mathrm{is}_c(\mathscr{Y}',\mathscr{Y})$ and $\|\Re(C^{-1})^{-1}\|_{\mathcal{L}(\mathscr{Y},\mathscr{Y}')}\leq\|C\|^2_{\mathcal{L}(\mathscr{Y},\mathscr{Y}')}\|\Re(C)^{-1}\|_{\mathcal{L}(\mathscr{Y}',\mathscr{Y})}$.

Given a family of operators $C_i\in\mathcal{L}\mathrm{is}(\mathscr{Y}_i,\mathscr{Z}_i)$ ($\mathcal{L}\mathrm{is}_c(\mathscr{Y}_i,\mathscr{Z}_i)$), we will write $C_i\in\mathcal{L}\mathrm{is}(\mathscr{Y}_i,\mathscr{Z}_i)$ ($\mathcal{L}\mathrm{is}_c(\mathscr{Y}_i,\mathscr{Z}_i)$) uniformly in i, or simply 'uniform', when

$$\sup_i\max(\|C_i\|_{\mathcal{L}(\mathscr{Y}_i,\mathscr{Z}_i)},\|C_i^{-1}\|_{\mathcal{L}(\mathscr{Z}_i,\mathscr{Y}_i)})<\infty,$$

or

$$\sup_i\max(\|C_i\|_{\mathcal{L}(\mathscr{Y}_i,\mathscr{Z}_i)},\|\Re(C_i)^{-1}\|_{\mathcal{L}(\mathscr{Z}_i,\mathscr{Y}_i)})<\infty.$$

Given a finite collection $\Upsilon=\{\upsilon\}$ in a linear space, we set the *synthesis operator*

$$\mathcal{F}_\Upsilon:\mathbb{R}^{\#\Upsilon}\to\operatorname{span}\Upsilon:\mathbf{c}\mapsto\mathbf{c}^\top\Upsilon:=\sum_{\upsilon\in\Upsilon}c_\upsilon\upsilon.$$

Equipping $\mathbb{R}^{\#\Upsilon}$ with the Euclidean scalar product $\langle\cdot,\cdot\rangle$, and identifying $(\mathbb{R}^{\#\Upsilon})'$ with $\mathbb{R}^{\#\Upsilon}$ using the corresponding Riesz map, we infer that the adjoint of $\mathcal{F}_\Upsilon$, known as the *analysis operator*, satisfies

$$\mathcal{F}'_\Upsilon:(\operatorname{span}\Upsilon)'\to\mathbb{R}^{\#\Upsilon}:f\mapsto f(\Upsilon):=[f(\upsilon)]_{\upsilon\in\Upsilon}.$$

A collection Υ is a *basis* for its span when $\mathcal{F}_\Upsilon\in\mathcal{L}\mathrm{is}(\mathbb{R}^{\#\Upsilon},\operatorname{span}\Upsilon)$ (and so $\mathcal{F}'_\Upsilon\in\mathcal{L}\mathrm{is}((\operatorname{span}\Upsilon)',\mathbb{R}^{\#\Upsilon})$.)

Two countable collections $\Upsilon=(\upsilon_i)_i$ and $\tilde\Upsilon=(\tilde\upsilon_i)_i$ in a Hilbert space will be called *biorthogonal* when $\langle\Upsilon,\tilde\Upsilon\rangle=[\langle\upsilon_j,\tilde\upsilon_i\rangle]_{ij}$ is an *invertible diagonal* matrix, and *biorthonormal* when it is the *identity* matrix.

4.2 Operator preconditioning

We shortly recap the idea of opposite order preconditioning, which is based on the following result, see [Hip06, Sect. 2].

Proposition 4.2.1. *Let $\mathscr{V},\mathscr{W}$ be reflexive Banach spaces.*
If $B\in\mathcal{L}\mathrm{is}(\mathscr{W},\mathscr{W}')$ and $D\in\mathcal{L}\mathrm{is}(\mathscr{V},\mathscr{W}')$, then

$$G:=D^{-1}B(D')^{-1}\in\mathcal{L}\mathrm{is}(\mathscr{V}',\mathscr{V}),$$

and

$$\begin{aligned}\|G\|_{\mathcal{L}(\mathscr{V}',\mathscr{V})}&\leq\|D^{-1}\|^2_{\mathcal{L}(\mathscr{W}',\mathscr{V})}\|B\|_{\mathcal{L}(\mathscr{W},\mathscr{W}')},\\ \|G^{-1}\|_{\mathcal{L}(\mathscr{V},\mathscr{V}')}&\leq\|D\|^2_{\mathcal{L}(\mathscr{V},\mathscr{W}')}\|B^{-1}\|_{\mathcal{L}(\mathscr{W}',\mathscr{W})}.\end{aligned}$$

If additionally $B\in\mathcal{L}\mathrm{is}_c(\mathscr{W},\mathscr{W}')$, then $G\in\mathcal{L}\mathrm{is}_c(\mathscr{V}',\mathscr{V})$, and

$$\|\Re(G)^{-1}\|_{\mathcal{L}(\mathscr{V},\mathscr{V}')}\leq\|D\|^2_{\mathcal{L}(\mathscr{V},\mathscr{W}')}\|\Re(B)^{-1}\|_{\mathcal{L}(\mathscr{W}',\mathscr{W})}.$$

Let be given families of finite dimensional spaces $\mathscr{V}_\mathcal{T}$ for $\mathcal{T} \in \mathbb{T}$, and operators $A_\mathcal{T} \in \mathcal{L}\mathrm{is}(\mathscr{V}_\mathcal{T}, \mathscr{V}_\mathcal{T}')$ uniformly in $\mathcal{T} \in \mathbb{T}$. Then in light of Proposition 4.2.1 we will seek preconditioners for $A_\mathcal{T}$ of the form

$$G_\mathcal{T} = D_\mathcal{T}^{-1} B_\mathcal{T} (D_\mathcal{T}')^{-1},$$

where $B_\mathcal{T} \in \mathcal{L}\mathrm{is}(\mathscr{W}_\mathcal{T}, \mathscr{W}_\mathcal{T}')$ and $D_\mathcal{T} \in \mathcal{L}\mathrm{is}(\mathscr{V}_\mathcal{T}, \mathscr{W}_\mathcal{T}')$ (both uniformly in $\mathcal{T} \in \mathbb{T}$), and

$$\dim \mathscr{W}_\mathcal{T} = \dim \mathscr{V}_\mathcal{T}. \tag{4.1}$$

A typical situation is that for some reflexive Banach space $\mathscr{V}$ and $A \in \mathcal{L}\mathrm{is}_c(\mathscr{V}, \mathscr{V}')$, it holds that $\mathscr{V}_\mathcal{T} \subset \mathscr{V}$ (thus *equipped with* $\|\cdot\|_\mathscr{V}$) and $(A_\mathcal{T} u)(v) := (Au)(v)$ $(u, v \in \mathscr{V}_\mathcal{T})$, so that indeed $A_\mathcal{T} \in \mathcal{L}\mathrm{is}_c(\mathscr{V}_\mathcal{T}, \mathscr{V}_\mathcal{T}')$ uniformly in $\mathcal{T} \in \mathbb{T}$. Then for a suitable reflexive Banach space $\mathscr{W}$, an operator $B \in \mathcal{L}\mathrm{is}_c(\mathscr{W}, \mathscr{W}')$, and a subspace $\mathscr{W}_\mathcal{T} \subset \mathscr{W}$ (thus *equipped with* $\|\cdot\|_\mathscr{W}$), one can take $(B_\mathcal{T} w)(z) := (Bw)(z)$ $(w, z \in \mathscr{W}_\mathcal{T})$, giving $B_\mathcal{T} \in \mathcal{L}\mathrm{is}_c(\mathscr{W}_\mathcal{T}, \mathscr{W}_\mathcal{T}')$ uniformly. A possible construction of $D_\mathcal{T} \in \mathcal{L}\mathrm{is}(\mathscr{V}_\mathcal{T}, \mathscr{W}_\mathcal{T}')$ uniformly is discussed in the next proposition.

Proposition 4.2.2 (Fortin projector ([For77])). *For some $D \in \mathcal{L}\mathrm{is}(\mathscr{V}, \mathscr{W}')$, let $D_\mathcal{T} \in \mathcal{L}(\mathscr{V}_\mathcal{T}, \mathscr{W}_\mathcal{T}')$ be defined by $(D_\mathcal{T} v)(w) := (Dv)(w)$. Then*

$$\|D_\mathcal{T}\|_{\mathcal{L}(\mathscr{V}_\mathcal{T}, \mathscr{W}_\mathcal{T}')} \leq \|D\|_{\mathcal{L}(\mathscr{V}, \mathscr{W}')}.$$

Assuming (4.1)*, additionally one has $D_\mathcal{T} \in \mathcal{L}\mathrm{is}(\mathscr{V}_\mathcal{T}, \mathscr{W}_\mathcal{T}')$* if*, and for $\mathscr{W}$ being a Hilbert space,* only if *there exists a projector $P_\mathcal{T} \in \mathcal{L}(\mathscr{W}, \mathscr{W})$ onto $\mathscr{W}_\mathcal{T}$ with $(D\mathscr{V}_\mathcal{T})((\mathrm{Id} - P_\mathcal{T})\mathscr{W}) = 0$, in which case*

$$\|D_\mathcal{T}^{-1}\|_{\mathcal{L}(\mathscr{W}_\mathcal{T}', \mathscr{V}_\mathcal{T})} \leq \|P_\mathcal{T}\|_{\mathcal{L}(\mathscr{W}, \mathscr{W})} \|D^{-1}\|_{\mathcal{L}(\mathscr{W}', \mathscr{V})}.$$

In our applications, the choices for $\mathscr{W}$ and D will be obvious, and the key ingredient for the construction of a uniform preconditioner $G_\mathcal{T}$ will be the selection of $\mathscr{W}_\mathcal{T}$ that allows for a uniformly bounded Fortin projector $P_\mathcal{T}$.

4.2.1 Implementation

Let $\Phi_\mathcal{T} = (\phi_i)_i$ and $\Psi_\mathcal{T} = (\psi_i)_i$ be bases for $\mathscr{V}_\mathcal{T}$ and $\mathscr{W}_\mathcal{T}$, respectively. Then in coordinates the preconditioned system reads as

$$\mathcal{F}_{\Phi_\mathcal{T}}^{-1} G_\mathcal{T} A_\mathcal{T} \mathcal{F}_{\Phi_\mathcal{T}} = \boldsymbol{G}_\mathcal{T} \boldsymbol{A}_\mathcal{T} := \boldsymbol{D}_\mathcal{T}^{-1} \boldsymbol{B}_\mathcal{T} \boldsymbol{D}_\mathcal{T}^{-\top} \boldsymbol{A}_\mathcal{T},$$

where

$$\boldsymbol{A}_\mathcal{T} := \mathcal{F}_{\Phi_\mathcal{T}}' A_\mathcal{T} \mathcal{F}_{\Phi_\mathcal{T}}, \quad \boldsymbol{B}_\mathcal{T} := \mathcal{F}_{\Psi_\mathcal{T}}' B_\mathcal{T} \mathcal{F}_{\Psi_\mathcal{T}}, \quad \boldsymbol{D}_\mathcal{T} := \mathcal{F}_{\Psi_\mathcal{T}}' D_\mathcal{T} \mathcal{F}_{\Phi_\mathcal{T}}.$$

By identifying a map in $\mathcal{L}(\mathbb{R}^{\#\Phi_\mathcal{T}}, \mathbb{R}^{\#\Phi_\mathcal{T}})$ with a $\#\Phi_\mathcal{T} \times \#\Phi_\mathcal{T}$ matrix by equipping $\mathbb{R}^{\#\Phi_\mathcal{T}}$ with the canonical basis $(\boldsymbol{e}_i)_i$ one has,

$$(\boldsymbol{A}_\mathcal{T})_{ij} = \langle \mathcal{F}_{\Phi_\mathcal{T}}' A_\mathcal{T} \mathcal{F}_{\Phi_\mathcal{T}} \boldsymbol{e}_j, \boldsymbol{e}_i \rangle = (A_\mathcal{T} \mathcal{F}_{\Phi_\mathcal{T}} \boldsymbol{e}_j)(\mathcal{F}_{\Phi_\mathcal{T}} \boldsymbol{e}_i) = (A_\mathcal{T} \phi_j)(\phi_i),$$

and similarly,

$$(\boldsymbol{B}_{\mathcal{T}})_{ij} = (B_{\mathcal{T}}\psi_j)(\psi_i), \quad (\boldsymbol{D}_{\mathcal{T}})_{ij} = (D_{\mathcal{T}}\phi_j)(\psi_i).$$

Preferably $\boldsymbol{D}_{\mathcal{T}}$ is such that its inverse can be applied in linear complexity, as is the case when $\boldsymbol{D}_{\mathcal{T}}$ is *diagonal*. A goal of this work is to construct such a diagonal $\boldsymbol{D}_{\mathcal{T}}$.

Remark 4.2.3. Using $\sigma(\cdot)$ and $\rho(\cdot)$ to denote the spectrum and spectral radius of an operator, clearly $\sigma(\boldsymbol{G}_{\mathcal{T}}\boldsymbol{A}_{\mathcal{T}}) = \sigma(G_{\mathcal{T}}A_{\mathcal{T}})$. So for the spectral condition number we have

$$\kappa_S(\boldsymbol{G}_{\mathcal{T}}\boldsymbol{A}_{\mathcal{T}}) := \rho(\boldsymbol{G}_{\mathcal{T}}\boldsymbol{A}_{\mathcal{T}})\rho((\boldsymbol{G}_{\mathcal{T}}\boldsymbol{A}_{\mathcal{T}})^{-1}) \leq \kappa_{\mathscr{V}_{\mathcal{T}},\mathscr{V}_{\mathcal{T}}}(G_{\mathcal{T}}A_{\mathcal{T}}),$$

which thus holds true *independently* of the choice of the basis $\Phi_{\mathcal{T}}$ for $\mathscr{V}_{\mathcal{T}}$. Furthermore, in view of an application of Conjugate Gradients, if $A_{\mathcal{T}}$ and $B_{\mathcal{T}}$ are coercive and *self-adjoint*, then $\boldsymbol{A}_{\mathcal{T}}$ and $\boldsymbol{G}_{\mathcal{T}}$ are positive definite and symmetric. Equipping $\mathbb{R}^{\dim \mathscr{V}_{\mathcal{T}}}$ with $|\!|\!|\cdot|\!|\!| := \|(\boldsymbol{G}_{\mathcal{T}})^{-\frac{1}{2}} \cdot \|$ or $|\!|\!|\cdot|\!|\!| := \|(\boldsymbol{A}_{\mathcal{T}})^{\frac{1}{2}} \cdot \|$, in that case we have

$$\kappa_{(\mathbb{R}^{\dim \mathscr{V}_{\mathcal{T}}},|\!|\!|\cdot|\!|\!|),(\mathbb{R}^{\dim \mathscr{V}_{\mathcal{T}}},|\!|\!|\cdot|\!|\!|)}(\boldsymbol{G}_{\mathcal{T}}\boldsymbol{A}_{\mathcal{T}}) = \kappa_S(\boldsymbol{G}_{\mathcal{T}}\boldsymbol{A}_{\mathcal{T}}).$$

4.3 Continuous piecewise linear discretization space

For a bounded polytopal domain $\Omega \subset \mathbb{R}^d$, a measurable, closed, possibly empty $\gamma \subset \partial\Omega$, and an $s \in [0,1]$, we take

$$\mathscr{V} := [L_2(\Omega), H^1_{0,\gamma}(\Omega)]_{s,2}, \quad \mathscr{W} := \mathscr{V}',$$

which forms the Gelfand triple $\mathscr{V} \hookrightarrow L_2(\Omega) \simeq L_2(\Omega)' \hookrightarrow \mathscr{W}$. We define the operator $D \in \mathcal{L}\mathrm{is}(\mathscr{V}, \mathscr{W}')$ as the unique extension to $\mathscr{V} \times \mathscr{W}$ of the duality pairing

$$(Dv)(w) := \langle v, w\rangle_{L_2(\Omega)},$$

which satisfies $\|D\|_{\mathcal{L}(\mathscr{V},\mathscr{W}')} = \|D^{-1}\|_{\mathcal{L}(\mathscr{W}',\mathscr{V})} = 1$.

Let $(\mathcal{T})_{\mathcal{T}\in\mathbb{T}}$ be a family of conforming partitions of Ω into closed uniformly shape regular d-simplices. Thanks to the conformity and the uniform shape regularity, for $d > 1$ we know that neighbouring $T, T' \in \mathcal{T}$, i.e. $T \cap T' \neq \emptyset$, have uniformly comparable sizes. For $d = 1$, we impose this uniform '*K-mesh property*' explicitly.

For some $\mathcal{T} \in \mathbb{T}$, denote $N^0_{\mathcal{T}}$ as the subset of vertices that are not on γ, where we assume that γ is the (possibly empty) union of $(d-1)$-faces of $T \in \mathcal{T}$. For $T \in \mathcal{T}$, write N_T for the set of its vertices, set $N^0_T := N^0_{\mathcal{T}} \cap N_T$, $h_T := |T|^{1/d}$, and the piecewise constant function $h_{\mathcal{T}}$ by $h_{\mathcal{T}}|_T = h_T$ ($T \in \mathcal{T}$). For any vertex $\nu \in N^0_{\mathcal{T}}$, define the patch $\omega_{\mathcal{T},\nu} := \bigcup_{\{T\in\mathcal{T}:\nu\in T\}} T$ and the local mesh size

$h_{\mathcal{T},\nu} := |\omega_{\mathcal{T},\nu}|^{1/d}$. We omit notational dependence on $\mathcal{T}$ if it is clear from the context, and simply write ω_ν and h_ν.

Let the discretization space $\mathscr{V}_\mathcal{T}$ be the space of continuous piecewise linears, zero on γ,

$$\mathscr{V}_\mathcal{T} = \mathscr{S}^{0,1}_\mathcal{T} := \{u \in H^1_{0,\gamma}(\Omega)\colon u|_T \in \mathcal{P}_1\,(T \in \mathcal{T})\} \subset \mathscr{V},$$

equipped with the nodal bases

$$\Phi_\mathcal{T} = \{\phi_\nu\colon \nu \in N^0_\mathcal{T}\}$$

defined by $\phi_\nu(\nu') := \delta_{\nu\nu'}$ $(\nu, \nu' \in N^0_\mathcal{T})$. For future reference, define the space of discontinuous piecewise constants by

$$\mathscr{S}^{-1,0}_\mathcal{T} := \{u \in L_2(\Omega)\colon u|_T \in \mathcal{P}_0\,(T \in \mathcal{T})\} \subset \mathscr{W},$$

equipped with the basis

$$\Sigma_\mathcal{T} := \{\mathbb{1}_T\colon T \in \mathcal{T}\}, \tag{4.2}$$

where $\mathbb{1}_K$ is defined by, for any $K \subseteq \Omega$, $\mathbb{1}_K := 1$ on K, and $\mathbb{1}_K := 0$ elsewhere.

4.3.1 The subspace $\mathscr{W}_\mathcal{T}$

We will construct the preconditioning space $\mathscr{W}_\mathcal{T}$ as

$$\mathscr{W}_\mathcal{T} := \operatorname{span}\Psi_\mathcal{T} \subset \mathscr{W}, \text{ with } \dim\mathscr{W}_\mathcal{T} = \dim\mathscr{V}_\mathcal{T}$$

for a collection $\Psi_\mathcal{T} \subset L_2(\Omega)$ that is biorthogonal to $\Phi_\mathcal{T}$, and for which the biorthogonal projector $P_\mathcal{T} \in \mathcal{L}(\mathscr{W},\mathscr{W})$ onto $\mathscr{W}_\mathcal{T}$ is uniformly bounded. We require the collection $\Psi_\mathcal{T} := \{\psi_\nu \in \mathscr{W} : \nu \in N^0_\mathcal{T}\}$ to satisfy

$$\begin{aligned} \big|\langle\phi_\nu, \psi_{\nu'}\rangle_{L_2(\Omega)}\big| &\eqsim \delta_{\nu\nu'}\|\phi_\nu\|_{L_2(\Omega)}\|\psi_{\nu'}\|_{L_2(\Omega)} \quad (\nu,\nu' \in N^0_\mathcal{T}),\\ \operatorname{supp}\psi_\nu &\subseteq \omega_\nu \quad (\nu \in N^0_\mathcal{T}). \end{aligned} \tag{4.3}$$

Existence of such collections will be shown later in Sect. 4.4.

4.3.2 Bounded Fortin projector

From (4.3) it follows that the biorthogonal Fortin projector $P_\mathcal{T}\colon H^1_{0,\gamma}(\Omega)' \to L_2(\Omega)$ onto $\mathscr{W}_\mathcal{T}$ with $\operatorname{ran}(I - P_\mathcal{T}) = \mathscr{V}_\mathcal{T}^{\perp_{L_2(\Omega)}}$ exists, and is given by

$$P_\mathcal{T} u = \sum_{\nu\in N^0_\mathcal{T}} \frac{\langle u, \phi_\nu\rangle_{L_2(\Omega)}}{\langle\phi_\nu, \psi_\nu\rangle_{L_2(\Omega)}}\psi_\nu.$$

Uniform boundedness of $\|P_\mathcal{T}\|_{\mathcal{L}(\mathscr{W},\mathscr{W})}$ follows from uniform boundedness of its adjoint $P'_\mathcal{T}$, which can be shown similarly as in Theorem 2.5.1[1]:

[1]Note that the roles of $\mathscr{V}$ and $\mathscr{W}$ are interchanged compared to Chapter 2.

Theorem 4.3.1. *It holds that* $\sup_{\mathcal{T}\in\mathbb{T}} \|P_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{W},\mathscr{W})} = \sup_{\mathcal{T}\in\mathbb{T}} \|P'_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{V},\mathscr{V})} < \infty$.

Proof. Let $\mathcal{T} \in \mathbb{T}$. Define $\omega_T^{(0)} := T$ for $T \in \mathcal{T}$, and for $i = 1, \ldots$, denote $\omega_T^{(i)} := \bigcup_{\{T'\in\mathcal{T}\colon T'\cap\omega_T^{(i-1)}\neq\emptyset\}} T'$. The adjoint $P'_{\mathcal{T}}\colon L_2(\Omega) \to H^1_{0,\gamma}(\Omega)$ onto $\mathscr{V}_{\mathcal{T}}$ is given by

$$P'_{\mathcal{T}}u = \sum_{\nu\in N_{\mathcal{T}}^0} \frac{\langle u, \psi_\nu\rangle_{L_2(\Omega)}}{\langle \phi_\nu, \psi_\nu\rangle_{L_2(\Omega)}}\phi_\nu.$$

Properties of the nodal basis functions, $\|\phi_\nu\|^2_{L_2(\Omega)} \eqsim h_\nu^d$ and $\|\phi_\nu\|^2_{H^1(\Omega)} \lesssim h_\nu^{d-2}$, in combination with (4.3), can be used to show that, for $T \in \mathcal{T}$ and $k \in \{0, 1\}$,

$$\begin{aligned}\|P'_{\mathcal{T}}u\|_{H^k(T)} &\leq \sum_{\nu\in N_T^0} \|\phi_\nu\|_{H^k(T)} \frac{\|u\|_{L_2(\operatorname{supp}\psi_\nu)}\|\psi_\nu\|_{L_2(\Omega)}}{|\langle \phi_\nu, \psi_\nu\rangle_{L_2(\Omega)}|}\\ &\lesssim \sum_{\nu\in N_T^0} h_\nu^{-k}\|u\|_{L_2(\operatorname{supp}\psi_\nu)} \lesssim h_T^{-k}\|u\|_{L_2(\omega_T^{(1)})},\end{aligned} \tag{4.4}$$

from which we may directly conclude that

$$\sup_{\mathcal{T}\in\mathbb{T}} \|P'_{\mathcal{T}}\|_{\mathcal{L}(L_2(\Omega),L_2(\Omega))} < \infty.$$

For proving boundedness in $H^1_{0,\gamma}(\Omega)$, we consider the Scott-Zhang ([SZ90]) interpolator $\Pi_{\mathcal{T}} : H^1_{0,\gamma}(\Omega) \to \mathscr{V}_{\mathcal{T}}$. From (4.4) and properties of the $\Pi_{\mathcal{T}}$ [SZ90, (3.8) and (4.3)], we deduce that

$$\begin{aligned}\|P'_{\mathcal{T}}u\|_{H^1(T)} &= \|\Pi_{\mathcal{T}}u + P'_{\mathcal{T}}(\mathrm{Id} - \Pi_{\mathcal{T}})u\|_{H^1(T)}\\ &\lesssim \|u\|_{H^1(\omega_{\mathcal{T}}^{(1)}(T))} + h_T^{-1}\|(\mathrm{Id} - \Pi_{\mathcal{T}})u\|_{L_2(\omega_{\mathcal{T}}^{(1)}(T))}\\ &\lesssim \|u\|_{H^1(\omega_{\mathcal{T}}^{(2)}(T))},\end{aligned}$$

and consequently

$$\sup_{\mathcal{T}\in\mathbb{T}} \|P'_{\mathcal{T}}\|_{\mathcal{L}(H^1_{0,\gamma}(\Omega),H^1_{0,\gamma}(\Omega))} < \infty.$$

An application of the Riesz-Thorin interpolation theorem yields the result. □

The basis $\Psi_{\mathcal{T}}$ has the crucial benefit that the matrix representation of $D_{\mathcal{T}}$, i.e.

$$\boldsymbol{D}_{\mathcal{T}} = \langle \Phi_{\mathcal{T}}, \Psi_{\mathcal{T}}\rangle_{L_2(\Omega)},$$

is *diagonal*, and thus easily invertible, cf. Sect. 4.2.1.

Combining the theorem with Proposition 4.2.2 gives the following corollary (without requiring additional assumptions on the family of partitions $\mathbb{T}$).

Corollary 4.3.2. *Suppose we have* $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{W}_{\mathcal{T}}, \mathscr{W}'_{\mathcal{T}})$ *uniformly. With* $D_{\mathcal{T}}\colon \mathscr{V}_{\mathcal{T}} \to \mathscr{W}_{\mathcal{T}}$ *defined by* $(D_{\mathcal{T}}v)(w) := \langle v, w\rangle_{L_2(\Omega)}$, *we find that* $G_{\mathcal{T}} = D_{\mathcal{T}}^{-1}B_{\mathcal{T}}(D'_{\mathcal{T}})^{-1} \in \mathcal{L}\mathrm{is}_c(\mathscr{V}'_{\mathcal{T}}, \mathscr{V}_{\mathcal{T}})$ *is a uniform preconditioner of* $A_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{V}_{\mathcal{T}}, \mathscr{V}'_{\mathcal{T}})$.

Given some $B \in \mathcal{L}\mathrm{is}_c(\mathscr{W}, \mathscr{W}')$, a possible choice for $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{W}_{\mathcal{T}}, \mathscr{W}_{\mathcal{T}}')$ uniformly in $\mathcal{T} \in \mathbb{T}$, is $(B_{\mathcal{T}}u)(v) := (Bu)(v)$ $(u, v \in \mathscr{W}_{\mathcal{T}})$. For $d \in \{2, 3\}$ and $\mathscr{W}' = \mathscr{V} = H^{\frac{1}{2}}(\Omega)$, a suitable B is given by the Weakly Singular Integral operator, whereas for $\mathscr{W}' = \mathscr{V} = H^{\frac{1}{2}}_{00}(\Omega) := [L_2(\Omega), H^1_0(\Omega)]_{\frac{1}{2},2}$, the recently in [HJHUT18] introduced *Modified* Weakly Singular Integral operator can be applied. Similar comments apply to screens.

4.4 Construction of $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{W}_{\mathcal{T}}, \mathscr{W}_{\mathcal{T}}')$

We expect it to be impossible to construct a basis $\Psi_{\mathcal{T}}$ in the (standard) spaces $\mathscr{S}^{-1,0}_{\mathcal{T}}$ or $\mathscr{S}^{0,1}_{\mathcal{T}}$ that is *local* and *biorthogonal* to $\Phi_{\mathcal{T}}$ as required in (4.3). One remedy is to construct $\Psi_{\mathcal{T}}$ in a (finite element) space w.r.t. a refined partition $\mathcal{T}_* \succ \mathcal{T}$. However, this implies that some opposite order operator $B \in \mathcal{L}\mathrm{is}_c(\mathscr{W}, \mathscr{W}')$ has to be discretized on a space w.r.t. the *refined* partition $\mathcal{T}_*$. This increases the cost of the preconditioner, and moreover, increases implementational complexity as one has to actually construct this refined partition.

To circumvent (explicit) dependence on the refined partition $\mathcal{T}_*$, we shall apply the idea described in Sect. 2.3. That is, we will construct an operator $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{W}_{\mathcal{T}}, \mathscr{W}_{\mathcal{T}}')$ by decomposing an enclosing space of space $\mathscr{W}_{\mathcal{T}}$ into a a standard finite element space $\mathscr{U}_{\mathcal{T}}$, either $\mathscr{S}^{-1,0}_{\mathcal{T}}$ (in Sect. 4.4.2) or $\mathscr{S}^{0,1}_{\mathcal{T}}$ (in Sect. 4.4.3), and some bubble space $\mathscr{B}_{\mathcal{T}}$. On $\mathscr{U}_{\mathcal{T}}$ we will apply the Galerkin discretization operator of the opposite order operator B, whereas on the bubble space $\mathscr{B}_{\mathcal{T}}$ a diagonal scaling will suffice.

In the first subsection we present this construction of $B_{\mathcal{T}}$ for some abstract $\mathscr{W}_{\mathcal{T}}$. In the subsequent subsections, we will present two viable options for $\mathscr{W}_{\mathcal{T}}$, leading to two different preconditioners.

4.4.1 Stable decomposition

The role of the space '$\mathscr{Y}$' is the next proposition is going to be played by $\mathscr{W}_{\mathcal{T}}$.

Proposition 4.4.1. *Let $\mathscr{Z}$ be an inner product space, $Q \in \mathcal{L}(\mathscr{Z}, \mathscr{Z})$ a projector, and with $\mathscr{U} := \operatorname{ran} Q$, let $\mathscr{B} := \operatorname{ran}(\mathrm{Id} - Q)$, $B^{\mathscr{U}} \in \mathcal{L}\mathrm{is}_c(\mathscr{U}, \mathscr{U}')$, and $B^{\mathscr{B}} \in \mathcal{L}\mathrm{is}_c(\mathscr{B}, \mathscr{B}')$. Then for any subspace $\mathscr{Y} \subset \mathscr{Z}$,*

$$(By)(\tilde{y}) := (B^{\mathscr{U}} Qy)(Q\tilde{y}) + (B^{\mathscr{B}}(\mathrm{Id} - Q)y)((\mathrm{Id} - Q)\tilde{y}) \quad (y, \tilde{y} \in \mathscr{Y}),$$

is bounded and coercive — $B \in \mathcal{L}\mathrm{is}_c(\mathscr{Y}, \mathscr{Y}')$ — with

$$\begin{aligned}
\|B\|_{\mathcal{L}(\mathscr{Y},\mathscr{Y}')} &\leq \\
&\left(\|Q\|^2 + \sqrt{\|Q\|^4 - \|Q\|^2}\right) \max(\|B^{\mathscr{U}}\|_{\mathcal{L}(\mathscr{U},\mathscr{U}')}, \|B^{\mathscr{B}}\|_{\mathcal{L}(\mathscr{B},\mathscr{B}')}), \\
\|\Re(B)^{-1}\|_{\mathcal{L}(\mathscr{Y}',\mathscr{Y})} &\leq \\
&\left(1 + \sqrt{1 - \|Q\|^{-2}}\right) \max(\|\Re(B^{\mathscr{U}})^{-1}\|_{\mathcal{L}(\mathscr{U}',\mathscr{U})}, \|\Re(B^{\mathscr{B}})^{-1}\|_{\mathcal{L}(\mathscr{B}',\mathscr{B})}),
\end{aligned}$$

where $\|Q\| := \|Q\|_{\mathcal{L}(\mathscr{Z},\mathscr{Z})}$.

Proof. Let $y, \tilde{y} \in \mathscr{Y}$. Write $u = Qy$, $b = (\mathrm{Id} - Q)y$, and similarly $\tilde{u} = Q\tilde{y}$, $\tilde{b} = (\mathrm{Id} - Q)\tilde{y}$. We have

$$\begin{aligned}|(B(y))(\tilde{y})| &\leq \max\left(\|B^{\mathscr{U}}\|_{\mathcal{L}(\mathscr{U},\mathscr{U}')}, \|B^{\mathscr{B}}\|_{\mathcal{L}(\mathscr{B},\mathscr{B}')}\right) \cdot \left(\|u\|_{\mathscr{Z}}\|\tilde{u}\|_{\mathscr{Z}} + \|b\|_{\mathscr{Z}}\|\tilde{b}\|_{\mathscr{Z}}\right) \\ &\leq \max(\cdots)\sqrt{\|u\|_{\mathscr{Z}}^2 + \|b\|_{\mathscr{Z}}^2} \cdot \sqrt{\|\tilde{u}\|_{\mathscr{Z}}^2 + \|\tilde{b}\|_{\mathscr{Z}}^2},\end{aligned}$$

and

$$|(B(y))(y)| \geq \min\left(\|\Re(B^{\mathscr{U}})^{-1}\|^{-1}_{\mathcal{L}(\mathscr{U}',\mathscr{U})}, \|\Re(B^{\mathscr{B}})^{-1}\|^{-1}_{\mathcal{L}(\mathscr{B}',\mathscr{B})}\right) \cdot (\|u\|_{\mathscr{Z}}^2 + \|b\|_{\mathscr{Z}}^2).$$

With $\gamma := \sup_{0\neq(u,b)\in\mathscr{U}\times\mathscr{B}} \frac{|\langle u,b\rangle_{\mathscr{Z}}|}{\|u\|_{\mathscr{Z}}\|b\|_{\mathscr{Z}}}$, for $0 \neq (u,b) \in \mathscr{U}\times\mathscr{B}$ we have $\frac{\|u+b\|_{\mathscr{Z}}^2}{\|u\|_{\mathscr{Z}}^2+\|b\|_{\mathscr{Z}}^2} \in [1-\gamma, 1+\gamma]$. Using that $\sqrt{\frac{1}{1-\gamma^2}} = \|Q\|$ (see e.g. [Szy06, (5.5), (5.7), (6.2)]), the proof is easily completed. □

Remark 4.4.2. For a quantitatively weaker result as Proposition 4.4.1 to hold it is actually sufficient when Q is only defined on $\mathscr{Y}$, and neither is it needed that it is a projector. Under these relaxed conditions, obvious estimates show bounds as in Proposition 4.4.1 with the factors $\|Q\|^2 + \sqrt{\|Q\|^4 - \|Q\|^2}$ and $1 + \sqrt{1 - \|Q\|^{-2}}$ reading as $\|Q|_{\mathscr{Y}}\|^2 + (1 + \|Q|_{\mathscr{Y}}\|)^2$ and 2, respectively. Both original factors are equal to 1 when Q is an orthogonal projector.

We are going to apply this abstract proposition with '$\mathscr{Y}$'= $\mathscr{W}_{\mathcal{T}}$, '$\mathscr{U}$'= $\mathscr{U}_{\mathcal{T}}$ being a standard finite element space, '$\mathscr{B}$'= $\mathscr{B}_{\mathcal{T}}$ being a suitably constructed 'bubble space', and '$\mathscr{Z}$'= $\mathscr{Z}_{\mathcal{T}} := \mathscr{U}_{\mathcal{T}} + \mathscr{B}_{\mathcal{T}}$, all equipped with the norm on $\mathscr{W}$. The resulting 'B' will be the $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{W}_{\mathcal{T}}, \mathscr{W}_{\mathcal{T}}')$ we are seeking.

In order to apply above proposition, what is left is the construction of a (uniformly) bounded projector defined on $\mathscr{Z}_{\mathcal{T}}$. Furthermore, to allow for a simple preconditioner on $\mathscr{B}_{\mathcal{T}}$ we would like to find a setting in which on this bubble space the $\mathscr{W}$-norm is equivalent to a weighted L_2-norm. Both issues will be dealt with in the next two lemmas. The operator $Q_{\mathcal{T}}|_{\mathscr{Z}_{\mathcal{T}}}$ in the first lemma will play the role of 'Q' in Proposition 4.4.1.

Lemma 4.4.3. *Let $Q_{\mathcal{T}} \in \mathcal{L}(L_2(\Omega), H^1_{0,\gamma}(\Omega)')$ be a projector, $\mathscr{U}_{\mathcal{T}} \subseteq \operatorname{ran} Q_{\mathcal{T}}$ and $\mathscr{B}_{\mathcal{T}} \subseteq \operatorname{ran}(\mathrm{Id} - Q_{\mathcal{T}})$ be subspaces of $L_2(\Omega)$, and with $\mathscr{Z}_{\mathcal{T}} := \mathscr{U}_{\mathcal{T}} + \mathscr{B}_{\mathcal{T}}$, let*

(1) $\|h_{\mathcal{T}}^{-1}(\mathrm{Id} - Q'_{\mathcal{T}})\|_{\mathcal{L}(H^1_{0,\gamma}(\Omega),L_2(\Omega))} \lesssim 1$, *(approximation property)*

(2) $\sup_{\mathcal{T}\in\mathbb{T}} \|Q_{\mathcal{T}}|_{\mathscr{Z}_{\mathcal{T}}}\|_{\mathcal{L}((\mathscr{Z}_{\mathcal{T}},\|\cdot\|_{L_2(\Omega)}),L_2(\Omega))} \lesssim 1$, *(boundedness in $L_2(\Omega)$)*

(3) $\|h_{\mathcal{T}} \cdot\|_{L_2(\Omega)} \lesssim \|\cdot\|_{H^1_{0,\gamma}(\Omega)'}$ *on* $\mathscr{Z}_{\mathcal{T}}$. *(inverse inequality)*

Then $Q_{\mathcal{T}}|_{\mathscr{Z}_{\mathcal{T}}} \colon \mathscr{Z}_{\mathcal{T}} \to \mathscr{Z}_{\mathcal{T}}$ is a projector, $\operatorname{ran} Q_{\mathcal{T}}|_{\mathscr{Z}_{\mathcal{T}}} = \mathscr{U}_{\mathcal{T}}$, $\operatorname{ran}(\mathrm{Id} - Q_{\mathcal{T}}|_{\mathscr{Z}_{\mathcal{T}}}) = \mathscr{B}_{\mathcal{T}}$, *and*

(i) $\sup_{\mathcal{T}\in\mathbb{T}} \|Q_{\mathcal{T}}|_{\mathscr{Z}_{\mathcal{T}}}\|_{\mathcal{L}((\mathscr{Z}_{\mathcal{T}},\|\cdot\|_{\mathscr{W}}),\mathscr{W})} < \infty$,

(ii) $\|\cdot\|_{\mathscr{W}} \eqsim \|h_{\mathcal{T}}^s \cdot\|_{L_2(\Omega)}$ *on* $\mathscr{B}_{\mathcal{T}}$.

Proof. The first three statements are easily verified. From (1) it follows that for $u \in H^1_{0,\gamma}(\Omega)'$:

$$\begin{aligned}\|(\mathrm{Id}-Q_{\mathcal{T}})u\|_{H^1_{0,\gamma}(\Omega)'} &= \sup_{v\in H^1_{0,\gamma}(\Omega)} \frac{\langle u, (\mathrm{Id}-Q'_{\mathcal{T}})v\rangle_{L_2(\Omega)}}{\|v\|_{H^1_{0,\gamma}(\Omega)}}\\ &\le \sup_{v\in H^1_{0,\gamma}(\Omega)} \frac{\|h_{\mathcal{T}}u\|_{L_2(\Omega)}\|h_{\mathcal{T}}^{-1}(\mathrm{Id}-Q'_{\mathcal{T}})v\|_{L_2(\Omega)}}{\|v\|_{H^1_{0,\gamma}(\Omega)}}\\ &\lesssim \|h_{\mathcal{T}}u\|_{L_2(\Omega)}.\end{aligned}$$

Together with the inverse inequality on $\mathscr{Z}_{\mathcal{T}}$, this gives boundedness of $\|(\mathrm{Id}-Q_{\mathcal{T}})|_{\mathscr{Z}_{\mathcal{T}}}\|_{\mathcal{L}((\mathscr{Z}_{\mathcal{T}},\|\cdot\|_{H^1_{0,\gamma}(\Omega)'}),H^1_{0,\gamma}(\Omega)')}$ and thus of $\|Q_{\mathcal{T}}|_{\mathscr{Z}_{\mathcal{T}}}\|_{\mathcal{L}((\mathscr{Z}_{\mathcal{T}},\|\cdot\|_{H^1_{0,\gamma}(\Omega)'}),H^1_{0,\gamma}(\Omega)')}$. The first result then follows from (2) and an interpolation argument.

By the inverse inequality on $\mathscr{B}_{\mathcal{T}}$ and the previously derived inequality, we have for $b_{\mathcal{T}} \in \mathscr{B}_{\mathcal{T}} \subseteq \mathrm{ran}(\mathrm{Id}-Q_{\mathcal{T}})$ that

$$\|b_{\mathcal{T}}\|_{H^1_{0,\gamma}(\Omega)'} = \|(\mathrm{Id}-Q_{\mathcal{T}})b_{\mathcal{T}}\|_{H^1_{0,\gamma}(\Omega)'} \lesssim \|h_{\mathcal{T}}b_{\mathcal{T}}\|_{L_2(\Omega)} \lesssim \|b_{\mathcal{T}}\|_{H^1_{0,\gamma}(\Omega)'}.$$

Another interpolation argument yields the second result. □

Lemma 4.4.4. *Suppose that* $\|\cdot\|_{\mathscr{W}} \eqsim \|h_{\mathcal{T}}^s \cdot\|_{L_2(\Omega)}$ *holds on* $\mathscr{B}_{\mathcal{T}}$*, and that* $\Theta_{\mathcal{T}}$ *is a* uniformly $\|h_{\mathcal{T}}^s \cdot\|_{L_2(\Omega)}$-stable basis for $\mathscr{B}_{\mathcal{T}}$*, i.e.*

$$\mathscr{B}_{\mathcal{T}} = \operatorname{span}\Theta_{\mathcal{T}} \quad \textit{and} \quad \big\|h_{\mathcal{T}}^s \sum\nolimits_{\theta\in\Theta_{\mathcal{T}}} c_\theta\theta\big\|^2_{L_2(\Omega)} \eqsim \sum\nolimits_{\theta\in\Theta_{\mathcal{T}}} |c_\theta|^2 \|h_{\mathcal{T}}^s\theta\|^2_{L_2(\Omega)},$$

then, for any $\beta_1 > 0$*, an operator* $B^{\mathscr{B}}_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{B}_{\mathcal{T}}, \mathscr{B}'_{\mathcal{T}})$ *is given by*

$$\big(B^{\mathscr{B}}_{\mathcal{T}} \sum_{\theta\in\Theta_{\mathcal{T}}} c_\theta\theta\big)\big(\sum_{\theta\in\Theta_{\mathcal{T}}} d_\theta\theta\big) = \beta_1 \sum_{\theta\in\Theta_{\mathcal{T}}} c_\theta d_\theta \|h_{\mathcal{T}}^s\theta\|^2_{L_2(\Omega)}. \tag{4.5}$$

Remark 4.4.5. It is not possible to construct $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}(\mathscr{W}_{\mathcal{T}}, \mathscr{W}'_{\mathcal{T}})$ directly as a diagonal scaling operator. Indeed, this would require $\|w_{\mathcal{T}}\|_{\mathscr{W}} \lesssim \|h_{\mathcal{T}}^s w_{\mathcal{T}}\|_{L_2(\Omega)}$ for $w_{\mathcal{T}} \in \mathscr{W}_{\mathcal{T}}$. Suppose this to be true, then by $L_2(\Omega)$-boundedness of the biorthogonal projector $P_{\mathcal{T}}$, we would find for $v_{\mathcal{T}} \in \mathscr{V}_{\mathcal{T}}$ that

$$\begin{aligned}\|h_{\mathcal{T}}^{-s}v_{\mathcal{T}}\|_{L_2(\Omega)} &= \sup_{w\in L_2(\Omega)} \frac{\langle h_{\mathcal{T}}^{-s}v_{\mathcal{T}}, P_{\mathcal{T}}w\rangle_{L_2(\Omega)}}{\|w\|_{L_2(\Omega)}} \lesssim \sup_{w\in L_2(\Omega)} \frac{\langle h_{\mathcal{T}}^{-s}v_{\mathcal{T}}, P_{\mathcal{T}}w\rangle_{L_2(\Omega)}}{\|P_{\mathcal{T}}w\|_{L_2(\Omega)}}\\ &= \sup_{w_{\mathcal{T}}\in\mathscr{W}_{\mathcal{T}}} \frac{\langle v_{\mathcal{T}}, w_{\mathcal{T}}\rangle_{L_2(\Omega)}}{\|h_{\mathcal{T}}^s w_{\mathcal{T}}\|_{L_2(\Omega)}} \lesssim \sup_{w_{\mathcal{T}}\in\mathscr{W}_{\mathcal{T}}} \frac{\langle v_{\mathcal{T}}, w_{\mathcal{T}}\rangle_{L_2(\Omega)}}{\|w_{\mathcal{T}}\|_{\mathscr{W}}} \le \|v_{\mathcal{T}}\|_{\mathscr{V}},\end{aligned}$$

which is known not to be true for smooth functions in $\mathscr{V}_{\mathcal{T}}$.

Concluding: If, given a family of subspaces $\mathscr{W}_\mathcal{T} \subset L_2(\Omega)$, one can find a family of projectors $Q_\mathcal{T} \in \mathcal{L}(L_2(\Omega), H^1_{0,\gamma}(\Omega)')$, subspaces $\mathscr{U}_\mathcal{T} \subseteq \operatorname{ran} Q_\mathcal{T}$ (of finite element type) and $\mathscr{B}_\mathcal{T} \subseteq \operatorname{ran}(\mathrm{Id} - Q_\mathcal{T})$ such that

$$\mathscr{W}_\mathcal{T} \subset \mathscr{Z}_\mathcal{T} := \mathscr{U}_\mathcal{T} + \mathscr{B}_\mathcal{T} \tag{4.6}$$

(with these spaces equipped with $\|\cdot\|_\mathscr{W}$-norm) and the conditions of Lemma 4.4.3 are satisfied, then given $B^\mathscr{U}_\mathcal{T} \in \mathcal{L}\mathrm{is}_c(\mathscr{U}_\mathcal{T}, \mathscr{U}'_\mathcal{T})$ and $B^\mathscr{B}_\mathcal{T} \in \mathcal{L}\mathrm{is}_c(\mathscr{B}_\mathcal{T}, \mathscr{B}'_\mathcal{T})$, the operator $B^\mathscr{W}_\mathcal{T}$ defined by

$$(B_\mathcal{T} w)(\tilde{w}) := (B^\mathscr{U}_\mathcal{T} Q_\mathcal{T} w)(Q_\mathcal{T}\tilde{w}) + (B^\mathscr{B}_\mathcal{T}(\mathrm{Id}-Q_\mathcal{T})w)((\mathrm{Id}-Q_\mathcal{T})\tilde{w}) \quad (w, \tilde{w} \in \mathscr{W}_\mathcal{T}), \tag{4.7}$$

is in $\mathcal{L}\mathrm{is}_c(\mathscr{W}_\mathcal{T}, \mathscr{W}'_\mathcal{T})$. Moreover, assuming a uniformly $\|h^s_\mathcal{T} \cdot \|_{L_2(\Omega)}$-stable basis for $\mathscr{B}_\mathcal{T}$, the operator $B^\mathscr{B}_\mathcal{T}$ can be of simple diagonal scaling type, where a natural definition for $B^\mathscr{U}_\mathcal{T}$ is by $(B_\mathcal{T} u)(\tilde{u}) := (Bu)(\tilde{u})$ $(u, \tilde{u} \in \mathscr{U}_\mathcal{T})$ for some opposite order operator $B \in \mathcal{L}\mathrm{is}_c(\mathscr{W}, \mathscr{W}')$. Finally, since $Q_\mathcal{T}$ enters the implementation, we search this projector to be of local type.

4.4.2 A space $\mathscr{W}_\mathcal{T}$ enclosed in a space decomposable into the piecewise constants and bubbles

In this subsection, we construct $\mathscr{W}_\mathcal{T} = \operatorname{span} \Psi_\mathcal{T}$ such that both $\Psi_\mathcal{T}$ is biorthogonal to $\Phi_\mathcal{T}$ ((4.3)), and $\mathscr{W}_\mathcal{T}$ is enclosed in a space that allows an appropriate decomposition into the space of piecewise constants $\mathscr{U}_\mathcal{T} := \mathscr{S}^{-1,0}_\mathcal{T}$ and a bubble space $\mathscr{B}_\mathcal{T}$.

Fix $\mathcal{T} \in \mathbb{T}$ and let $\mathcal{T}_* \succ \mathcal{T}$ be a uniform red-refinement, i.e. every simplex $T \in \mathcal{T}$ is subdivided into 2^d subsimplices.[2] We define $\Psi_\mathcal{T} = \{\psi_{\mathcal{T},\nu} \colon \nu \in N^0_\mathcal{T}\} \subset \mathscr{S}^{-1,0}_{\mathcal{T}_*}$ by taking a weighted difference of 'patch indicator' functions:

$$\psi_{\mathcal{T},\nu} := 2^{d+1} \mathbb{1}_{\omega_{\mathcal{T}_*,\nu}} - \mathbb{1}_{\omega_{\mathcal{T},\nu}} \quad (\nu \in N^0_\mathcal{T}). \tag{4.8}$$

Lemma 4.4.6. *The collection $\Psi_\mathcal{T}$ satisfies* (4.3) *with* $\operatorname{supp} \psi_{\mathcal{T},\nu} = \omega_{\mathcal{T},\nu}$ *and*

$$\langle \psi_{\mathcal{T},\nu}, \phi_{\mathcal{T},\nu'} \rangle_{L_2(\Omega)} = \delta_{\nu\nu'} |\omega_{\mathcal{T},\nu}| \quad (\nu, \nu' \in N^0_\mathcal{T}). \tag{4.9}$$

Proof. Clearly $\operatorname{supp} \psi_{\mathcal{T},\nu} = \omega_{\mathcal{T},\nu}$, so we are left to show the biorthogonality condition. Fix some vertex $\nu \in N^0_\mathcal{T}$. For a simplex $T_\nu \in \mathcal{T}$ with $\nu \in T_\nu$, we have

$$\langle \mathbb{1}_{T_\nu}, \phi_{\mathcal{T},\nu} \rangle_{L_2(\Omega)} = \frac{|T_\nu|}{d+1}.$$

[2] Red-refinement is not uniquely defined for $d \geq 3$, but the refined simplices at the corners of the 'parent simplex' are uniquely determined which suffices for our goal.

Let $T_{*,\nu} \in \mathcal{T}_*$ be the (unique) simplex with $\nu \in T_{*,\nu} \subset T_\nu$. From the refinement equation satisfied by the nodal hats, and $|T_{*,\nu}| = 2^{-d}|T_\nu|$, it follows that

$$\langle \mathbb{1}_{T_{*,\nu}}, \phi_{\mathcal{T},\nu}\rangle_{L_2(\Omega)} = \langle \mathbb{1}_{T_{*,\nu}}, \phi_{\mathcal{T}_*,\nu} + \sum_{\nu\neq\tilde\nu\in N_{T_{*,\nu}}} 2^{-1}\phi_{\mathcal{T}_*,\tilde\nu}\rangle_{L_2(\Omega)} = \frac{2^{-d}|T_\nu|}{d+1}(1+2^{-1}d),$$

$$\langle \mathbb{1}_{T_{*,\nu}}, \phi_{\mathcal{T},\nu'}\rangle_{L_2(\Omega)} = \cdots = \frac{2^{-d}|T_\nu|}{d+1}2^{-1} \quad (\nu\neq\nu'\in N^0_{T_\nu}).$$

From these relations (4.9) follows. □

By Lemma 4.4.6 it has been established that the Fortin interpolator is uniformly bounded, and that $D_{\mathcal{T}}$ is represented by a diagonal matrix. The next proposition verifies the conditions imposed in Sect. 4.4.1 for the construction of $B_{\mathcal{T}}$.

Proposition 4.4.7. *Let $\mathscr{U}_{\mathcal{T}} := \mathscr{S}^{-1,0}_{\mathcal{T}}$, $\mathscr{W}_{\mathcal{T}} := \operatorname{span}\Psi_{\mathcal{T}}$ as constructed above, $Q_{\mathcal{T}}$ be the $L_2(\Omega)$-orthogonal projector onto $\mathscr{U}_{\mathcal{T}}$, $\Theta_{\mathcal{T}} := (\mathrm{Id}-Q_{\mathcal{T}})\Psi_{\mathcal{T}}$, and $\mathscr{B}_{\mathcal{T}} := \operatorname{span}\Theta_{\mathcal{T}}$. Then $\mathscr{W}_{\mathcal{T}} \subset \mathscr{Z}_{\mathcal{T}} := \mathscr{U}_{\mathcal{T}} + \mathscr{B}_{\mathcal{T}}$ ((4.6)), the conditions of Lemma 4.4.3 are satisfied, in particular $Q_{\mathcal{T}}\psi_\nu = \mathbb{1}_{\omega_\nu}$, and $\Theta_{\mathcal{T}}$ is a uniformly $\|h^s_{\mathcal{T}}\cdot\|_{L_2(\Omega)}$-stable basis for $\mathscr{B}_{\mathcal{T}}$ as required for Lemma 4.4.4.*

Proof. The first statement follows from $\mathscr{W}_{\mathcal{T}} \subset L_2(\Omega)$. The first two conditions of Lemma 4.4.3 are obviously valid. Concerning the third condition, the inverse inequality $\|h_{\mathcal{T}}\cdot\|_{L_2(\Omega)} \lesssim \|\cdot\|_{H^1_{0,\gamma}(\Omega)'}$ holds, for general d, on $\mathscr{S}^{-1,0}_{\mathcal{T}_*}$, see e.g. Lemma 2.3.4, and thus in particular on $\mathscr{Z}_{\mathcal{T}}$. The property $Q_{\mathcal{T}}\psi_\nu = \mathbb{1}_{\omega_\nu}$ is easily checked.

We are left to show that the collection of bubbles $\{\theta_\nu := (\mathrm{Id}-Q_{\mathcal{T}})\psi_\nu : \nu \in N^0_{\mathcal{T}}\}$ is $\|h^s_{\mathcal{T}}\cdot\|_{L_2(\Omega)}$-stable. Pick some $T\in\mathcal{T}$, then the normalized 'bubble element matrix' satisfies

$$\begin{aligned}\tfrac14|T|^{-1}\langle\theta_\nu,\theta_{\nu'}\rangle_{L_2(T)} &= |T|^{-1}\langle 2^d\mathbb{1}_{\omega_{\mathcal{T}_*,\nu}} - \mathbb{1}_{\omega_{\mathcal{T},\nu}}, 2^d\mathbb{1}_{\omega_{\mathcal{T}_*,\nu'}} - \mathbb{1}_{\omega_{\mathcal{T},\nu'}}\rangle_{L_2(T)}\\ &= \begin{cases} 2^d-1 & \nu=\nu'\in N^0_T,\\ -1 & \nu\neq\nu'\in N^0_T.\end{cases}\end{aligned} \tag{4.10}$$

For $d\geq 2$, this constant (symmetric) $(d+1)\times(d+1)$ matrix is strictly diagonally dominant, and therefore positive definite. We conclude this proposition by

$$\Big\|\sum_{\nu\in N^0_{\mathcal{T}}} h^s_{\mathcal{T}} c_\nu\theta_\nu\Big\|^2_{L_2(\Omega)} = \sum_{T\in\mathcal{T}} h_T^{2s}\Big\|\sum_{\nu\in N^0_T} c_\nu\theta_\nu\Big\|^2_{L_2(T)} \eqsim \sum_{T\in\mathcal{T}} h_T^{2s}\sum_{\nu\in N^0_T}|c_\nu|^2\|\theta_\nu\|^2_{L_2(T)}$$
$$= \sum_{\nu\in N^0_{\mathcal{T}}}|c_\nu|^2\sum_{T\in\mathcal{T}}\|h^s_T\theta_\nu\|^2_{L_2(T)} = \sum_{\nu\in N^0_{\mathcal{T}}}|c_\nu|^2\|h^s_{\mathcal{T}}\theta_\nu\|^2_{L_2(\Omega)}.$$ □

Remark 4.4.8. For $d = 1$, the bubbles arising from $\Psi_{\mathcal{T}}$ as given in (4.8) do not form a $\|h_{\mathcal{T}}^s \cdot\|_{L_2(\Omega)}$-stable collection. Instead, with $\mathcal{T}_{**} \succ \mathcal{T}$ being the two times uniform red-refinement, one can consider $\psi_{\mathcal{T},\nu} = \frac{16}{3}\mathbb{1}_{\omega_{\mathcal{T}_{**},\nu}} - \frac{1}{3}\mathbb{1}_{\omega_{\mathcal{T},\nu}}$ for which the statements of Lemma 4.4.6 and Proposition 4.4.7 are again valid.

Implementation

The matrix representation of preconditioner $\mathcal{F}_{\Phi_{\mathcal{T}}}^{-1} G_{\mathcal{T}} (\mathcal{F}'_{\Phi_{\mathcal{T}}})^{-1}$ is given by

$$\boldsymbol{G}_{\mathcal{T}} = \boldsymbol{D}_{\mathcal{T}}^{-1} \boldsymbol{B}_{\mathcal{T}} \boldsymbol{D}_{\mathcal{T}}^{-\top}.$$

With $\Psi_{\mathcal{T}}$ as constructed in (4.8), we find that $\boldsymbol{D}_{\mathcal{T}} = \mathcal{F}'_{\Psi_{\mathcal{T}}} D_{\mathcal{T}} \mathcal{F}_{\Phi_{\mathcal{T}}}$ is given by

$$\boldsymbol{D}_{\mathcal{T}} = \operatorname{diag}\{|\omega_\nu| \colon \nu \in N_{\mathcal{T}}^0\}.$$

Given some $B_{\mathcal{T}}^{\mathscr{U}} \in \mathcal{L}\mathrm{is}_c(\mathscr{U}_{\mathcal{T}}, \mathscr{U}'_{\mathcal{T}})$ (recall that $\mathscr{U}_{\mathcal{T}} = \mathscr{S}_{\mathcal{T}}^{-1,0}$), then by taking $B_{\mathcal{T}}$ as described in (4.7), we have

$$\begin{aligned}
\boldsymbol{B}_{\mathcal{T}} &:= \mathcal{F}'_{\Psi_{\mathcal{T}}} B_{\mathcal{T}} \mathcal{F}_{\Psi_{\mathcal{T}}} \\
&= \mathcal{F}'_{\Psi_{\mathcal{T}}} (Q'_{\mathcal{T}} B_{\mathcal{T}}^{\mathscr{U}} Q_{\mathcal{T}} + (\mathrm{Id} - Q_{\mathcal{T}})' B_{\mathcal{T}}^{\mathscr{B}} (\mathrm{Id} - Q_{\mathcal{T}})) \mathcal{F}_{\Psi_{\mathcal{T}}} \\
&= \boldsymbol{p}_{\mathcal{T}}^\top \boldsymbol{B}_{\mathcal{T}}^{\mathscr{U}} \boldsymbol{p}_{\mathcal{T}} + \boldsymbol{B}_{\mathcal{T}}^{\mathscr{B}},
\end{aligned}$$

where, using that $\mathcal{F}_{\Theta_{\mathcal{T}}}^{-1}(\mathrm{Id} - Q_{\mathcal{T}})\mathcal{F}_{\Psi_{\mathcal{T}}} = \mathrm{Id}$ by $\Theta_{\mathcal{T}} = (I - Q_{\mathcal{T}})\Psi_{\mathcal{T}}$,

$$\boldsymbol{B}_{\mathcal{T}}^{\mathscr{U}} := \mathcal{F}'_{\Sigma_{\mathcal{T}}} B_{\mathcal{T}}^{\mathscr{U}} \mathcal{F}_{\Sigma_{\mathcal{T}}}, \quad \boldsymbol{p}_{\mathcal{T}} := \mathcal{F}_{\Sigma_{\mathcal{T}}}^{-1} Q_{\mathcal{T}} \mathcal{F}_{\Psi_{\mathcal{T}}}, \quad \boldsymbol{B}_{\mathcal{T}}^{\mathscr{B}} := \mathcal{F}'_{\Theta_{\mathcal{T}}} B_{\mathcal{T}}^{\mathscr{B}} \mathcal{F}_{\Theta_{\mathcal{T}}},$$

Recall the canonical basis $\Sigma_{\mathcal{T}}$ for $\mathscr{U}_{\mathcal{T}}$ from (4.2). Using $Q_{\mathcal{T}}\psi_\nu = \mathbb{1}_{\omega_\nu}$ shows that

$$(\boldsymbol{p}_{\mathcal{T}})_{T\nu} = \begin{cases} 1 & \text{if } T \subset \omega_\nu, \\ 0 & \text{else.} \end{cases}$$

From (4.10), we infer that $\|h_{\mathcal{T}}^s \theta_\nu\|_{L_2(\Omega)}^2 \eqsim |\omega_\nu|^{1+\frac{2s}{d}}$. By making a harmless modification to the definition of $B_{\mathcal{T}}^{\mathscr{B}}$ in (4.5) based on this equivalency, we obtain that

$$\boldsymbol{B}_{\mathcal{T}}^{\mathscr{B}} = \beta_1 \boldsymbol{D}_{\mathcal{T}}^{1+\frac{2s}{d}}.$$

The matrix $\boldsymbol{B}_{\mathcal{T}}^{\mathscr{U}}$ depends on the operator $B_{\mathcal{T}}^{\mathscr{U}} \in \mathcal{L}\mathrm{is}_c(\mathscr{U}_{\mathcal{T}}, \mathscr{U}'_{\mathcal{T}})$ that is selected. The canonical choice is the Galerkin discretization operator on $\mathscr{U}_{\mathcal{T}}$ of a $B \in \mathcal{L}\mathrm{is}_c(\mathscr{W}, \mathscr{W}')$. The cost of the application of $\boldsymbol{G}_{\mathcal{T}}$ is the cost of the application of $\boldsymbol{B}_{\mathcal{T}}^{\mathscr{U}}$ plus cost that scales linearly in $\#\mathcal{T}$.

4.4.3 A space $\mathscr{W}_{\mathcal{T}}$ that is enclosed in a space decomposable into the continuous piecewise linears and bubbles

We follow the same program as in the previous subsection Sect. 4.4.2 but now with $\mathscr{U}_{\mathcal{T}} := \mathscr{S}_{\mathcal{T}}^{0,1}$, being the space of continuous piecewise linears.

Other than in Sect. 4.4.2, we cannot apply Proposition 4.4.1 for $Q_{\mathcal{T}}$ being the orthogonal projector onto $\mathscr{U}_{\mathcal{T}}$, since with the current choice of this space it will not be a local projector. As an alternative, we take $Q_{\mathcal{T}}$ to be some biorthogonal projector. The question whether it enjoys an approximation property is answered in the following lemma.

Lemma 4.4.9. *For $\nu \in N_{\mathcal{T}}$, so* including *vertices on γ, let $\widetilde{\phi}_\nu \in L_2(\Omega)$ be such that*

$$\|\widetilde{\phi}_\nu\|_{L_2(\Omega)} \lesssim h_\nu^{d/2}, \quad \sum\nolimits_{\nu\in N_{\mathcal{T}}} \widetilde{\phi}_\nu = \mathbb{1}_\Omega, \quad \operatorname{supp}\widetilde{\phi}_\nu \subset B(\nu; Rh_\nu) \tag{4.11}$$

for some constant $R > 0$, and

$$\big|\langle \widetilde{\phi}_\nu, \phi_{\nu'}\rangle_{L_2(\Omega)}\big| \eqsim \delta_{\nu\nu'}|\omega_\nu| \quad (\nu, \nu' \in N_{\mathcal{T}}^0).$$

Denote $\widetilde{\mathscr{U}_{\mathcal{T}}} := \operatorname{span}\{\widetilde{\phi}_\nu \colon \nu \in N_{\mathcal{T}}^0\}$, so without *vertices on γ. The biorthogonal projector $Q_{\mathcal{T}} \colon u \mapsto \sum_{\nu\in N_{\mathcal{T}}^0} \frac{\langle u, \widetilde{\phi}_\nu\rangle_{L_2(\Omega)}}{\langle \phi_\nu, \widetilde{\phi}_\nu\rangle_{L_2(\Omega)}}\phi_\nu$, for which $\operatorname{ran} Q_{\mathcal{T}} = \mathscr{S}_{\mathcal{T}}^{0,1}$ and $\operatorname{ran}(\mathrm{Id} - Q_{\mathcal{T}}) = \widetilde{\mathscr{U}_{\mathcal{T}}}^{\perp_{L_2(\Omega)}}$, satisfies the approximation property*

$$\|h_{\mathcal{T}}^{-1}(\mathrm{Id} - Q'_{\mathcal{T}})v\|_{\mathcal{L}(H^1_{0,\gamma}(\Omega), L_2(\Omega))} \lesssim 1,$$

and $\|Q_{\mathcal{T}}\|_{\mathcal{L}(L_2(\Omega), L_2(\Omega))} \lesssim 1$.

Proof. We use the same strategy as in Chapter 2. That is, we define a Scott-Zhang type quasi-interpolator $\Pi_{\mathcal{T}} \colon H^1(\Omega) \to L_2(\Omega)$, cf. [SZ90]. For every $\nu \in N_{\mathcal{T}}$, select a $(d-1)$-face e_ν of some $T \in \mathcal{T}$ with $\nu \in e_\nu$ and $e_\nu \subset \gamma$ if $\nu \in \gamma$. We define $\Pi_{\mathcal{T}}$ by

$$\Pi_{\mathcal{T}} u := \sum_{\nu\in N_{\mathcal{T}}} g_{\mathcal{T},\nu}(u)\widetilde{\phi}_{\mathcal{T},\nu}, \quad g_{\mathcal{T},\nu}(u) := \fint_{e_\nu} u \,\mathrm{d}s.$$

Since $g_{\mathcal{T},\nu}(\mathbb{1}) = 1$, using the properties from (4.11) one can show, cf. proof of Theorem 2.5.1 for details, that

$$\|h_{\mathcal{T}}^{-1}(\mathrm{Id} - \Pi_{\mathcal{T}})(u)\|_{L_2(\Omega)} \lesssim \|u\|_{H^1(\Omega)} \quad (u \in H^1(\Omega)).$$

By construction, $g_{\mathcal{T},\nu}(u) = 0$ for ν on γ and $u \in H^1_{0,\gamma}(\Omega)$, and therefore $\operatorname{ran}\Pi_{\mathcal{T}}|_{H^1_{0,\gamma}(\Omega)} \subset \widetilde{\mathscr{U}_{\mathcal{T}}}$. Finally, combined with $L_2(\Omega)$-boundedness and locality of $Q'_{\mathcal{T}}$, and the fact that $Q'_{\mathcal{T}}$ reproduces $\widetilde{\mathscr{U}_{\mathcal{T}}}$, we find that

$$\begin{aligned}\|h_{\mathcal{T}}^{-1}(\mathrm{Id} - Q'_{\mathcal{T}})v\|_{L_2(\Omega)} &= \inf_{w_{\mathcal{T}}\in\widetilde{\mathscr{U}_{\mathcal{T}}}} \|h_{\mathcal{T}}^{-1}(\mathrm{Id} - Q'_{\mathcal{T}})(v - w_{\mathcal{T}})\|_{L_2(\Omega)} \\ &\lesssim \|h_{\mathcal{T}}^{-1}(\mathrm{Id} - \Pi_{\mathcal{T}})(v)\|_{L_2(\Omega)} \lesssim \|v\|_{H^1_{0,\gamma}(\Omega)} \ \ (v \in H^1_{0,\gamma}(\Omega)).\end{aligned}$$

The last statement can be proven similarly as in the proof of Theorem 4.3.1. □

As before, let $\mathcal{T}_* \succ \mathcal{T}$ denote a uniform red-refinement of $\mathcal{T}$, and for any $T \in \mathcal{T}$ and $\nu \in N_T$, let $T_{*,\nu} \in \mathcal{T}_*$ denote the simplex with $\nu \in T_{*,\nu} \subset T$. For $\nu \in N_{\mathcal{T}}$, so including boundary vertices, define

$$\widetilde{\phi}_{\mathcal{T},\nu} := \tfrac{1}{d+1} \sum_{\substack{T\in\mathcal{T}\\ T\subset\omega_\nu}} \big(\mathbb{1}_T + \tfrac{d2^{1+d}}{d+1}\mathbb{1}_{T_{*,\nu}} - \tfrac{2^{1+d}}{d+1} \sum_{\substack{\nu'\in N_T\\ \nu'\neq\nu}} \mathbb{1}_{T_{*,\nu'}}\big) \in \mathscr{S}^{-1,0}_{\mathcal{T}_*}.$$

These functions satisfy (4.11), and

$$\langle \widetilde{\phi}_{\mathcal{T},\nu}, \phi_{\mathcal{T},\nu'}\rangle_{L_2(\Omega)} = \delta_{\nu\nu'}(d+1)^{-1}|\omega_{\mathcal{T},\nu}|,$$

and so determine a valid biorthogonal projector $Q_{\mathcal{T}}$ via Lemma 4.4.9.

For $\mathcal{T}_{**} \succ \mathcal{T}_*$ a uniform red-refinement of $\mathcal{T}_*$, we define $\Theta_{\mathcal{T}} := \{\theta_{\mathcal{T},\nu} \colon \nu \in N^0_{\mathcal{T}}\}$ by

$$\theta_{\mathcal{T},\nu} := \tfrac{2^{d+2}}{d+2}\big(2^d\mathbb{1}_{\omega_{\mathcal{T}_{**},\nu}} - \mathbb{1}_{\omega_{\mathcal{T}_*,\nu}}\big).$$

Since red-refinement subdivides each simplex into d subsimplices, one infers that

$$\mathscr{B}_{\mathcal{T}} := \operatorname{span}\Theta_{\mathcal{T}} \perp_{L_2(\Omega)} \mathscr{S}^{-1,0}_{\mathcal{T}_*}, \tag{4.12}$$

so that in particular $\mathscr{B}_{\mathcal{T}} \subset \ker Q_{\mathcal{T}}$.

Defining $\Psi_{\mathcal{T}} := \{\psi_{\mathcal{T},\nu} \colon \nu \in N^0_{\mathcal{T}}\}$ by

$$\psi_{\mathcal{T},\nu} := \phi_{\mathcal{T},\nu} + \theta_{\mathcal{T},\nu},$$

calculations as in the proof of Lemma 4.4.6 show the following result.

Lemma 4.4.10. *The collection $\Psi_{\mathcal{T}}$ satisfies* (4.3) *with* $\operatorname{supp}\psi_{\mathcal{T},\nu} = \omega_{\mathcal{T},\nu}$ *and*

$$\langle \psi_{\mathcal{T},\nu}, \phi_{\mathcal{T},\nu'}\rangle_{L_2(\Omega)} = \delta_{\nu\nu'}(d+1)^{-1}|\omega_{\mathcal{T},\nu}| \quad (\nu,\nu' \in N^0_{\mathcal{T}}).$$

So the Fortin interpolator is uniformly bounded, and $D_{\mathcal{T}}$ is represented by a diagonal matrix. Next we verify the conditions imposed in Sect. 4.4.1 for the construction of $B_{\mathcal{T}}$.

Proposition 4.4.11. *Let $\mathscr{U}_{\mathcal{T}}$, $Q_{\mathcal{T}}$, $\mathscr{B}_{\mathcal{T}}$, and $\mathscr{W}_{\mathcal{T}} := \operatorname{span}\Psi_{\mathcal{T}}$ be defined as above. Then $\mathscr{W}_{\mathcal{T}} \subset \mathscr{Z}_{\mathcal{T}} := \mathscr{U}_{\mathcal{T}} + \mathscr{B}_{\mathcal{T}}$* ((4.6)), *the conditions of Lemma 4.4.3 are satisfied, in particular $\Phi_{\mathcal{T}} = Q_{\mathcal{T}}\Psi_{\mathcal{T}}$ and so $\Theta_{\mathcal{T}} = (\mathrm{Id} - Q_{\mathcal{T}})\Psi_{\mathcal{T}}$, and lastly, $\Theta_{\mathcal{T}}$ is an $\|h_{\mathcal{T}}^s \cdot\|_{L_2(\Omega)}$-orthogonal basis for $\mathscr{B}_{\mathcal{T}}$ as required for Lemma 4.4.4.*

Proof. The first statement is obviously true. We have already verified the first two conditions of Lemma 4.4.3. The third condition follows from this inverse inequality on $\mathscr{S}^{-1,1}_{\mathcal{T}_{**}}$ (see e.g. (2.44)), and $\Phi_{\mathcal{T}} = Q_{\mathcal{T}}\Psi_{\mathcal{T}}$ is a consequence of (4.12). The last statement follows from $|\operatorname{supp}\theta_\nu \cap \operatorname{supp}\theta_{\nu'}| = 0$ when $\nu \neq \nu'$. □

Implementation

Suppose that we have some operator $B_{\mathcal{T}}^{\mathscr{U}} \in \mathcal{L}\mathrm{is}_c(\mathscr{U}_{\mathcal{T}}, \mathscr{U}_{\mathcal{T}}')$ uniformly (here $\mathscr{U}_{\mathcal{T}} = \mathscr{S}_{\mathcal{T}}^{0,1}$). The matrix representation of the preconditioner $G_{\mathcal{T}}$, with $B_{\mathcal{T}}$ from (4.7) and the bases from Proposition 4.4.11, becomes

$$\begin{aligned}\boldsymbol{G}_{\mathcal{T}} &= \boldsymbol{D}_{\mathcal{T}}^{-1}\boldsymbol{B}_{\mathcal{T}}\boldsymbol{D}_{\mathcal{T}}^{-\top},\\ \boldsymbol{B}_{\mathcal{T}} &:= \mathcal{F}_{\Psi_{\mathcal{T}}}'(Q_{\mathcal{T}}'B_{\mathcal{T}}^{\mathscr{U}}Q_{\mathcal{T}} + (\mathrm{Id}-Q_{\mathcal{T}})'B_{\mathcal{T}}^{\mathscr{B}}(\mathrm{Id}-Q_{\mathcal{T}}))\mathcal{F}_{\Psi_{\mathcal{T}}}\\ &= \boldsymbol{B}_{\mathcal{T}}^{\mathscr{U}} + \boldsymbol{B}_{\mathcal{T}}^{\mathscr{B}},\end{aligned}$$

with these matrices given by

$$\begin{aligned}\boldsymbol{D}_{\mathcal{T}} &:= \mathcal{F}_{\Psi_{\mathcal{T}}}'D_{\mathcal{T}}\mathcal{F}_{\Phi_{\mathcal{T}}} = \mathrm{diag}\left\{\tfrac{|\omega_\nu|}{d+1} : \nu \in N_{\mathcal{T}}^0\right\},\\ \boldsymbol{B}_{\mathcal{T}}^{\mathscr{U}} &:= \mathcal{F}_{\Phi_{\mathcal{T}}}'B_{\mathcal{T}}^{\mathscr{U}}\mathcal{F}_{\Phi_{\mathcal{T}}}, \quad \boldsymbol{B}_{\mathcal{T}}^{\mathscr{B}} := \mathcal{F}_{\Theta_{\mathcal{T}}}'B_{\mathcal{T}}^{\mathscr{B}}\mathcal{F}_{\Theta_{\mathcal{T}}} = \beta_1\boldsymbol{D}_{\mathcal{T}}^{1+\frac{2s}{d}},\end{aligned}$$

where we used that $\mathcal{F}_{\Phi_{\mathcal{T}}}^{-1}Q_{\mathcal{T}}\mathcal{F}_{\Psi_{\mathcal{T}}} = \mathbf{Id}$ and $\mathcal{F}_{\Theta_{\mathcal{T}}}^{-1}(\mathrm{Id}-Q_{\mathcal{T}})\mathcal{F}_{\Psi_{\mathcal{T}}} = \mathbf{Id}$, and where, based on $\|h_{\mathcal{T}}^s\theta_\nu\|_{L_2(\Omega)}^2 \eqsim |\omega_\nu|^{1+\frac{2s}{d}}$, we made an harmless modification to the operator $B_{\mathcal{T}}^{\mathscr{B}}$ from Lemma 4.4.4.

4.5 Extensions

4.5.1 Higher order

Add the superscript 1 to the spaces defined so far, e.g. write $\mathscr{V}_{\mathcal{T}}^1$ for $\mathscr{S}_{\mathcal{T}}^{0,1}$ with its nodal basis $\Phi_{\mathcal{T}}^1$, and similarly use $G_{\mathcal{T}}^1$ for the associated preconditioner from either Sect. 4.4.2 or Sect. 4.4.3.

We will now consider a (family of) higher order continuous piecewise polynomials, i.e. for some $\ell \in \{2, 3, \dots\}$ let

$$\mathscr{V}_{\mathcal{T}} = \mathscr{S}_{\mathcal{T}}^{0,\ell} := \{u \in H_{0,\gamma}^1(\Omega) : u|_T \in \mathcal{P}_\ell\,(T \in \mathcal{T})\} \subset \mathscr{V}.$$

Because we have an inverse inequality on $\mathscr{V}_{\mathcal{T}}$, we can construct a uniform preconditioner $G_{\mathcal{T}} \in \mathcal{L}\mathrm{is}(\mathscr{V}_{\mathcal{T}}', \mathscr{V}_{\mathcal{T}})$ using an additive subspace correction method. That is, we consider the overlapping decomposition $\mathscr{V}_{\mathcal{T}} = \mathscr{V}_{\mathcal{T}}^1 + \mathscr{V}_{\mathcal{T}}^2$, where these spaces are given by

$$\mathscr{V}_{\mathcal{T}} = (\mathscr{V}_{\mathcal{T}}, \|\cdot\|_{\mathscr{V}}), \quad \mathscr{V}_{\mathcal{T}}^1 = (\mathscr{V}_{\mathcal{T}}^1, \|\cdot\|_{\mathscr{V}}), \quad \mathscr{V}_{\mathcal{T}}^2 = (\mathscr{V}_{\mathcal{T}}, \|h_{\mathcal{T}}^{-s}\cdot\|_{L_2(\Omega)}).$$

Proposition 4.5.1. *For $k \in \{1,2\}$, let $G_{\mathcal{T}}^k \in \mathcal{L}\mathrm{is}_c((\mathscr{V}_{\mathcal{T}}^k)', \mathscr{V}_{\mathcal{T}}^k)$, then for $I_{\mathcal{T}}^k : \mathscr{V}_{\mathcal{T}}^k \to \mathscr{V}_{\mathcal{T}}$ the trivial embedding, we find that $G_{\mathcal{T}} := \sum_{k=1}^2 I_{\mathcal{T}}^k G_{\mathcal{T}}^k (I_{\mathcal{T}}^k)' \in \mathcal{L}\mathrm{is}_c(\mathscr{V}_{\mathcal{T}}', \mathscr{V}_{\mathcal{T}})$, with*

$$\begin{aligned}\|G_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{V}_{\mathcal{T}}', \mathscr{V}_{\mathcal{T}})} &\lesssim \max_{k=1,2}\|G_{\mathcal{T}}^k\|_{\mathcal{L}((\mathscr{V}_{\mathcal{T}}^k)', \mathscr{V}_{\mathcal{T}}^k)},\\ \|\Re(G_{\mathcal{T}})^{-1}\|_{\mathcal{L}(\mathscr{V}_{\mathcal{T}}, \mathscr{V}_{\mathcal{T}}')} &\lesssim \max_{k=1,2}\|\Re(G_{\mathcal{T}}^k)^{-1}\|_{\mathcal{L}(\mathscr{V}_{\mathcal{T}}^k, (\mathscr{V}_{\mathcal{T}}^k)')}.\end{aligned}$$

Proof. We have the (standard) inverse inequality $\|u\|_{\mathscr{V}} \lesssim \|h_{\mathcal{T}}^{-s}u\|_{L_2(\Omega)}$ for $u \in \mathscr{V}_{\mathcal{T}}$. Let $u \in \mathscr{V}_{\mathcal{T}}$, then for any $(u_1, u_2) \in \mathscr{V}_{\mathcal{T}}^1 \times \mathscr{V}_{\mathcal{T}}^2$ with $u_1 + u_2 = u$ we find

$$\|u\|_{\mathscr{V}} \leq \|u_1\|_{\mathscr{V}} + \|u_2\|_{\mathscr{V}} \lesssim \|u_1\|_{\mathscr{V}} + \|h_{\mathcal{T}}^{-s}u_2\|_{L_2(\Omega)}.$$

Denote $\Pi_{\mathcal{T}}^1 \colon H_{0,\gamma}^1(\Omega) \to \mathscr{V}_{\mathcal{T}}^1$ for the Scott-Zhang interpolator ([SZ90]). For $u \in \mathscr{V}_{\mathcal{T}}$, take $u_1 = \Pi_{\mathcal{T}}^1 u \in \mathscr{V}_{\mathcal{T}}^1$ and $u_2 = u - \Pi_{\mathcal{T}}^1 u \in \mathscr{V}_{\mathcal{T}}^2$, then from approximation properties of the interpolator we infer

$$\begin{aligned}\|u_1\|_{\mathscr{V}} + \|h_{\mathcal{T}}^{-s}u_2\|_{L_2(\Omega)} &\leq \|u\|_{\mathscr{V}} + \|u_2\|_{\mathscr{V}} + \|h_{\mathcal{T}}^{-s}u_2\|_{L_2(\Omega)} \\ &\lesssim \|u\|_{\mathscr{V}} + \|h_{\mathcal{T}}^{-s}u_2\|_{L_2(\Omega)} \lesssim \|u\|_{\mathscr{V}}.\end{aligned}$$

Since apparently for $u \in \mathscr{V}_{\mathcal{T}}$,

$$\|u\|_{\mathscr{V}} \eqsim \inf\big\{\|u_1\|_{\mathscr{V}} + \|h_{\mathcal{T}}^{-s}u_2\|_{L_2(\Omega)} \colon u_1 \in \mathscr{V}_1,\, u_2 \in \mathscr{V}_2,\, u_1 + u_2 = u\big\},$$

the result follows from subspace correction methods theory, e.g. [Osw94]. □

On the space $\mathscr{V}_{\mathcal{T}}^1$ we can apply the operator $G_{\mathcal{T}}^1$ constructed earlier, whereas on $\mathscr{V}_{\mathcal{T}}^2$ a simple scaling operator suffices. Denote $N_{\mathcal{T}}^{0,\ell}$ for the set of canonical Lagrange evaluation points of $\mathscr{S}_{\mathcal{T}}^{0,\ell}$, and let $\Phi_{\mathcal{T}}^\ell = \{\phi_\nu^\ell \colon \nu \in N_{\mathcal{T}}^{0,\ell}\}$ be the corresponding nodal basis. For some constant $\beta_2 > 0$, define an operator $R_{\mathcal{T}} \colon \mathscr{V}_{\mathcal{T}}^2 \to (\mathscr{V}_{\mathcal{T}}^2)'$ by

$$(R_{\mathcal{T}}u)(w) := \beta_2^{-1} \sum_{\nu \in N_{\mathcal{T}}^{0,\ell}} \|h_{\mathcal{T}}^{-s}\phi_\nu^\ell\|_{L_2(\Omega)}^2 u(\nu)w(\nu).$$

Proposition 4.5.2. *The operator $G_{\mathcal{T}}^2 := R_{\mathcal{T}}^{-1}$ satisfies $G_{\mathcal{T}}^2 \in \mathcal{L}\mathrm{is}_c((\mathscr{V}_{\mathcal{T}}^2)', \mathscr{V}_{\mathcal{T}}^2)$ uniformly.*

Proof. It is not hard to see that the result follows if $\Phi_{\mathcal{T}}^\ell$ is a (uniformly) $\|h_{\mathcal{T}}^{-s} \cdot \|_{L_2(\Omega)}$-stable basis. Writing $N_T^{0,\ell} := N_{\mathcal{T}}^{0,\ell} \cap T$, this stability can be deduced from

$$\begin{aligned}\Big\|h_{\mathcal{T}}^{-s}\sum_{\nu \in N_{\mathcal{T}}^{0,\ell}} c_\nu \phi_\nu^\ell\Big\|_{L_2(\Omega)}^2 &= \sum_{T \in \mathcal{T}} h_T^{-2s}\Big\|\sum_{\nu \in N_T^{0,\ell}} c_\nu \phi_\nu^\ell\Big\|_{L_2(T)}^2 \eqsim \sum_{T \in \mathcal{T}} h_T^{-2s}\sum_{\nu \in N_T^{0,\ell}} |c_\nu|^2 \|\phi_\nu^\ell\|_{L_2(T)}^2 \\ &= \sum_{\nu \in N_{\mathcal{T}}^{0,\ell}} |c_\nu|^2 \|h_{\mathcal{T}}^{-s}\phi_\nu^\ell\|_{L_2(\Omega)}^2.\end{aligned}$$

□

Implementation

Equipping $\mathscr{V}_{\mathcal{T}}$ and $\mathscr{V}_{\mathcal{T}}^2$ with $\Phi_{\mathcal{T}}^\ell$, and $\mathscr{V}_{\mathcal{T}}^1$ with $\Phi_{\mathcal{T}}^1$, the matrix representation of $G_{\mathcal{T}} := \sum_{k=1}^2 I_{\mathcal{T}}^k G_{\mathcal{T}}^k (I_{\mathcal{T}}^k)' \in \mathcal{L}\mathrm{is}_c(\mathscr{V}_{\mathcal{T}}', \mathscr{V}_{\mathcal{T}})$ is given by

$$\boldsymbol{G}_{\mathcal{T}} = \boldsymbol{q}_{\mathcal{T}} \boldsymbol{G}_{\mathcal{T}}^1 \boldsymbol{q}_{\mathcal{T}}^\top + \boldsymbol{G}_{\mathcal{T}}^2,$$

with $\boldsymbol{G}^1_{\mathcal{T}}$ either from Sect. 4.4.2 or Sect. 4.4.3,

$$(\boldsymbol{q}_{\mathcal{T}})_{\nu'\nu} = \phi^\ell_{\nu'}(\nu) \quad (\nu' \in N^{0,\ell}_{\mathcal{T}}, \nu \in N^{0,1}_{\mathcal{T}}).$$

and

$$\boldsymbol{G}^2_{\mathcal{T}} = \beta_2 \operatorname{diag}\{\|h^{-s}_{\mathcal{T}}\phi^\ell_\nu\|^{-2}_{L_2(\Omega)} : \nu \in N^{0,\ell}_{\mathcal{T}}\}.$$

4.5.2 Manifolds

Let Γ be a compact d-dimensional Lipschitz, piecewise smooth manifold in $\mathbb{R}^{d'}$ for some $d' \geq d$ with or without boundary $\partial\Gamma$. For some closed measurable $\gamma \subset \partial\Gamma$ and $s \in [0,1]$, let

$$\mathscr{V} := [L_2(\Gamma), H^1_{0,\gamma}(\Gamma)]_{s,2}, \quad \mathscr{W} := \mathscr{V}'.$$

We assume that Γ is given as the closure of the disjoint union of $\cup_{i=1}^p \chi_i(\Omega_i)$, with, for $1 \leq i \leq p$, $\chi_i \colon \mathbb{R}^d \to \mathbb{R}^{d'}$ being some smooth regular parametrization, and $\Omega_i \subset \mathbb{R}^d$ an open polytope. W.l.o.g. assuming that for $i \neq j$, $\overline{\Omega}_i \cap \overline{\Omega}_j = \emptyset$, we define

$$\chi \colon \Omega := \cup_{i=1}^p \Omega_i \to \cup_{i=1}^p \chi_i(\Omega_i) \text{ by } \chi|_{\Omega_i} = \chi_i.$$

Let $\mathbb{T}$ be a family of conforming partitions $\mathcal{T}$ of Γ into 'panels' such that, for $1 \leq i \leq p$, $\chi^{-1}(\mathcal{T}) \cap \Omega_i$ is a uniformly shape regular conforming partition of Ω_i into d-simplices (that for $d = 1$ satisfies a uniform K-mesh property). We assume that γ is a (possibly empty) union of 'faces' of $T \in \mathcal{T}$ (i.e., sets of type $\chi_i(e)$, where e is a $(d-1)$-dimensional face of $\chi_i^{-1}(T)$).

The usual lowest order boundary element spaces are defined by

$$\begin{aligned}\mathscr{S}^{-1,0}_{\mathcal{T}} &:= \{u \in L_2(\Gamma) \colon u \circ \chi|_{\chi^{-1}(T)} \in \mathcal{P}_0 \ (T \in \mathcal{T})\},, \\ \mathscr{S}^{0,1}_{\mathcal{T}} &:= \{u \in H^1_{0,\gamma}(\Gamma) \colon u \circ \chi|_{\chi^{-1}(T)} \in \mathcal{P}_1 \ (T \in \mathcal{T})\},\end{aligned}$$

with their canonical bases denoted as $\Sigma_{\mathcal{T}} = \{\mathbb{1}_T \colon T \in \mathcal{T}\}$ and $\Phi_{\mathcal{T}} = \{\phi_\nu \colon \nu \in N^0_{\mathcal{T}}\}$, respectively, with $N^0_{\mathcal{T}}$ the vertices of $\mathcal{T}$ not on γ.

The construction of the preconditioners in the domain case relied on the explicit construction of a collection $\Psi_{\mathcal{T}}$ biorthogonal to $\Phi_{\mathcal{T}}$, and on the explicit computation of a (bi)orthogonal projection of $\mathscr{W}_{\mathcal{T}} := \operatorname{span} \Psi_{\mathcal{T}}$ onto either $\mathscr{S}^{-1,0}_{\mathcal{T}}$ or $\mathscr{S}^{0,1}_{\mathcal{T}}$, where orthogonality was interpreted w.r.t. the $L_2(\Omega)$-scalar product. Both the construction of $\Psi_{\mathcal{T}}$ and the computation of the (bi)orthogonal projection could be reduced to computations on the individual elements in the partition, which yielded explicit expressions.

When attempting to transfer everything to the manifold case, a problem is the appearance of a generally non-constant weight $x \mapsto |\partial\chi(x)|$ in the $L_2(\Gamma)$-scalar product

$$\langle u, v\rangle_{L_2(\Gamma)} = \int_\Omega u(\chi(x))v(\chi(x))|\partial\chi(x)|\,dx.$$

To deal with this, following Sect. 2.3.2, on $L_2(\Gamma)$ we define an additional 'mesh-dependent' scalar product

$$\langle u, v\rangle_{\mathcal{T}} := \sum_{T\in\mathcal{T}} \frac{|T|}{|\chi^{-1}(T)|}\int_{\chi^{-1}(T)} u(\chi(x))v(\chi(x))dx.$$

which is constructed by replacing on each $\chi^{-1}(T)$, the Jacobian $|\partial\chi|$ by its average $\frac{|T|}{|\chi^{-1}(T)|}$ over $\chi^{-1}(T)$, and interpret (bi)orthogonality with respect to this scalar product.

Now all steps in the *construction* of the preconditioners carry over, and yield preconditioners for the manifold case whose implementations are exactly as described in Sect. 4.4.2 and Sect. 4.4.3, where the patch volumes $|\omega_{\mathcal{T},\nu}|$ now should be read as the volumes of the patches on Γ.

To *prove* that the constructed preconditioners are indeed uniform preconditioners requires additional work due to the use of the mesh-dependent scalar product. We refer to Chapter 2 for details. The key ingredient is that not only the norm associated to $\langle\cdot,\cdot\rangle_{L_2(\Gamma)}$ is uniformly equivalent to $\|\cdot\|_{L_2(\Gamma)}$, but also that $\langle\cdot,\cdot\rangle_{L_2(\Gamma)}$ and $\langle\cdot,\cdot\rangle_{\mathcal{T}}$ are close in the sense that

$$|\langle v,u\rangle_{\mathcal{T}} - \langle v,u\rangle_{L_2(\Gamma)}| \lesssim \|h_{\mathcal{T}}v\|_{L_2(\Gamma)}\|u\|_{L_2(\Gamma)} \quad (v,u\in L_2(\Gamma)).$$

4.6 Numerical experiments

Let $\Gamma = \partial[0,1]^3 \subset \mathbb{R}^3$ be the boundary of the unit cube, $\mathscr{V} := H^{1/2}(\Gamma)$, $\mathscr{W} := H^{-1/2}(\Gamma)$, and $\mathscr{V}_{\mathcal{T}} = \mathscr{S}^{0,\ell}_{\mathcal{T}} \subset \mathscr{V}$ the trial space of continuous piecewise polynomials of degree ℓ w.r.t. a partition $\mathcal{T}$. We shall evaluate preconditioning of essentially a discretized Hypersingular Integral operator.

The Hypersingular Integral operator $\tilde{A} \in \mathcal{L}(\mathscr{V},\mathscr{V}')$ is only semi-coercive, since it has a non-trivial kernel equal to $\mathrm{span}\{\mathbb{1}\}$. Solving $\tilde{A}u = f$ for f with $f(\mathbb{1}) = 0$ is, however, equivalent to solving $Au = f$ with A given by $(Au)(v) = (\tilde{A}u)(v) + \alpha\langle u,\mathbb{1}\rangle_{L_2(\Gamma)}\langle v,\mathbb{1}\rangle_{L_2(\Gamma)}$ for some $\alpha > 0$. This operator A is in $\mathcal{L}\mathrm{is}_c(\mathscr{V},\mathscr{V}')$, and we shall consider preconditioning discretizations $A_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{V}_{\mathcal{T}},\mathscr{V}'_{\mathcal{T}})$ of A. By comparing different values numerically, we found $\alpha = 0.05$ to give good results in our examples.

As opposite order operator B we take the Weakly Singular integral operator, which on compact 2-dimensional manifolds is known to be in $\mathcal{L}\mathrm{is}_c(\mathscr{W},\mathscr{W}')$. We will compare preconditioners $G_{\mathcal{T}}$ based on the discretizations $B^{\mathscr{U}}_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{U}_{\mathcal{T}},\mathscr{U}'_{\mathcal{T}})$ of B, for $\mathscr{U}_{\mathcal{T}} = \mathscr{S}^{-1,0}_{\mathcal{T}}$ or $\mathscr{U}_{\mathcal{T}} = \mathscr{S}^{0,1}_{\mathcal{T}}$ equipped with the canonical bases $\Sigma_{\mathcal{T}} = \{\mathbb{1}_T : T\in\mathcal{T}\}$ and $\Phi_{\mathcal{T}} = \{\phi_\nu : \nu\in N_{\mathcal{T}}\}$, respectively, cf. Sect. 4.4.2 or Sect. 4.4.3.

For $\ell = 1$ (the lowest order case) and $\mathscr{V}_{\mathcal{T}}$ being equipped with $\Phi_{\mathcal{T}}$, the matrix representation of the preconditioner $G_{\mathcal{T}}$ reads either as (Sect. 4.4.2)

$$\boldsymbol{G}_{\mathcal{T}} = \boldsymbol{G}^{-1,0}_{\mathcal{T}} = \boldsymbol{D}^{-1}_{\mathcal{T}}\big(\boldsymbol{p}^{\top}_{\mathcal{T}}\boldsymbol{B}^{\mathscr{U}}_{\mathcal{T}}\boldsymbol{p}_{\mathcal{T}} + \beta_1\boldsymbol{D}^{3/2}_{\mathcal{T}}\big)\boldsymbol{D}^{-1}_{\mathcal{T}}$$

where $\boldsymbol{B}_{\mathcal{T}}^{\mathscr{U}} = (B\Sigma_{\mathcal{T}})(\Sigma_{\mathcal{T}})$, $\boldsymbol{D}_{\mathcal{T}} = \operatorname{diag}\{|\omega_\nu| \colon \nu \in N_{\mathcal{T}}\}$, $(\boldsymbol{p}_{\mathcal{T}})_{T\nu} = \begin{cases} 1 & \text{if } T \subset \omega_\nu, \\ 0 & \text{otherwise,} \end{cases}$

and $\beta_1 > 0$ is some constant, or as (Sect. 4.4.3)

$$\boldsymbol{G}_{\mathcal{T}} = \boldsymbol{G}_{\mathcal{T}}^{0,1} = \boldsymbol{D}_{\mathcal{T}}^{-1}\big(\boldsymbol{B}_{\mathcal{T}}^{\mathscr{U}} + \beta_1 \boldsymbol{D}_{\mathcal{T}}^{3/2}\big)\boldsymbol{D}_{\mathcal{T}}^{-1}$$

where $\boldsymbol{B}_{\mathcal{T}}^{\mathscr{U}} = (B\Phi_{\mathcal{T}})(\Phi_{\mathcal{T}})$, $\boldsymbol{D}_{\mathcal{T}} = \operatorname{diag}\{|\frac{\omega_\nu}{d+1}| \colon \nu \in N_{\mathcal{T}}\}$, and $\beta_1 > 0$ is some constant.

For $\ell > 1$ denote the above $\boldsymbol{G}_{\mathcal{T}}$ by either $\boldsymbol{G}_{\mathcal{T}}^{1,-1,0}$ or $\boldsymbol{G}_{\mathcal{T}}^{1,0,1}$, then, with $\mathscr{V}_{\mathcal{T}} = \mathscr{S}_{\mathcal{T}}^{0,\ell}$ being equipped with the standard nodal basis $\{\phi_\nu^\ell \colon \nu \in N_{\mathcal{T}}^\ell\}$, the matrix representation of the preconditioner $G_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c((\mathscr{S}_{\mathcal{T}}^{0,\ell})', \mathscr{S}_{\mathcal{T}}^{0,\ell})$ from Sect. 4.5.1 is

$$\boldsymbol{G}_{\mathcal{T}}^{*} = \boldsymbol{q}_{\mathcal{T}}\boldsymbol{G}_{\mathcal{T}}^{1,*}\boldsymbol{q}_{\mathcal{T}}^{\top} + \beta_2 \operatorname{diag}\{\|h_{\mathcal{T}}^{-\frac{1}{2}}\phi_\nu^\ell\|_{L_2(\Omega)}^{-2} \colon \nu \in N_{\mathcal{T}}^\ell\},$$

where either $* = -1, 0$ or $* = 0, 1$, and $(\boldsymbol{q}_{\mathcal{T}})_{\nu'\nu} = \phi_{\nu'}^\ell(\nu)$ ($\nu' \in N_{\mathcal{T}}^\ell, \nu \in N_{\mathcal{T}}^1$).

The (full) matrix representations of the discretized singular integral operators $\boldsymbol{A}_{\mathcal{T}}$ and $\boldsymbol{B}_{\mathcal{T}}^{\mathscr{U}}$ are calculated using the BEM++ software package [ŚBA+15]. Condition numbers are determined using Lanczos iteration with respect to $|\!|\!|\cdot|\!|\!| := \|\boldsymbol{A}_{\mathcal{T}}^{\frac{1}{2}} \cdot \|$.

4.6.1 Uniform refinements

Consider a conforming triangulation $\mathcal{T}_1$ of Γ consisting of 2 triangles per side, so 12 triangles with 8 vertices in total. We let $\mathbb{T}$ be the sequence $\{\mathcal{T}_k\}_{k\geq 1}$ of uniform newest vertex bisections, where $\mathcal{T}_k \succ \mathcal{T}_{k-1}$ is found by bisecting each triangle from $\mathcal{T}_{k-1}$.

With $\mathscr{V}_{\mathcal{T}} = \mathscr{S}_{\mathcal{T}}^{0,1}$, Table 4.1 compares the condition numbers for the preconditioned system given by Sect. 4.4.2 ($\mathscr{U}_{\mathcal{T}} = \mathscr{S}_{\mathcal{T}}^{-1,0}$) and by Sect. 4.4.3 ($\mathscr{U}_{\mathcal{T}} = \mathscr{S}_{\mathcal{T}}^{0,1}$). We see that the condition numbers remain nicely bounded, and that both choices give similar condition numbers.

Instead of using the 'full matrices', we can consider compressed hierarchical matrices to approximate the stiffness matrices $\mathbf{A}_{\mathcal{T}}$ and $\mathbf{B}_{\mathcal{T}}^{\mathscr{U}}$ for finer partitions. Table 4.2 gives the condition numbers, again for uniform refinements, but now using hierarchical matrices based on adaptive cross approximation [Hac99, Beb00]. We see that even for large systems, our preconditioner gives very satisfactory results.

Finally, consider the (higher order) trial space $\mathscr{V}_{\mathcal{T}} = \mathscr{S}_{\mathcal{T}}^{0,3}$. Table 4.3 gives condition numbers for the preconditioned system, using the method described in Sect. 4.5.1.

Table 4.1. Spectral condition numbers of the preconditioned hypersingular system, using uniform refinements, discretized by continuous piecewise linears $\mathscr{S}_{\mathcal{T}}^{0,1}$, with $\alpha = 0.05$. The preconditioners $\boldsymbol{G}_{\mathcal{T}}^{-1,0}$ and $\boldsymbol{G}_{\mathcal{T}}^{0,1}$ are constructed using the single layer operator discretized on $\mathscr{U}_{\mathcal{T}} = \mathscr{S}_{\mathcal{T}}^{-1,0}$ (Sect. 4.4.2) and $\mathscr{U}_{\mathcal{T}} = \mathscr{S}_{\mathcal{T}}^{0,1}$ (Sec 4.4.3), respectively, where have used $\beta_1 = 0.65$ in the first case and $\beta_1 = 0.34$ in the second case.

dofs	$\kappa_S(\boldsymbol{A}_{\mathcal{T}})$	$\kappa_S(\boldsymbol{G}_{\mathcal{T}}^{-1,0}\boldsymbol{A}_{\mathcal{T}})$	$\kappa_S(\boldsymbol{G}_{\mathcal{T}}^{0,1}\boldsymbol{A}_{\mathcal{T}})$
14	3.0	2.71	2.64
50	7.1	2.36	2.37
194	14.2	2.25	2.26
770	28.7	2.30	2.27
3074	57.8	2.29	2.27
12290	115.7	2.29	2.27
49154	231.4	2.30	2.27

Table 4.2. In the same setting as Table 4.1, but using compressed hierarchical matrices.

dofs	$\kappa_S(\boldsymbol{A}_{\mathcal{T}})$	$\kappa_S(\boldsymbol{G}_{\mathcal{T}}^{-1,0}\boldsymbol{A}_{\mathcal{T}})$	$\kappa_S(\boldsymbol{G}_{\mathcal{T}}^{0,1}\boldsymbol{A}_{\mathcal{T}})$
12290	115.6	2.29	2.27
24578	168.7	2.24	2.24
49154	231.3	2.30	2.27
98306	336.9	2.25	2.25
196610	461.7	2.30	2.28
393218	671.9	2.27	2.28
786434	751.6	2.30	2.30

Table 4.3. Spectral condition numbers of the preconditioned hypersingular system, using uniform refinements, discretized by continuous piecewise cubics $\mathscr{S}_{\mathcal{T}}^{0,3}$, with $\alpha = 0.05$. The higher order preconditioners $\boldsymbol{G}_{\mathcal{T}}^{-1,0}$ and $\boldsymbol{G}_{\mathcal{T}}^{0,1}$ are constructed as described in Sect. 4.5.1, by using the preconditioners from Table 4.1 with constants $\beta_1 = 0.65, \beta_2 = 0.065$ in the first case and $\beta_1 = 0.34, \beta_2 = 0.065$ in the second case.

dofs	$\kappa_S(\boldsymbol{A}_{\mathcal{T}})$	$\kappa_S(\boldsymbol{G}_{\mathcal{T}}^{-1,0}\boldsymbol{A}_{\mathcal{T}})$	$\kappa_S(\boldsymbol{G}_{\mathcal{T}}^{0,1}\boldsymbol{A}_{\mathcal{T}})$
56	19.49	4.75	4.72
218	36.27	5.18	5.17
866	74.78	6.23	6.20
3458	150.73	6.55	6.48
13826	301.97	6.63	6.57
55298	603.86	6.65	6.58

TABLE 4.4. Spectral condition numbers of the preconditioned hypersingular system discretized by $\mathscr{S}_{\mathcal{T}}^{0,1}$ using local refinements at each of the eight cube corners. Both preconditioners $\boldsymbol{G}_{\mathcal{T}}^{-1,0}$ and $\boldsymbol{G}_{\mathcal{T}}^{0,1}$ are constructed with same parameters as in Table 4.1, and are compared against diagonal preconditioning. The second column is defined by $h_{\mathcal{T},min} := \min_{T\in\mathcal{T}} h_T$.

dofs	$h_{\mathcal{T},min}$	$\kappa_S(\mathrm{diag}(\boldsymbol{A}_{\mathcal{T}})^{-1}\boldsymbol{A}_{\mathcal{T}})$	$\kappa_S(\boldsymbol{G}_{\mathcal{T}}^{-1,0}\boldsymbol{A}_{\mathcal{T}})$	$\kappa_S(\boldsymbol{G}_{\mathcal{T}}^{0,1}\boldsymbol{A}_{\mathcal{T}})$
8	$1.4 \cdot 10^{0}$	2.15	2.83	2.68
14	$1.0 \cdot 10^{0}$	2.79	2.71	2.64
314	$1.1 \cdot 10^{-2}$	12.11	2.21	2.20
626	$1.2 \cdot 10^{-4}$	13.18	2.31	2.30
938	$1.3 \cdot 10^{-6}$	13.43	2.36	2.36
1250	$1.4 \cdot 10^{-8}$	13.51	2.39	2.38
1562	$1.6 \cdot 10^{-10}$	13.53	2.41	2.39
1850	$2.5 \cdot 10^{-12}$	13.55	2.41	2.40

4.6.2 Local refinements

Here we take $\mathbb{T}$ to be the sequence $\{\mathcal{T}_k\}_{k\geq 1}$ of locally refined triangulations, where $\mathcal{T}_k \succ \mathcal{T}_{k-1}$ is constructed using conforming newest vertex bisection to refine all triangles in $\mathcal{T}_{k-1}$ that touch a corner of the cube.

Table 4.4 gives condition numbers of the preconditioned hypersingular system discretized by continuous piecewise linears, i.e. $\mathscr{V}_{\mathcal{T}} = \mathscr{S}_{\mathcal{T}}^{0,1}$. The condition numbers remain bounded under local refinements, confirming uniformity of the preconditioner w.r.t. $\mathbb{T}$.

4.7 Conclusion

Using the framework of operator preconditioning, we have constructed uniform preconditioners for elliptic operators of orders $2s \in [0, 2]$ discretized by continuous finite (or boundary) elements. The evaluation of the preconditioners requires the application of an opposite order operator plus minor cost of linear complexity. Compared to earlier proposals, both the construction of a so-called dual-mesh and the inversion of a non-diagonal matrix are avoided, and our results are valid without constraints on the mesh-grading. For lowest order finite elements the computed condition numbers of the preconditioned system are below 2.5.

5 The simplest case

5.1 Introduction

This chapter deals with the construction of uniform preconditioners for negative and positive order operators, discretized by continuous piecewise polynomial trial spaces, using the framework of 'operator preconditioning' [Hip06], see also [SW98, Ste02, BC07, HJHUT20].

For some d-dimensional closed domain (or manifold) Ω and an $s \in [0,1]$, we consider the (fractional) Sobolev space $H^s(\Omega)$ and its dual that we denote by $H^{-s}(\Omega)$. Let $(\mathscr{S}_\mathcal{T})_{\mathcal{T}\in\mathbb{T}}$ be a family of *continuous piecewise* polynomials of some fixed degree ℓ w.r.t. uniformly shape regular, possibly locally refined, partitions.

Given some families of uniformly boundedly invertible operators

$$A_\mathcal{T}\colon \big(\mathscr{S}_\mathcal{T}, \|\cdot\|_{H^{-s}(\Omega)}\big) \to \big(\mathscr{S}_\mathcal{T}, \|\cdot\|_{H^{-s}(\Omega)}\big)',$$
$$B_\mathcal{T}\colon \big(\mathscr{S}_\mathcal{T}, \|\cdot\|_{H^{s}(\Omega)}\big) \to \big(\mathscr{S}_\mathcal{T}, \|\cdot\|_{H^{s}(\Omega)}\big)',$$

we are interested in constructing a *preconditioner* for $A_\mathcal{T}$ using operator preconditioning with $B_\mathcal{T}$, *and* vice versa. To this end, we introduce a uniformly boundedly invertible operator $D_\mathcal{T}\colon \big(\mathscr{S}_\mathcal{T}, \|\cdot\|_{H^{-s}(\Omega)}\big) \to \big(\mathscr{S}_\mathcal{T}, \|\cdot\|_{H^{s}(\Omega)}\big)'$, yielding preconditioned systems $D_\mathcal{T}^{-1}B_\mathcal{T}(D_\mathcal{T}')^{-1}A_\mathcal{T}$ and $(D_\mathcal{T}')^{-1}A_\mathcal{T}D_\mathcal{T}^{-1}B_\mathcal{T}$ that are uniformly boundedly invertible.

In Chapters 2 and 4 we already constructed such preconditioners in a more general setting where *different* ansatz spaces were used to define $A_\mathcal{T}$ and $B_\mathcal{T}$. The setting studied in the current work, however, allows for preconditioners with a remarkably simple implementation.

A typical setting is that for some $A\colon H^{-s}(\Omega) \to H^s(\Omega)$ and $B\colon H^s(\Omega) \to H^{-s}(\Omega)$, both boundedly invertible and coercive, it holds that $(A_\mathcal{T}u)(v) := (Au)(v)$ and $(B_\mathcal{T}u)(v) := (Bu)(v)$ with $u, v \in \mathscr{S}_\mathcal{T}$. An example for $s = \frac{1}{2}$ is that A is the Single Layer Integral operator and B is the Hypersingular Integral operator. For this case, continuity of piecewise polynomial trial functions is required for discretizing B, but not for A, for which often discontinuous piecewise polynomials are employed. Nevertheless, when the solution of

the Single Layer Integral equation is expected to be smooth, e.g., when Ω is a smooth manifold, then it is advantageous to take an ansatz space of continuous (or even smoother) functions also for A.

An obvious choice for $D_{\mathcal{T}}$ would be to consider $(D_{\mathcal{T}}u)(v) := \langle u, v\rangle_{L_2(\Omega)}$. However, a problem becomes apparent when one considers the matrix representation $\boldsymbol{D}_{\mathcal{T}}$ of $D_{\mathcal{T}}$ in the standard basis being the mass matrix: the inverse matrix $\boldsymbol{D}_{\mathcal{T}}^{-1}$, that appears in the preconditioned system, is densely populated. In view of application cost, this inverse matrix has to be approximated, where it generally can be expected that, in order to obtain a uniform preconditioner, approximation errors have to decrease with a decreasing (minimal) mesh size, which will be confirmed in a numerical experiment. To circumvent this issue, we will introduce a $D_{\mathcal{T}}$ that has a *diagonal* matrix representation, so that its inverse can be exactly evaluated.

5.1.1 Notation

In this work, by $\lambda \lesssim \mu$ we mean that λ can be bounded by a multiple of μ, independently of parameters which λ and μ may depend on, with the sole exception of the space dimension d, or in the manifold case, on the parametrization of the manifold that is used to define the finite element spaces on it. Obviously, $\lambda \gtrsim \mu$ is defined as $\mu \lesssim \lambda$, and $\lambda \eqsim \mu$ as $\lambda \lesssim \mu$ and $\lambda \gtrsim \mu$.

For normed linear spaces $\mathscr{Y}$ and $\mathscr{Z}$, in this chapter for convenience over $\mathbb{R}$, $\mathcal{L}(\mathscr{Y}, \mathscr{Z})$ will denote the space of bounded linear mappings $\mathscr{Y} \to \mathscr{Z}$ endowed with the operator norm $\|\cdot\|_{\mathcal{L}(\mathscr{Y},\mathscr{Z})}$. The subset of invertible operators in $\mathcal{L}(\mathscr{Y}, \mathscr{Z})$ with inverses in $\mathcal{L}(\mathscr{Z}, \mathscr{Y})$ will be denoted as $\mathcal{L}\mathrm{is}(\mathscr{Y}, \mathscr{Z})$.

For $\mathscr{Y}$ a reflexive Banach space and $C \in \mathcal{L}(\mathscr{Y}, \mathscr{Y}')$ being *coercive*, i.e.,

$$\inf_{0\neq y\in\mathscr{Y}} \frac{(Cy)(y)}{\|y\|^2_{\mathscr{Y}}} > 0,$$

both C and $\Re(C) := \frac{1}{2}(C + C')$ are in $\mathcal{L}\mathrm{is}(\mathscr{Y}, \mathscr{Y}')$ with

$$\|\Re(C)\|_{\mathcal{L}(\mathscr{Y},\mathscr{Y}')} \leq \|C\|_{\mathcal{L}(\mathscr{Y},\mathscr{Y}')},$$
$$\|C^{-1}\|_{\mathcal{L}(\mathscr{Y}',\mathscr{Y})} \leq \|\Re(C)^{-1}\|_{\mathcal{L}(\mathscr{Y}',\mathscr{Y})} = \Big(\inf_{0\neq y\in\mathscr{Y}} \frac{(Cy)(y)}{\|y\|^2_{\mathscr{Y}}}\Big)^{-1}.$$

The subset of coercive operators in $\mathcal{L}\mathrm{is}(\mathscr{Y}, \mathscr{Y}')$ is denoted as $\mathcal{L}\mathrm{is}_c(\mathscr{Y}, \mathscr{Y}')$. If $C \in \mathcal{L}\mathrm{is}_c(\mathscr{Y}, \mathscr{Y}')$, then $C^{-1} \in \mathcal{L}\mathrm{is}_c(\mathscr{Y}', \mathscr{Y})$ and $\|\Re(C^{-1})^{-1}\|_{\mathcal{L}(\mathscr{Y},\mathscr{Y}')} \leq \|C\|^2_{\mathcal{L}(\mathscr{Y},\mathscr{Y}')}\|\Re(C)^{-1}\|_{\mathcal{L}(\mathscr{Y}',\mathscr{Y})}$.

Given a family of operators $C_i \in \mathcal{L}\mathrm{is}(\mathscr{Y}_i, \mathscr{Z}_i)$ ($\mathcal{L}\mathrm{is}_c(\mathscr{Y}_i, \mathscr{Z}_i)$), we will write $C_i \in \mathcal{L}\mathrm{is}(\mathscr{Y}_i, \mathscr{Z}_i)$ ($\mathcal{L}\mathrm{is}_c(\mathscr{Y}_i, \mathscr{Z}_i)$) uniformly in i, or simply 'uniform', when

$$\sup_i \max(\|C_i\|_{\mathcal{L}(\mathscr{Y}_i,\mathscr{Z}_i)}, \|C_i^{-1}\|_{\mathcal{L}(\mathscr{Z}_i,\mathscr{Y}_i)}) < \infty,$$

or

$$\sup_i \max(\|C_i\|_{\mathcal{L}(\mathscr{Y}_i,\mathscr{Z}_i)}, \|\Re(C_i)^{-1}\|_{\mathcal{L}(\mathscr{Z}_i,\mathscr{Y}_i)}) < \infty.$$

5.2 Construction of $D_{\mathcal{T}}$ in the domain case

For some d-dimensional domain Ω and an $s \in [0,1]$, we consider the Sobolev spaces

$$H^s(\Omega) := [L_2(\Omega), H^1(\Omega)]_{s,2}, \quad H^{-s}(\Omega) := H^s(\Omega)',$$

which form the Gelfand triple $H^s(\Omega) \hookrightarrow L_2(\Omega) \simeq L_2(\Omega)' \hookrightarrow H^{-s}(\Omega)$.

Remark 5.2.1. In this work, for convenience we restrict ourselves to Sobolev spaces with positive smoothness index which do not incorporate homogeneous Dirichlet boundary conditions and their duals. The proofs given below can however be extended to the setting with boundary conditions, see the arguments found in Chapters 2 and 4.

Let $(\mathcal{T})_{\mathcal{T}\in\mathbb{T}}$ be a family of *conforming* partitions of Ω into (open) *uniformly shape regular* d-simplices. Thanks to the conformity and the uniform shape regularity, for $d > 1$ we know that neighbouring $T, T' \in \mathcal{T}$, i.e. $\overline{T} \cap \overline{T'} \neq \emptyset$, have uniformly comparable sizes. For $d = 1$, we impose this uniform '*K-mesh property*' explicitly.

Fix $\ell > 0$. For $\mathcal{T} \in \mathbb{T}$, let $\mathscr{S}_{\mathcal{T}}$ denote the space of *continuous piecewise polynomials* of degree ℓ w.r.t. $\mathcal{T}$, i.e.,

$$\mathscr{S}_{\mathcal{T}} := \{u \in H^1(\Omega) : u|_T \in \mathcal{P}_\ell \, (T \in \mathcal{T})\}.$$

Additionally, for $r \in [-1,1]$, we will write $\mathscr{S}_{\mathcal{T},r}$ as shorthand notation for the normed linear space $\big(\mathscr{S}_{\mathcal{T}}, \|\cdot\|_{H^r(\Omega)}\big)$.

Denote $N_{\mathcal{T}}$ for the set of the usual Lagrange evaluation points of $\mathscr{S}_{\mathcal{T}}$, and equip the latter space with $\Phi_{\mathcal{T}} = \{\phi_{\mathcal{T},\nu} : \nu \in N_{\mathcal{T}}\}$, being the *canonical nodal basis* defined by $\phi_{\mathcal{T},\nu}(\nu') := \delta_{\nu\nu'}$ $(\nu, \nu' \in N_{\mathcal{T}})$. For $T \in \mathcal{T}$, set $h_T := |T|^{1/d}$ and let $N_T := \overline{T} \cap N_{\mathcal{T}}$ be the set of evaluation points in $\overline{T}$. We will omit notational dependence on $\mathcal{T}$ if it is clear from the context, e.g., we will simply write ϕ_ν.

5.2.1 Operator preconditioning

Given some family of opposite order operators $A_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}_{\mathcal{T},-s}, (\mathscr{S}_{\mathcal{T},-s})')$ and $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}_{\mathcal{T},s}, (\mathscr{S}_{\mathcal{T},s})')$, both uniformly in $\mathcal{T} \in \mathbb{T}$, we are interested in constructing *optimal* preconditioners for both $A_{\mathcal{T}}$ and $B_{\mathcal{T}}$, using the idea of opposite order preconditioning ([Hip06]).

That is, if one has an additional family of operators $D_{\mathcal{T}} \in \mathcal{L}\mathrm{is}(\mathscr{S}_{\mathcal{T},-s}, (\mathscr{S}_{\mathcal{T},s})')$ uniformly in $\mathcal{T} \in \mathbb{T}$, then uniformly preconditioned systems for $A_{\mathcal{T}}$ and $B_{\mathcal{T}}$ are given by

$$\begin{aligned} &D_{\mathcal{T}}^{-1} B_{\mathcal{T}} (D_{\mathcal{T}}')^{-1} A_{\mathcal{T}} \in \mathcal{L}\mathrm{is}(\mathscr{S}_{\mathcal{T},-s}, \mathscr{S}_{\mathcal{T},-s}), \\ &(D_{\mathcal{T}}')^{-1} A_{\mathcal{T}} D_{\mathcal{T}}^{-1} B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}(\mathscr{S}_{\mathcal{T},s}, \mathscr{S}_{\mathcal{T},s}), \end{aligned} \tag{5.1}$$

see the following diagram:

$$\begin{array}{ccc} \mathscr{S}_{\mathcal{T},-s} & \xrightarrow{A_{\mathcal{T}}} & (\mathscr{S}_{\mathcal{T},-s})' \\ {\scriptstyle D_{\mathcal{T}}^{-1}}\uparrow & & \downarrow{\scriptstyle (D'_{\mathcal{T}})^{-1}} \\ (\mathscr{S}_{\mathcal{T},s})' & \xleftarrow[B_{\mathcal{T}}]{} & \mathscr{S}_{\mathcal{T},s} \end{array}.$$

In the following we shall be concerned with constructing a suitable family $D_{\mathcal{T}}$.

An obvious but unsatisfactory choice for $D_{\mathcal{T}}$

An option would be to consider $(D_{\mathcal{T}}u)(v) := \langle u, v\rangle_{L_2(\Omega)}$ $(u, v \in \mathscr{S}_{\mathcal{T}})$, being uniformly in $\mathcal{L}(\mathscr{S}_{\mathcal{T},-s}, (\mathscr{S}_{\mathcal{T},s})')$. For showing boundedness of its inverse, let $Q_{\mathcal{T}}$ be the $L_2(\Omega)$-orthogonal projector onto $\mathscr{S}_{\mathcal{T}}$ then

$$\begin{aligned} \|D_{\mathcal{T}}^{-1}\|^{-1}_{\mathcal{L}((\mathscr{S}_{\mathcal{T},s})',\mathscr{S}_{\mathcal{T},-s})} &= \inf_{0\neq u\in\mathscr{S}_{\mathcal{T},-s}} \sup_{0\neq v\in H^s(\Omega)} \frac{\langle u, v\rangle_{L_2(\Omega)}}{\|u\|_{H^{-s}(\Omega)}\|Q_{\mathcal{T}}v\|_{H^s(\Omega)}} \\ &\geq \|Q_{\mathcal{T}}\|^{-1}_{\mathcal{L}(H^s(\Omega),H^s(\Omega))}, \end{aligned}$$

As follows from Proposition 2.2.3, the converse is also true, i.e., uniform boundedness of $\|D_{\mathcal{T}}^{-1}\|_{\mathcal{L}((\mathscr{S}_{\mathcal{T},s})',\mathscr{S}_{\mathcal{T},-s})}$ is actually *equivalent* to uniform boundedness of $\|Q_{\mathcal{T}}\|_{\mathcal{L}(H^s(\Omega),H^s(\Omega))}$.

This uniform boundedness of $\|Q_{\mathcal{T}}\|_{\mathcal{L}(H^s(\Omega),H^s(\Omega))}$ is well-known for families of *quasi-uniform*, uniformly shape regular conforming partitions of Ω into say d-simplices. It has also been demonstrated for families of locally refined partitions, for $d = 2$ including those that are generated by the newest vertex bisection (NVB) algorithm, see [Car02, GHS16, DST20]. On the other hand, in [BY14] a one-dimensional counterexample was presented in which the $L_2(\Omega)$-orthogonal projector on a family of sufficiently strongly graded, although uniform K meshes, is not $H^1(\Omega)$-stable. Thus, in any case uniform $H^1(\Omega)$-stability cannot hold without assuming some sufficiently mild grading of the meshes.

Aside from this latter theoretical shortcoming, more importantly, there is a computational problem with the current choice of $D_{\mathcal{T}}$. The matrix representation of $D_{\mathcal{T}}$ w.r.t. $\Phi_{\mathcal{T}}$ is the 'mass matrix' $\boldsymbol{D}_{\mathcal{T}} := \langle \Phi_{\mathcal{T}}, \Phi_{\mathcal{T}}\rangle_{L_2(\Omega)}$. Its inverse $\boldsymbol{D}_{\mathcal{T}}^{-1}$, appearing in the preconditioner, is densely populated, and therefore has to be approximated, where generally the error in such approximations has to decrease with a decreasing (minimal) mesh-size in order to arrive at a uniform preconditioner.

5.2.2 Constructing a practical $D_{\mathcal{T}}$

To avoid the preceding problems, we shall construct $D_{\mathcal{T}} \in \mathcal{L}\mathrm{is}(\mathscr{S}_{\mathcal{T},-s}, (\mathscr{S}_{\mathcal{T},s})')$ with a diagonal matrix representation. To this end, we require some auxiliary

space $\widetilde{\mathscr{S}_{\mathcal{T}}} \subset H^1(\Omega)$ equipped with a local basis $\widetilde{\Phi}_{\mathcal{T}}$ that is $L_2(\Omega)$-biorthogonal to $\Phi_{\mathcal{T}}$ and that has 'approximation properties'. To be precise, let $\widetilde{\Phi}_{\mathcal{T}} := \{\widetilde{\phi}_\nu \in H^1(\Omega) : \nu \in N_{\mathcal{T}}\}$ be some collection that satisfies:

$$\begin{aligned} \langle \widetilde{\phi}_\nu, \phi_{\nu'} \rangle_{L_2(\Omega)} = \delta_{\nu\nu'} \langle \mathbb{1}, \phi_\nu \rangle_{L_2(\Omega)}, \quad \sum_{\nu \in N_{\mathcal{T}}} \widetilde{\phi}_\nu = \mathbb{1}_\Omega, \\ \|\widetilde{\phi}_\nu\|_{H^k(\Omega)} \lesssim \|\phi_\nu\|_{H^k(\Omega)} \quad \big(k \in \{0,1\}\big), \quad \operatorname{supp} \widetilde{\phi}_\nu \subseteq \operatorname{supp} \phi_\nu.^1 \end{aligned} \tag{5.2}$$

We will take $D_{\mathcal{T}} := I'_{\mathcal{T}} \widetilde{D}_{\mathcal{T}}$ with $\widetilde{D}_{\mathcal{T}}$ and $I_{\mathcal{T}}$ being defined and analyzed in the next two theorems.

Theorem 5.2.2. *The operator $\widetilde{D}_{\mathcal{T}} \colon \mathscr{S}_{\mathcal{T},-s} \to (\widetilde{\mathscr{S}_{\mathcal{T},s}})'$, defined by $(\widetilde{D}_{\mathcal{T}} u)(v) := \langle u, v \rangle_{L_2(\Omega)}$, satisfies $\widetilde{D}_{\mathcal{T}} \in \mathcal{L}\mathrm{is}(\mathscr{S}_{\mathcal{T},-s}, (\widetilde{\mathscr{S}_{\mathcal{T},s}})')$ uniformly in $\mathcal{T} \in \mathbb{T}$.*

Proof. This proof largely follows Sect. 2.3, but because here we consider a Sobolev space $H^s(\Omega)$ that does not incorporate homogeneous boundary conditions, it allows for an easier proof.

From the assumptions (5.2), it follows that the biorthogonal 'Fortin' projector $P_{\mathcal{T}} \colon L_2(\Omega) \to H^1(\Omega)$ onto $\widetilde{\mathscr{S}_{\mathcal{T}}}$ with $\operatorname{ran}(\mathrm{Id} - P_{\mathcal{T}}) = \mathscr{S}_{\mathcal{T}}^{\perp_{L_2(\Omega)}}$ exists, and is given by

$$P_{\mathcal{T}} u = \sum_{\nu \in N_{\mathcal{T}}} \frac{\langle u, \phi_\nu \rangle_{L_2(\Omega)}}{\langle \widetilde{\phi}_\nu, \phi_\nu \rangle_{L_2(\Omega)}} \widetilde{\phi}_\nu.$$

Let $T \in \mathcal{T}$, by (5.2) and the fact that $\langle \mathbb{1}, \phi_\nu \rangle_{L_2(\Omega)} \eqsim \|\phi_\nu\|^2_{L_2(\Omega)}$, we find for $k \in \{0,1\}$

$$\|P_{\mathcal{T}} u\|_{H^k(T)} \lesssim \sum_{\nu \in N_T} \frac{\|\widetilde{\phi}_\nu\|_{H^k(T)}}{\|\phi_\nu\|_{L_2(\Omega)}} \|u\|_{L_2(\operatorname{supp} \phi_\nu)} \lesssim h_T^{-k} \|u\|_{L_2(\omega_{\mathcal{T}}(T))}, \tag{5.3}$$

with $\omega_{\mathcal{T}}(T) := \bigcup_{\{\nu \in N_T\}} \operatorname{supp} \phi_\nu$. This shows $\sup_{\mathcal{T} \in \mathbb{T}} \|P_{\mathcal{T}}\|_{\mathcal{L}(L_2(\Omega), L_2(\Omega))} < \infty$.

From the above inequality, and $\sum_{\nu \in N_{\mathcal{T}}} \widetilde{\phi}_\nu = \mathbb{1}$, we deduce that

$$\begin{aligned} \|(\mathrm{Id} - P_{\mathcal{T}}) u\|_{H^1(T)} &= \inf_{p \in \mathcal{P}_0} \|(\mathrm{Id} - P_{\mathcal{T}})(u - p)\|_{H^1(T)} \\ &\lesssim \inf_{p \in \mathcal{P}_0} \|u - p\|_{H^1(T)} + h_T^{-1} \|u - p\|_{L_2(\omega_{\mathcal{T}}(T))} \\ &\lesssim \inf_{p \in \mathcal{P}_0} h_T^{-1} \|u - p\|_{L_2(\omega_{\mathcal{T}}(T))} + |u|_{H^1(T)} \\ &\lesssim |u|_{H^1(\omega_{\mathcal{T}}(T))}, \end{aligned}$$

with the last step following from the Bramble-Hilbert lemma. We conclude that $\sup_{\mathcal{T} \in \mathbb{T}} \|P_{\mathcal{T}}\|_{\mathcal{L}(H^1(\Omega), H^1(\Omega))} < \infty$, and consequently by the Riesz-Thorin

[1] This last condition can be replaced by $\widetilde{\phi}_\nu$ having (uniformly) local support.

interpolation theorem, that

$$\sup_{\mathcal{T}\in\mathbb{T}} \|P_{\mathcal{T}}\|_{\mathcal{L}(H^s(\Omega),H^s(\Omega))} < \infty.$$

This latter property guarantees that $\widetilde{D}_{\mathcal{T}}$ is uniformly boundedly invertible:

$$\begin{aligned}
\|\widetilde{D}_{\mathcal{T}}\|_{\mathcal{L}(\mathscr{S}_{\mathcal{T},-s},(\widetilde{\mathscr{S}}_{\mathcal{T},s})')} &= \sup_{0\neq u\in\mathscr{S}_{\mathcal{T},-s}} \sup_{0\neq v\in\widetilde{\mathscr{S}}_{\mathcal{T},s}} \frac{\langle u,v\rangle_{L_2(\Omega)}}{\|u\|_{H^{-s}(\Omega)}\|v\|_{H^s(\Omega)}} \leq 1,\\
\|\widetilde{D}_{\mathcal{T}}^{-1}\|^{-1}_{\mathcal{L}((\widetilde{\mathscr{S}}_{\mathcal{T},s})',\mathscr{S}_{\mathcal{T},-s})} &= \inf_{0\neq u\in\mathscr{S}_{\mathcal{T},-s}} \sup_{0\neq v\in\widetilde{\mathscr{S}}_{\mathcal{T},s}} \frac{\langle u,v\rangle_{L_2(\Omega)}}{\|u\|_{H^{-s}(\Omega)}\|v\|_{H^s(\Omega)}}\\
&= \inf_{0\neq u\in\mathscr{S}_{\mathcal{T},-s}} \sup_{0\neq v\in H^s(\Omega)} \frac{\langle u,v\rangle_{L_2(\Omega)}}{\|u\|_{H^{-s}(\Omega)}\|P_{\mathcal{T}}v\|_{H^s(\Omega)}}\\
&\geq \|P_{\mathcal{T}}\|^{-1}_{\mathcal{L}(H^s(\Omega),H^s(\Omega))}. \qquad \square
\end{aligned}$$

Theorem 5.2.3. *For $I_{\mathcal{T}}\colon \mathscr{S}_{\mathcal{T},s} \to \widetilde{\mathscr{S}}_{\mathcal{T},s}$ being the bijection given by $I_{\mathcal{T}}\phi_\nu = \widetilde{\phi}_\nu$ ($\nu\in N_{\mathcal{T}}$), it holds that $I_{\mathcal{T}} \in \mathcal{L}\mathrm{is}(\mathscr{S}_{\mathcal{T},s},\widetilde{\mathscr{S}}_{\mathcal{T},s})$ uniformly in $\mathcal{T}\in\mathbb{T}$.*

Proof. Note that we may write

$$I_{\mathcal{T}}u = \sum_{\nu\in N_{\mathcal{T}}} \frac{\langle u,\widetilde{\phi}_\nu\rangle_{L_2(\Omega)}}{\langle \phi_\nu,\widetilde{\phi}_\nu\rangle_{L_2(\Omega)}}\widetilde{\phi}_\nu \quad\text{and}\quad I_{\mathcal{T}}^{-1}u = \sum_{\nu\in N_{\mathcal{T}}} \frac{\langle u,\phi_\nu\rangle_{L_2(\Omega)}}{\langle \widetilde{\phi}_\nu,\phi_\nu\rangle_{L_2(\Omega)}}\phi_\nu.$$

Equivalently to (5.3), we see for $k\in\{0,1\}$ that

$$\|I_{\mathcal{T}}u\|_{H^k(T)} \lesssim \sum_{\nu\in N_T} \frac{\|\widetilde{\phi}_\nu\|_{H^k(T)}\|\widetilde{\phi}_\nu\|_{L_2(\Omega)}}{\|\phi_\nu\|^2_{L_2(\Omega)}}\|u\|_{L_2(\operatorname{supp}\phi_\nu)} \lesssim h_T^{-k}\|u\|_{L_2(\omega_{\mathcal{T}}(T))}.$$

Following the same arguments as in the proof of Theorem 5.2.2, using that $I_{\mathcal{T}}\mathbb{1} = \mathbb{1}$, then reveals that $I_{\mathcal{T}}$ is uniformly bounded. Uniformly boundedness of $I_{\mathcal{T}}^{-1}$ follows similarly. $\square$

As announced earlier, we define $D_{\mathcal{T}} \in \mathcal{L}(\mathscr{S}_{\mathcal{T},-s},(\mathscr{S}_{\mathcal{T},s})')$ by $D_{\mathcal{T}} := I'_{\mathcal{T}}\widetilde{D}_{\mathcal{T}}$, so $(D_{\mathcal{T}}u)(v) := \langle u, I_{\mathcal{T}}v\rangle_{L_2(\Omega)}$ ($u,v\in\mathscr{S}_{\mathcal{T}}$). Combining the previous theorems gives the following corollary.

Corollary 5.2.4. *The operator $D_{\mathcal{T}}$ is in $\mathcal{L}\mathrm{is}(\mathscr{S}_{\mathcal{T},-s},(\mathscr{S}_{\mathcal{T},s})')$ uniformly in $\mathcal{T}\in\mathbb{T}$.*

Remark 5.2.5. The matrix representation of $D_{\mathcal{T}}$ w.r.t. $\Phi_{\mathcal{T}}$ given by

$$\boldsymbol{D}_{\mathcal{T}} = \langle \Phi_{\mathcal{T}}, I_{\mathcal{T}}\Phi_{\mathcal{T}}\rangle_{L_2(\Omega)} = \operatorname{diag}\{\langle \mathbb{1},\phi_\nu\rangle_{L_2(\Omega)} : \nu\in N_{\mathcal{T}}\},$$

which is *diagonal* and therefore easily invertible. The matrix $\boldsymbol{D}_{\mathcal{T}}$ is known as the *lumped mass matrix*.

Remark 5.2.6. The operator $D_{\mathcal{T}}$ depends merely on the *existence* of a biorthogonal basis $\widetilde{\Phi}_{\mathcal{T}}$ that satisfies (5.2). Indeed, this basis does not appear in the implementation of $\boldsymbol{D}_{\mathcal{T}}$.

A possible construction of $\widetilde{\Phi}_{\mathcal{T}}$ can be given using techniques from Chapter 2. Consider some collection of local 'bubble' functions $\Theta_{\mathcal{T}} = \{\theta_\nu \in H^1(\Omega) : \nu \in N_{\mathcal{T}}\}$ that satisfy: $\big|\langle \theta_\nu, \phi_{\nu'}\rangle_{L_2(\Omega)}\big| \eqsim \delta_{\nu\nu'}\|\phi_\nu\|^2_{L_2(\Omega)}$, $\|\theta_\nu\|_{H^k(\Omega)} \lesssim \|\phi_\nu\|_{H^k(\Omega)}$ $(k \in \{0,1\})$, and $\operatorname{supp}\theta_\nu \subseteq \operatorname{supp}\phi_\nu$. Existence of such a collection can be shown by a construction on a reference d-simplex, and then using an affine bijection to transfer it to general elements, see Sect. 2.4. A suitable $\widetilde{\Phi}_{\mathcal{T}}$ that satisfies (5.2) is then given by

$$\widetilde{\phi}_\nu := \phi_\nu + \frac{\langle \mathbb{1}, \phi_\nu\rangle_{L_2(\Omega)}}{\langle \theta_\nu, \phi_\nu\rangle_{L_2(\Omega)}}\theta_\nu - \sum_{\nu' \in N_{\mathcal{T}}} \frac{\langle \phi_\nu, \phi_{\nu'}\rangle_{L_2(\Omega)}}{\langle \theta_{\nu'}, \phi_{\nu'}\rangle_{L_2(\Omega)}}\theta_{\nu'}.$$

We emphasize that the construction of a uniform preconditioner outlined in this subsection does not assume some sufficiently mild grading of the meshes.

Implementation

Taking $\Phi_{\mathcal{T}}$ as basis for both $\mathscr{S}_{\mathcal{T},-s}$ and $\mathscr{S}_{\mathcal{T},s}$, the *matrix representation* of the preconditioned systems from (5.1) read as

$$\boldsymbol{D}_{\mathcal{T}}^{-1}\boldsymbol{B}_{\mathcal{T}}\boldsymbol{D}_{\mathcal{T}}^{-\top}\boldsymbol{A}_{\mathcal{T}} \quad \text{and} \quad \boldsymbol{D}_{\mathcal{T}}^{-\top}\boldsymbol{A}_{\mathcal{T}}\boldsymbol{D}_{\mathcal{T}}^{-1}\boldsymbol{B}_{\mathcal{T}},$$

where

$$\begin{aligned}\boldsymbol{A}_{\mathcal{T}} &:= (A_{\mathcal{T}}\Phi_{\mathcal{T}})(\Phi_{\mathcal{T}}), \quad \boldsymbol{B}_{\mathcal{T}} := (B_{\mathcal{T}}\Phi_{\mathcal{T}})(\Phi_{\mathcal{T}}),\\ \boldsymbol{D}_{\mathcal{T}} = \boldsymbol{D}_{\mathcal{T}}^{\top} &:= (D_{\mathcal{T}}\Phi_{\mathcal{T}})(\Phi_{\mathcal{T}}) = \operatorname{diag}\{\langle \mathbb{1}, \phi_\nu\rangle_{L_2(\Omega)} : \nu \in N_{\mathcal{T}}\}.\end{aligned}$$

Alternatively, we could equip the spaces with the *scaled* nodal basis $\breve{\Phi}_{\mathcal{T}} := \boldsymbol{D}_{\mathcal{T}}^{-\frac{1}{2}}\Phi_{\mathcal{T}}$, so that the $L_2(\Omega)$-norm of any basis function is proportional to 1, yielding

$$\begin{aligned}\breve{\boldsymbol{A}}_{\mathcal{T}} &:= (A_{\mathcal{T}}\breve{\Phi}_{\mathcal{T}})(\breve{\Phi}_{\mathcal{T}}) = (\boldsymbol{D}_{\mathcal{T}}^{-\frac{1}{2}})^{\top}\boldsymbol{A}_{\mathcal{T}}\boldsymbol{D}_{\mathcal{T}}^{-\frac{1}{2}},\\ \breve{\boldsymbol{B}}_{\mathcal{T}} &:= (B_{\mathcal{T}}\breve{\Phi}_{\mathcal{T}})(\breve{\Phi}_{\mathcal{T}}) = (\boldsymbol{D}_{\mathcal{T}}^{-\frac{1}{2}})^{\top}\boldsymbol{B}_{\mathcal{T}}\boldsymbol{D}_{\mathcal{T}}^{-\frac{1}{2}},\\ \breve{\boldsymbol{D}}_{\mathcal{T}} &:= (D_{\mathcal{T}}\breve{\Phi}_{\mathcal{T}})(\breve{\Phi}_{\mathcal{T}}) = (\boldsymbol{D}_{\mathcal{T}}^{-\frac{1}{2}})^{\top}\boldsymbol{D}_{\mathcal{T}}\boldsymbol{D}_{\mathcal{T}}^{-\frac{1}{2}} = \mathbf{Id},\end{aligned}$$

showing that $\breve{\boldsymbol{B}}_{\mathcal{T}}$ is a uniform preconditioner for $\breve{\boldsymbol{A}}_{\mathcal{T}}$ (and vice versa). To the best of our knowledge, so far this most easy form of operator preconditioning, where the stiffness matrix of some operator w.r.t. some basis is preconditioned by stiffness matrix of an opposite order operator w.r.t. the same basis, has not been shown to be optimal.

5.3 Manifold case

Let Γ be a compact d-dimensional Lipschitz, piecewise smooth manifold in $\mathbb{R}^{d'}$ for some $d' \geq d$ without boundary $\partial\Gamma$. For $s \in [0,1]$, we consider the Sobolev spaces

$$H^s(\Gamma) := [L_2(\Gamma), H^1(\Gamma)]_{s,2}, \quad H^{-s}(\Gamma) := H^s(\Gamma)'.$$

We assume that Γ is given as the closure of the disjoint union of $\cup_{i=1}^p \chi_i(\Omega_i)$, with, for $1 \leq i \leq p$, $\chi_i \colon \mathbb{R}^d \to \mathbb{R}^{d'}$ being some smooth regular parametrization, and $\Omega_i \subset \mathbb{R}^d$ an open polytope. W.l.o.g. assuming that for $i \neq j$, $\overline{\Omega}_i \cap \overline{\Omega}_j = \emptyset$, we define

$$\chi \colon \Omega := \cup_{i=1}^p \Omega_i \to \cup_{i=1}^p \chi_i(\Omega_i) \text{ by } \chi|_{\Omega_i} = \chi_i.$$

Let $\mathbb{T}$ be a family of conforming partitions $\mathcal{T}$ of Γ into 'panels' such that, for $1 \leq i \leq p$, $\chi^{-1}(\mathcal{T}) \cap \Omega_i$ is a uniformly shape regular conforming partition of Ω_i into d-simplices (that for $d = 1$ satisfies a uniform K-mesh property).

Fix $\ell > 0$, we set

$$\mathscr{S}_\mathcal{T} := \{u \in H^1(\Gamma) \colon u \circ \chi|_{\chi^{-1}(T)} \in \mathcal{P}_\ell \ (T \in \mathcal{T})\},$$

equipped with the canonical nodal basis $\Phi_\mathcal{T} = \{\phi_\nu : \nu \in N_\mathcal{T}\}$.

For construction of an operator $D_\mathcal{T} \in \mathcal{L}\text{is}(\mathscr{S}_{\mathcal{T},-s}, (\mathscr{S}_{\mathcal{T},s})')$ one can proceed as in the domain case. A suitable collection $\widetilde{\Phi}_\mathcal{T}$ that is $L_2(\Gamma)$-biorthogonal to $\Phi_\mathcal{T}$ exists. Moreover, the analysis from the domain case applies verbatim by only changing $\langle\cdot,\cdot\rangle_{L_2(\Omega)}$ into $\langle\cdot,\cdot,\rangle_{L_2(\Gamma)}$. A *hidden problem*, however, is that the computation of $\boldsymbol{D}_\mathcal{T} = \operatorname{diag}\{\langle \mathbb{1}, \phi_\nu\rangle_{L_2(\Gamma)} : \nu \in N_\mathcal{T}\}$ involves integrals over Γ that generally have to be approximated using numerical quadrature.

In Sect. 2.3.2 we solved this issue by defining an additional 'mesh-dependent' scalar product

$$\langle u, v\rangle_\mathcal{T} := \sum_{T\in\mathcal{T}} \frac{|T|}{|\chi^{-1}(T)|} \int_{\chi^{-1}(T)} u(\chi(x))v(\chi(x))dx.$$

This is constructed by replacing on each $\chi^{-1}(T)$, the Jacobian $|\partial\chi|$ by its average $\frac{|T|}{|\chi^{-1}(T)|}$ over $\chi^{-1}(T)$.

By considering $\widetilde{\Phi}_\mathcal{T}$ that is biorthogonal to $\Phi_\mathcal{T}$ with respect to $\langle\cdot,\cdot\rangle_\mathcal{T}$, and the linear bijection $I_\mathcal{T}$ given by $I_\mathcal{T}\phi_\nu = \widetilde{\phi}_\nu$, one is able to show that the operator $D_\mathcal{T}$ defined as $(D_\mathcal{T}u)(v) := \langle u, I_\mathcal{T}v\rangle_\mathcal{T}$ satisfies the necessary requirements. For details we refer to Chapter 2. The resulting matrix representation of $D_\mathcal{T}$ w.r.t. $\Phi_\mathcal{T}$ is then given by $\boldsymbol{D}_\mathcal{T} = \operatorname{diag}\{\langle \mathbb{1}, \phi_\nu\rangle_\mathcal{T} \colon \nu \in N_\mathcal{T}\}$.

5.4 Numerical experiments

Let $\Gamma = \partial[0,1]^3 \subset \mathbb{R}^3$ be the two-dimensional manifold without boundary given as the boundary of the unit cube, $s = \frac{1}{2}$, and $\mathscr{S}_\mathcal{T}$ the space of continuous piecewise polynomials of degree ℓ w.r.t. a partition $\mathcal{T}$. We will

evaluate preconditioning of the discretized Single Layer Integral operator $A_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}_{\mathcal{T},-s}, (\mathscr{S}_{\mathcal{T},-s})')$ and an (essentially) discretized Hypersingular Integral operator $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}_{\mathcal{T},s}, (\mathscr{S}_{\mathcal{T},s})')$.

The Hypersingular Integral operator $\tilde{B} \in \mathcal{L}(H^{\frac{1}{2}}(\Gamma), H^{-\frac{1}{2}}(\Gamma))$, is only-semi coercive, but solving $\tilde{B}u = f$ for f with $f(\mathbb{1}) = 0$ is equivalent to solving $Bu = f$ with B given by $(Bu)(v) = (\tilde{B}u)(v) + \alpha\langle u, \mathbb{1}\rangle_{L_2(\Gamma)}\langle v, \mathbb{1}\rangle_{L_2(\Gamma)}$, for some fixed $\alpha > 0$. This operator B is in $\mathcal{L}\mathrm{is}_c(H^{\frac{1}{2}}(\Gamma), H^{-\frac{1}{2}}(\Gamma))$, and we shall consider discretizations $B_{\mathcal{T}} \in \mathcal{L}\mathrm{is}_c(\mathscr{S}_{\mathcal{T},s}, (\mathscr{S}_{\mathcal{T},s})')$ of B. We found $\alpha = 0.05$ to give good results in our examples.

Equipping both $\mathscr{S}_{\mathcal{T},s}$ and $\mathscr{S}_{\mathcal{T},-s}$ with the standard nodal basis $\Phi_{\mathcal{T}} = \{\phi_\nu : \nu \in N_{\mathcal{T}}\}$, the matrix representations of the preconditioned systems from Sect. 5.2.2 read as

$$\boldsymbol{D}_{\mathcal{T}}^{-1}\boldsymbol{B}_{\mathcal{T}}\boldsymbol{D}_{\mathcal{T}}^{-\top}\boldsymbol{A}_{\mathcal{T}} \quad \text{and} \quad \boldsymbol{D}_{\mathcal{T}}^{-\top}\boldsymbol{A}_{\mathcal{T}}\boldsymbol{D}_{\mathcal{T}}^{-1}\boldsymbol{B}_{\mathcal{T}},$$

for $\boldsymbol{D}_{\mathcal{T}} = \mathrm{diag}\{\langle \mathbb{1}, \phi_\nu\rangle_{L_2(\Gamma)} : \nu \in N_{\mathcal{T}}\}$, $\boldsymbol{A}_{\mathcal{T}} = (A_{\mathcal{T}}\Phi_{\mathcal{T}})(\Phi_{\mathcal{T}})$ and $\boldsymbol{B}_{\mathcal{T}} := (B_{\mathcal{T}}\Phi_{\mathcal{T}})(\Phi_{\mathcal{T}})$.

We calculated (spectral) condition numbers of these preconditioned systems, where this condition number is given by $\kappa_S(\boldsymbol{X}) := \rho(\boldsymbol{X})\rho(\boldsymbol{X}^{-1})$ with $\rho(\cdot)$ denoting the spectral radius. Note that the condition numbers of the preconditioned systems coincide, i.e.,

$$\kappa_S(\boldsymbol{D}_{\mathcal{T}}^{-1}\boldsymbol{B}_{\mathcal{T}}\boldsymbol{D}_{\mathcal{T}}^{-\top}\boldsymbol{A}_{\mathcal{T}}) = \kappa_S(\boldsymbol{D}_{\mathcal{T}}^{-\top}\boldsymbol{A}_{\mathcal{T}}\boldsymbol{D}_{\mathcal{T}}^{-1}\boldsymbol{B}_{\mathcal{T}}),$$

so we may restrict ourselves to results for preconditioning of $\boldsymbol{A}_{\mathcal{T}}$.

We used the BEM++ software package [ŚBA+15] to approximate the matrix representation of $\boldsymbol{A}_{\mathcal{T}}$ and $\boldsymbol{B}_{\mathcal{T}}$ by hierarchical matrices based on adaptive cross approximation [Hac99, Beb00].

As initial partition $\mathcal{T}_\perp = \mathcal{T}_1$ of Γ we take a conforming partition consisting of 2 triangles per side, so 12 triangles in total, with an assignment of the newest vertices that satisfies the so-called matching condition. We let $\mathbb{T}$ be the sequence $\{\mathcal{T}_k\}_{k\geq 1}$ where the (conforming) partition $\mathcal{T}_k$ is found by applying both uniform and local refinements. To be precise, $\mathcal{T}_k$ is constructed by first applying k uniform bisections to $\mathcal{T}_\perp$, and then $4k$ local refinements by repeatedly applying NVB to all triangles that touch a corner of the cube. These partitions share both the difficulties of locally refined partitions (the presence of triangles with strongly different sizes) and that of uniform partitions (the diagonally scaled stiffness matrix has a condition number $\gtrsim 2^{k|s|}$).

5.4.1 Comparison preconditioners

Write $\boldsymbol{G}_{\mathcal{T}}^D := \boldsymbol{D}_{\mathcal{T}}^{-1}\boldsymbol{B}_{\mathcal{T}}\boldsymbol{D}_{\mathcal{T}}^{-\top}$ for the preconditioner constructed in Sect. 5.2.2. We will compare this with the preconditioner described in Sect. 5.2.1, for which the matrix representation is given by $\boldsymbol{G}_{\mathcal{T}}^M := \boldsymbol{M}_{\mathcal{T}}^{-1}\boldsymbol{B}_{\mathcal{T}}\boldsymbol{M}_{\mathcal{T}}^{-\top}$ with mass matrix $\boldsymbol{M}_{\mathcal{T}} = \boldsymbol{M}_{\mathcal{T}}^\top = \langle \Phi_{\mathcal{T}}, \Phi_{\mathcal{T}}\rangle_{L_2(\Gamma)}$. Because our partitions of the two-dimensional

Table 5.1. Spectral condition numbers, $\kappa_S(\boldsymbol{G}_{\mathcal{T}}^{\circ}\boldsymbol{A}_{\mathcal{T}})$ for $\circ \in \{D, M\}$, of the preconditioned Single Layer system discretized on $\{\mathcal{T}_k\}_{k\geq 1}$, by continuous piecewise linears ($\ell = 1$) in the middle columns and discretized by continuous piecewise cubics ($\ell = 3$) in the right columns. Here $\boldsymbol{G}_{\mathcal{T}}^{D}$ is the preconditioner introduced in Sect. 5.2.2, whereas $\boldsymbol{G}_{\mathcal{T}}^{M}$ is the preconditioner described in Sect. 5.2.1 whose application requires an application of $\boldsymbol{M}_{\mathcal{T}}^{-1}$, which we implemented using an LU-factorization.

Partition $\mathcal{T}$		Linears ($\ell = 1$)			Cubics ($\ell = 3$)		
h_{min}	h_{max}	dofs	$\boldsymbol{G}_{\mathcal{T}}^{D}\boldsymbol{A}_{\mathcal{T}}$	$\boldsymbol{G}_{\mathcal{T}}^{M}\boldsymbol{A}_{\mathcal{T}}$	dofs	$\boldsymbol{G}_{\mathcal{T}}^{D}\boldsymbol{A}_{\mathcal{T}}$	$\boldsymbol{G}_{\mathcal{T}}^{M}\boldsymbol{A}_{\mathcal{T}}$
$1.4 \cdot 10^{0}$	$1.4 \cdot 10^{0}$	8	16.2	1.20	56	90.5	1.68
$4.4 \cdot 10^{-2}$	$5.0 \cdot 10^{-01}$	218	14.9	1.91	1946	87.9	2.08
$1.3 \cdot 10^{-3}$	$3.5 \cdot 10^{-01}$	482	14.7	2.04	4322	86.1	2.17
$4.3 \cdot 10^{-5}$	$1.7 \cdot 10^{-01}$	962	14.7	2.10	8642	85.0	2.21
$1.3 \cdot 10^{-6}$	$8.8 \cdot 10^{-02}$	2306	15.4	2.14	20738	84.9	2.23
$4.2 \cdot 10^{-8}$	$4.4 \cdot 10^{-02}$	7106	15.6	2.16	63938	84.9	2.24
$1.3 \cdot 10^{-9}$	$2.2 \cdot 10^{-02}$	25730	15.8	2.17	231554	84.8	2.25
$4.1 \cdot 10^{-11}$	$1.1 \cdot 10^{-02}$	99650	15.8	2.17	896834	84.7	2.25

surface are created with NVB, we know that also the latter preconditioner provides uniformly bounded condition numbers. In contrast to $\boldsymbol{D}_{\mathcal{T}}^{-1}$, the inverse $\boldsymbol{M}_{\mathcal{T}}^{-1}$ cannot be evaluated in linear complexity. We implemented the application of $\boldsymbol{M}_{\mathcal{T}}^{-1}$ by computing an LU-factorization of $\boldsymbol{M}_{\mathcal{T}}$.

Table 5.1 compares the spectral condition numbers for the preconditioned Single Layer systems with trial spaces given by continuous piecewise linears and those by continuous piecewise cubics. The condition numbers $\kappa_S(\boldsymbol{G}_{\mathcal{T}}^{D}\boldsymbol{A}_{\mathcal{T}})$ are uniformly bounded, but quantitatively the condition numbers $\kappa_S(\boldsymbol{G}_{\mathcal{T}}^{M}\boldsymbol{A}_{\mathcal{T}})$ are better.

5.4.2 Improving the preconditioner quality

As observed in Table 5.1, the preconditioner $\boldsymbol{G}_{\mathcal{T}}^{M}$ appears to be of superior quality, but it has unfavourable computational complexity. It does suggest a way for improving $\boldsymbol{G}_{\mathcal{T}}^{D}$: by replacing $\boldsymbol{D}_{\mathcal{T}}^{-1}$ with a better approximation of $\boldsymbol{M}_{\mathcal{T}}^{-1}$, one may hope to improve the quality. To this end, we introduce damped (preconditioned) Richardson. Let $0 < \lambda_{-} \leq \lambda_{min}(\boldsymbol{D}_{\mathcal{T}}^{-1}\boldsymbol{M}_{\mathcal{T}})$, $\lambda_{max}(\boldsymbol{D}_{\mathcal{T}}^{-1}\boldsymbol{M}_{\mathcal{T}}) \leq \lambda_{+}$, $\boldsymbol{R}_{\mathcal{T}}^{(0)} := 0$ and for $k \geq 0$ define

$$\boldsymbol{R}_{\mathcal{T}}^{(k+1)} := \boldsymbol{R}_{\mathcal{T}}^{(k)} + \omega \boldsymbol{D}_{\mathcal{T}}^{-1}(\mathrm{Id} - \boldsymbol{M}_{\mathcal{T}}\boldsymbol{R}_{\mathcal{T}}^{(k)}), \quad \omega = \frac{2}{\lambda_{-} + \lambda_{+}},$$

being the result of k Richardson iterations. Correspondingly define

$$\boldsymbol{G}_{\mathcal{T}}^{(k)} := \boldsymbol{R}_{\mathcal{T}}^{(k)}\boldsymbol{B}_{\mathcal{T}}\boldsymbol{R}_{\mathcal{T}}^{(k)}. \tag{5.4}$$

TABLE 5.2. Spectral condition numbers $\kappa_S(\boldsymbol{G}_{\mathcal{T}}^{(k)}\boldsymbol{A}_{\mathcal{T}})$ with $\boldsymbol{G}_{\mathcal{T}}^{(k)}$ the preconditioner from (5.4) that incorporates k Richardson iterations. The systems are discretized by continuous piecewise linears in the left columns and discretized by continuous piecewise cubics in the right columns.

Linears ($\ell = 1$)				Cubics ($\ell = 3$)			
dofs	$k=2$	$k=4$	$k=6$	dofs	$k=2$	$k=4$	$k=6$
8	2.26	1.29	1.22	56	10.1	3.99	2.65
218	3.05	2.07	1.94	1946	8.96	3.57	2.52
482	3.53	2.28	2.08	4322	8.80	3.59	2.52
962	3.79	2.44	2.19	8642	8.63	3.59	2.52
2306	3.98	2.52	2.24	20738	8.54	3.59	2.52
7106	4.18	2.57	2.27	63938	8.54	3.59	2.52
25730	4.35	2.61	2.28	231554	8.54	3.59	2.52
99650	4.47	2.65	2.29	896834	8.54	3.59	2.52

It follows that $\boldsymbol{G}_{\mathcal{T}}^{(1)} = \boldsymbol{G}_{\mathcal{T}}^{D}$ and $\lim_{k\to\infty} \boldsymbol{G}_{\mathcal{T}}^{(k)} = \boldsymbol{G}_{\mathcal{T}}^{M}$. Although we have no proof, we suspect that $\boldsymbol{G}_{\mathcal{T}}^{(k)}$ provides a uniform preconditioner for $\boldsymbol{A}_{\mathcal{T}}$ due to the fact that $\boldsymbol{R}_{\mathcal{T}}^{(k)}$ approximates $\boldsymbol{M}_{\mathcal{T}}^{-1}$, while preserving constant functions, being a key ingredient in the proofs of Theorems 5.2.2 and 5.2.3.

Values for λ_- and λ_+ can be found by calculating the extremal eigenvalues of the corresponding preconditioned mass matrix on a reference simplex, see e.g. [Wat87]. For $\ell = 1$ this gives $\omega = \frac{2(d+2)}{d+3}$, whereas for $\ell = 3$ and $d = 2$ we computed $\omega = 0.836$.

Table 5.2 compares the condition numbers $\kappa_S(\boldsymbol{G}_{\mathcal{T}}^{(k)}\boldsymbol{A}_{\mathcal{T}})$ for $k \in \{2, 4, 6\}$. We see that a few Richardson iterations drastically improves our preconditioner, making its quality on par with that of $\boldsymbol{G}_{\mathcal{T}}^{M}$ while having a favourable linear application cost.

Finally, to show that one cannot simply use any (iterative) method for approximating $\boldsymbol{M}_{\mathcal{T}}^{-1}$, we consider the case where one approximates this inverse using a Jacobi preconditioner. The resulting preconditioner is then given by

$$\boldsymbol{G}_{\mathcal{T}}^{J} := (\operatorname{diag} \boldsymbol{M}_{\mathcal{T}})^{-1}\boldsymbol{B}_{\mathcal{T}}(\operatorname{diag} \boldsymbol{M}_{\mathcal{T}})^{-\top}. \tag{5.5}$$

Table 5.3 clearly displays that this is not a uniformly bounded preconditioner, which we assume is due to the fact that $(\operatorname{diag} \boldsymbol{M}_{\mathcal{T}})^{-1}$ does not preserve constant functions for $\ell > 1$.

5.5 Conclusion

Considering discretized opposite order operators $\boldsymbol{A}_{\mathcal{T}}$ and $\boldsymbol{B}_{\mathcal{T}}$ using the same ansatz space of continuous piecewise polynomial w.r.t. a possibly locally re-

Table 5.3. Spectral condition numbers $\kappa_S(\boldsymbol{G}_{\mathcal{T}}^J \boldsymbol{A}_{\mathcal{T}})$ with $\boldsymbol{G}_{\mathcal{T}}^J$ from (5.5), and systems discretized by continuous piecewise cubics ($\ell = 3$).

dofs	$\boldsymbol{G}_{\mathcal{T}}^J \boldsymbol{A}_{\mathcal{T}}$
56	62.6
1946	377.1
4322	495.6
8642	1016.9
20738	3067.8
63938	10928.3

fined partition $\mathcal{T}$, we consider matrices $\boldsymbol{D}_{\mathcal{T}}$ such that $\boldsymbol{D}_{\mathcal{T}}^{-1} \boldsymbol{B}_{\mathcal{T}} \boldsymbol{D}_{\mathcal{T}}^{-\top}$ is a uniform preconditioner for $\boldsymbol{A}_{\mathcal{T}}$, and $\boldsymbol{D}_{\mathcal{T}}^{-\top} \boldsymbol{A}_{\mathcal{T}} \boldsymbol{D}_{\mathcal{T}}^{-1}$ for $\boldsymbol{B}_{\mathcal{T}}$. The obvious choice for $\boldsymbol{D}_{\mathcal{T}}$ would be the mass matrix, however, it yields uniformly bounded condition numbers only under a mildly grading assumption on the mesh, and more importantly, it has the disadvantage that its inverse is dense. We proved that when taking $\boldsymbol{D}_{\mathcal{T}}$ as the lumped mass matrix the condition numbers are uniformly bounded, remarkably without a sufficiently mild grading assumption on the mesh, while obviously its inverse can be applied in linear cost.

In our experiments with locally refined meshes generated by Newest Vertex Bisection, the condition numbers with $\boldsymbol{D}_{\mathcal{T}}$ being the mass matrix are quantitatively better than those found with $\boldsymbol{D}_{\mathcal{T}}$ being the lumped mass matrix though. Constructing $\boldsymbol{D}_{\mathcal{T}}^{-1}$ as an approximation for the inverse mass matrix by a few preconditioned damped Richardson steps with the lumped mass matrix as a preconditioner, both the resulting matrix can be applied at linear cost and the observed condition numbers are essentially as good as with the inverse mass matrix.

Part II

Space-time methods for parabolic evolution equations

6 An adaptive method

6.1 Introduction

This chapter is about the adaptive numerical solution of parabolic evolution equations written in a simultaneous space-time variational formulation. In comparison to the usually applied time-marching schemes, simultaneous space-time solvers offer the following potential advantages:

- local, adaptive refinements simultaneous in space and time ([SY18, RS19, GS19]),
- quasi-best approximation from the selected trial space ('Cea's lemma') ([And13, LM17, SW21b]), being a necessary requirement for proving optimal rates for adaptive routines ([CS11, KSU16, RS19]),
- superior parallel performance ([DGVdZ18, NS19, HLNS19, vVW21a]),
- using the product structure of the space-time cylinder, sparse tensor product approximation ([GO07, CS11, KSU16, RS19]) which allows to solve the whole time evolution at a complexity of solving the corresponding stationary problem.

Other relevant publications on space-time solvers include [Ste15, LMN16, SZ20, Dev20, DS20].

In any case without applying sparse tensor product approximation, a disadvantage of the space-time approach is the larger memory consumption because instead of solving a sequence of PDEs on a d-dimensional space, one has to solve one PDE on a $(d+1)$-dimensional space. This disadvantage, however, disappears when one needs simultaneously the whole time evolution as for example with problems of optimal control ([GK11, BRU20]) or data-assimilation ([DSW21]).

6.1.1 Parabolic problem in a simultaneous space-time variational formulation

For some separable Hilbert spaces $V \hookrightarrow H$ with dense embedding (e.g. $H^1_0(\Omega)$ and $L_2(\Omega)$ for the model problem of the heat equation on a spatial domain

$\Omega \subset \mathbb{R}^d$), and a boundedly invertible $A(t) = A(t)' \colon V \to V'$ with $(A(t)\cdot)(\cdot) \eqsim \|\cdot\|_V^2$ (e.g. $(A(t)\eta)(\zeta) = \int_\Omega \nabla\eta \cdot \nabla\zeta \, d\mathbf{x}$), we consider

$$\begin{cases} \frac{du}{dt}(t) + A(t)u(t) &= g(t) \quad (t \in (0,T)), \\ u(0) &= u_0. \end{cases}$$

An application of a variational formulation of the PDE over space and time leads to an equation

$$\begin{bmatrix} B \\ \gamma_0 \end{bmatrix} u = \begin{bmatrix} g \\ u_0 \end{bmatrix} \tag{6.1}$$

where, with $X := L_2(I;V) \cap H^1(I;V')$ and $Y := L_2(I;V)$, the operator at the left hand side is boundedly invertible $X \to Y' \times H$.

6.1.2 Our previous work

In [CS11, RS19] we equipped X and Y with Riesz bases being tensor products of wavelet bases in space in time, and H with some spatial Riesz basis. Consequently, the equation (6.1) got an equivalent formulation as a bi-infinite well-posed matrix-vector equation $\begin{bmatrix} \mathbf{B} \\ \boldsymbol{\gamma}_0 \end{bmatrix} \mathbf{u} = \begin{bmatrix} \mathbf{g} \\ \mathbf{u}_0 \end{bmatrix}$ (actually, in [RS19], we considered a formulation of first order, and in [CS11] we used a variational formulation with essentially interchanged roles of X and Y, which however is irrelevant for the current discussion). To get a coercive bilinear form we formed normal equations to which we applied an adaptive wavelet scheme ([CDD01]). With such a scheme the norm of a sufficiently accurate approximation of the (infinite) residual vector of a current approximation is used as an a posteriori error estimator. The coefficients in modulus of this vector are applied as local error indictors in a bulk chasing (or Dörfler) marking procedure. The resulting adaptive algorithm converges at the best possible rate in linear computational complexity.

The goal of the current work is to investigate to what extent similar optimal theoretical results can be shown for finite element discretizations, whilst realizing a quantitatively superior implementation.

6.1.3 Least squares minimization

Without having Riesz bases for X and Y, already the step of first discretizing and then forming normal equations does not apply, and we reverse their order. A problem equivalent to (6.1) is to compute

$$u = \operatorname*{argmin}_{w \in X} \|Bw - g\|_{Y'}^2 + \|\gamma_0 w - u_0\|_H^2. \tag{6.2}$$

An obvious approach for the numerical approximation is to consider the minimization over finite dimensional subspaces X^δ of X, which however is not feasible because of the presence of the dual norm.

For trial spaces X^δ that are 'full' (or 'sparse') tensor products of finite element spaces in space and time, in [And13] it was shown how to construct corresponding test spaces $Y^\delta \subset Y$ of similar type and dimension, such that (X^δ, Y^δ) is uniformly *inf-sup stable* meaning that when the continuous dual norm $\|\cdot\|_{Y'}$ is replaced by the discrete dual norm $\|\cdot\|_{Y^{\delta\prime}}$, a minimization over X^δ yields a quasi-best approximation to u from X^δ. Such a family of trial spaces however does not allow to create a nested sequence of trial spaces by adaptive *local* refinements.

6.1.4 Family of inf-sup stable pairs of trial and test spaces

To construct an alternative, essentially larger family, let Σ be a wavelet Riesz basis for $L_2(0,T)$ that, after renormalization, is also a Riesz basis for $H^1(0,T)$. We equip this basis with a tree structure where every wavelet that is not on the coarsest level has a parent on the next coarser level. In space, we consider the collection of all linear finite element spaces that can be generated by conforming newest vertex bisection starting from an initial conforming partition of a polytopal Ω into d-simplices. The restriction to linear finite elements is not essential and is made for simplicity only. Now we consider trial spaces X^δ that are spanned by a number of wavelets each of them tensorized with a finite element space from the aforementioned collection. In order to be able to apply the arising system matrices in linear complexity, see [KS14] or Chapter 7, we impose the condition that if a wavelet tensorized with a finite element space is in the spanning set, then so is its parent wavelet tensorized with a finite element space that includes the former one.

The infinite collection of finite element spaces can be associated to a hierarchical 'basis' that can be equipped with a tree structure. Each hierarchical basis function, except those on the coarsest level, is associated to a node ν that was inserted as the midpoint of an edge connecting two nodes on the next coarser level, which nodes we call the parents of ν. With this definition there is a one-to-one correspondence between the finite element spaces from our collection and the spans of the sets of hierarchical basis functions that form trees. Consequently, our collection of trial spaces X^δ consists of the spans of sets of tensor products of wavelets-in-time and hierarchical basis functions-in-space which sets are *downwards closed*, also known as *lower*, in the sense that if a pair of a wavelet and a hierarchical basis function is in the set, then so are all its parents in time and space. Spaces from this collection can be 'locally' expanded by adding the span of a tensor product of a wavelet and hierarchical basis function one-by-one.

For this family of spaces X^δ we construct a corresponding family of spaces $Y^\delta \subset Y$ of similar type such that each pair (X^δ, Y^δ) is uniformly inf-sup stable,

with the dimension of Y^δ being proportional to that of X^δ. Furthermore, using the properties of the wavelets in time and by applying multigrid preconditioners in space we construct optimal preconditioners at X and Y-side which allow a fast solution of the discrete problems $\operatorname{argmin}_{w\in X^\delta} \|Bw-g\|^2_{Y^{\delta\prime}}+\|\gamma_0 w-u_0\|^2_H$.

6.1.5 Adaptive algorithm

Having fixed the family of trial spaces, it remains to develop an algorithm that selects a suitable, preferably quasi-optimal, nested sequence of spaces from the family adapted to the solution u of (6.2). The theory about adaptive (Ritz-) Galerkin approximations for such quadratic minimization problems is in a mature state. As noticed before, however, Galerkin approximations for (6.2) are not computable.

Therefore given X^δ, let $X^{\underline{\delta}} \supset X^\delta$ be such that saturation holds, i.e., for some constant $\zeta < 1$, it holds that $\inf_{w\in X^{\underline{\delta}}} \|u-w\|_X \leq \zeta \inf_{w\in X^\delta} \|u-w\|_X$. We now replace problem (6.2) by

$$u^{\delta\underline{\delta}} = \operatorname*{argmin}_{w\in X^{\underline{\delta}}} \|Bw-g\|^2_{Y^{\underline{\delta}\prime}} + \|\gamma_0 w-u_0\|^2_H, \tag{6.3}$$

where in the notation $u^{\delta\underline{\delta}}$ the first instance of $\underline{\delta}$ refers to the space $Y^{\underline{\delta}}$ and the second to the space $X^{\underline{\delta}}$. Its (computable) Galerkin approximation from X^δ is given by

$$u^{\delta\delta} = \operatorname*{argmin}_{w\in X^{\delta}} \|Bw-g\|^2_{Y^{\underline{\delta}\prime}} + \|\gamma_0 w-u_0\|^2_H.$$

By a standard adaptive procedure, described below, we expand X^δ to some $X^{\tilde{\delta}} \subseteq X^{\underline{\delta}}$ such that $u^{\underline{\delta}\tilde{\delta}}$ is closer to $u^{\underline{\delta}\underline{\delta}}$ than $u^{\underline{\delta}\delta}$. Next, we replace $Y^{\underline{\delta}}$ by $Y^{\tilde{\underline{\delta}}}$ (being the test space corresponding to $X^{\tilde{\underline{\delta}}}$) and repeat (i.e. consider (6.3) with $(\underline{\delta},\underline{\delta})$ reading as $(\tilde{\underline{\delta}},\tilde{\underline{\delta}})$, and improve its Galerkin approximation $u^{\tilde{\underline{\delta}}\tilde{\delta}}$ from $X^{\tilde{\delta}}$ by an adaptive enlargement of the latter space).

The adaptive expansion of the trial space X^δ to $X^{\tilde{\delta}}$ will be by the application of the usual solve-estimate-mark-refine paradigm, where the error indicators are the coefficients of the residual vector w.r.t. (modified) tensor product basis functions that were added to X^δ to create $X^{\underline{\delta}}$. In order for this collection of additional tensor product basis functions to be stable in X-norm, for this step we modify the hierarchical basis functions such that they get a vanishing moment, and therefore become closer to 'real' wavelets.

Under the aforementioned saturation assumption, we prove that the overall adaptive procedure produces an r-linearly converging sequence to the solution.

6.1.6 Numerical results

We tested the adaptive algorithm in several examples with a two-dimensional spatial domain. In all but one case, we observed a convergence rate equal to

$1/2$, being the best that can be expected in view of the piecewise polynomial degree of the trial functions *and* the tensor product construction, and for non-smooth solutions improving upon usual non-adaptive approximation. Only for the case where $u_0 = 1$ and homogenous Dirichlet boundary conditions are prescribed, the observed rate was reduced to 0.4. It is unknown whether or not this is the best non-linear approximation rate for our family of trial spaces.

Thanks to the use of optimal preconditioners and that of a carefully designed matrix-vector multiplication routine, which generalizes such a routine for sparse-grids introduced in [BZ96] to adaptive settings, we observe that throughout the whole execution of the adaptive loop the total runtime remains proportional to the current number of unknowns.

Recently in [FK21], see also [GS21], a first order system least squares (FOSLS) of second order parabolic PDEs was proven to be well-posed. This formulation has the very attractive property that the several components of the residual are all measured in L_2-norms. So other than with (6.2) there is no need to discretize a dual-norm, and so to guarantee an inf-sup condition. Minimization over any conforming trial space yields a quasi-best approximation from that space in the corresponding 'energy-norm'. This norm, however, is stronger than the norm on X. For the aforementioned example of a discontinuity between initial and boundary conditions, and with the application of continuous piecewise linear finite elements w.r.t. tetrahedral meshes of the space-time cylinder it results in a convergence rate of 0.07 for uniform refinements, which is not visibly improved using adaptive refinements.

6.1.7 Organization

This chapter is organized as follows: In Sect. 6.2 the well-posed space-time variational formulation of the parabolic problem is discussed, and in Sect. 6.3 we discuss its inf-sup stable discretisation. The adaptive solution procedure is presented in Sect. 6.4, and its convergence is proven. The construction of the trial and test spaces is detailed in Sect. 6.5, and optimal preconditioners are presented. In Sect. 6.6, the definition of the enlarged space $X^{\underline{\delta}}$ is given, and the construction of a stable basis of a stable complement space of X^δ in $X^{\underline{\delta}}$ is outlined. Numerical results are presented in Sect. 6.7, and a conclusion is formulated in Sect. 6.8.

6.1.8 Notations

In this work, by $C \lesssim D$ we will mean that C can be bounded by a multiple of D, independently of parameters which C and D may depend on. Obviously, $C \gtrsim D$ is defined as $D \lesssim C$, and $C \eqsim D$ as $C \lesssim D$ and $C \gtrsim D$.

For normed linear spaces E and F, by $\mathcal{L}(E,F)$ we will denote the normed linear space of bounded linear mappings $E \to F$, and by $\mathcal{L}\mathrm{is}(E,F)$ its subset of boundedly invertible linear mappings $E \to F$. We write $E \hookrightarrow F$ to denote

that E is continuously embedded into F. For simplicity only, we exclusively consider linear spaces over the scalar field $\mathbb{R}$.

6.2 Space-time formulations of a parabolic evolution problem

Let V, H be separable Hilbert spaces of functions on some "spatial domain" such that $V \hookrightarrow H$ with dense embedding. Identifying H with its dual, we obtain the Gelfand triple $V \hookrightarrow H \simeq H' \hookrightarrow V'$.

For a.e.

$$t \in I := (0, T),$$

let $a(t; \cdot, \cdot)$ denote a bilinear form on $V \times V$ such that for any $\eta, \zeta \in V$, $t \mapsto a(t; \eta, \zeta)$ is measurable on I, and such that for some $\varrho \in \mathbb{R}$, for a.e. $t \in I$,

$$|a(t; \eta, \zeta)| \lesssim \|\eta\|_V \|\zeta\|_V \quad (\eta, \zeta \in V) \quad (\textit{boundedness}), \tag{6.4}$$

$$a(t; \eta, \eta) + \varrho \langle \eta, \eta \rangle_H \gtrsim \|\eta\|_V^2 \quad (\eta \in V) \quad (\textit{Gårding inequality}). \tag{6.5}$$

With $A(t) \in \mathcal{L}\mathrm{is}(V, V')$ being defined by $(A(t)\eta)(\zeta) = a(t; \eta, \zeta)$, given a forcing function g and an initial value u_0, we are interested in solving the *parabolic initial value problem* to finding u such that

$$\begin{cases} \frac{du}{dt}(t) + A(t)u(t) = g(t) & (t \in I), \\ u(0) = u_0. \end{cases} \tag{6.6}$$

In a simultaneous space-time variational formulation, the parabolic PDE reads as finding u from a suitable space of functions of time and space such that

$$(Bw)(v) := \int_I \langle \tfrac{dw}{dt}(t), v(t) \rangle + a(t; w(t), v(t)) dt = \int_I \langle g(t), v(t) \rangle =: g(v)$$

for all v from another suitable space of functions of time and space. One possibility to enforce the initial condition is by testing it against additional test functions. A proof of the following result can be found in [SS09], cf. [DL92, Ch.XVIII, §3] and [Wlo82, Ch. IV, §26] for slightly different statements.

Theorem 6.2.1. *With $X := L_2(I; V) \cap H^1(I; V')$, $Y := L_2(I; V)$, under conditions* (6.4) *and* (6.5) *it holds that*

$$\begin{bmatrix} B \\ \gamma_0 \end{bmatrix} \in \mathcal{L}\mathrm{is}(X, Y' \times H),$$

where for $t \in \bar{I}$, $\gamma_t \colon u \mapsto u(t, \cdot)$ denotes the trace map. That is, assuming $g \in Y'$ and $u_0 \in H$, finding $u \in X$ such that

$$\begin{bmatrix} B \\ \gamma_0 \end{bmatrix} u = \begin{bmatrix} g \\ u_0 \end{bmatrix} \tag{6.7}$$

is a well-posed simultaneous space-time variational formulation of (6.6).

With $\tilde{u}(t) := u(t)e^{-\varrho t}$, (6.6) is equivalent to $\frac{d\tilde{u}}{dt}(t) + (A(t) + \varrho\mathrm{Id})\tilde{u}(t) = g(t)e^{-\varrho t}$ ($t \in I$), $\tilde{u}(0) = u_0$. Since $((A(t)+\varrho\mathrm{Id})\eta)(\eta) \gtrsim \|\eta\|_V^2$, w.l.o.g. we assume that (6.5) is valid for $\varrho = 0$, i.e., $a(t;\cdot,\cdot)$ is *coercive* uniformly for a.e. $t \in I$.

For simplicity, cf. discussion in Remark 6.3.5, additionally we assume that $a(t;\cdot,\cdot)$ is *symmetric*, and define $A = A' \in \mathcal{L}\mathrm{is}(Y, Y')$ by $(Aw)(v) = \int_I (A(t)w(t))v(t)\,dt$.

Because $\begin{bmatrix} A & 0 \\ 0 & \mathrm{Id} \end{bmatrix} \in \mathcal{L}\mathrm{is}(Y \times H, Y' \times H)$, an equivalent formulation of (6.7) as a self-adjoint saddle point equation reads as finding $(\mu, \sigma, u) \in Y \times H \times X$ (where $\mu = 0 = \sigma$) such that

$$\begin{bmatrix} A & 0 & B \\ 0 & \mathrm{Id} & \gamma_0 \\ B' & \gamma_0' & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \sigma \\ u \end{bmatrix} = \begin{bmatrix} g \\ u_0 \\ 0 \end{bmatrix}, \tag{6.8}$$

or equivalently

$$\begin{bmatrix} A & B \\ B' & -\gamma_0'\gamma_0 \end{bmatrix} \begin{bmatrix} \mu \\ u \end{bmatrix} = \begin{bmatrix} g \\ -\gamma_0' u_0 \end{bmatrix}, \tag{6.9}$$

or

$$\underbrace{(B'A^{-1}B + \gamma_0'\gamma_0)}_{S:=} u = \underbrace{B'A^{-1}g + \gamma_0' u_0}_{f:=}. \tag{6.10}$$

We equip Y and X with *'energy'-norms*

$$\|\cdot\|_Y^2 := (A\cdot)(\cdot), \quad \|\cdot\|_X^2 := \|\cdot\|_Y^2 + \|\partial_t \cdot\|_{Y'}^2 + \|\gamma_T \cdot\|_H^2,$$

which are equivalent to the canonical norms on Y and X. Notice that (6.8)–(6.10) are the Euler-Langrange equations that result from the minimization problem

$$u = \operatorname*{argmin}_{w \in X} \|Bw - f\|_{Y'}^2 + \|\gamma_0 w - u_0\|_H^2.$$

Lemma 6.2.2. *We have* $\|\cdot\|_X^2 = (S\cdot)(\cdot)$.

Proof. It holds that

$$\begin{aligned} \|w\|_X^2 &= \sup_{0 \neq v_1 \in Y} \frac{(Bw)(v_1)}{\|v_1\|_Y^2} + \|\gamma_0 w\|_H^2 = \sup_{0 \neq (v_1, v_2) \in Y \times H} \frac{((Bw)(v_1) + \langle \gamma_0 w, v_2 \rangle_H)^2}{\|v_1\|_Y^2 + \|v_2\|_H^2} \\ &= (Sw)(w), \end{aligned}$$

where the first equality can be found in e.g. [ESV17, Thm. 2.1], and, when realising that S is the Schur complement of the operator in (6.8), the last one in e.g. [KS08, Lemma 2.2]. □

6.3 Discretizations

6.3.1 Galerkin discretization of the Schur complement equation

Let $(X^\delta)_{\delta\in\Delta}$ be a collection of closed, e.g, finite dimensional, subspaces of X, so equipped with $\|\cdot\|_X$. We will specify such a family in Sect. 6.5-6.6.1. We define a *partial order* on Δ by

$$\delta \preceq \tilde{\delta} \Longleftrightarrow X^\delta \subseteq X^{\tilde{\delta}}.$$

For $\delta \in \Delta$, let $u_\delta \in X^\delta$ denote the *Galerkin approximation to the solution u of* (6.10), i.e., the solution of

$$(Su_\delta)(v) = f(v) \quad (v \in X^\delta), \tag{6.11}$$

being the best approximation to u from X^δ w.r.t. $\|\cdot\|_X$.

For proving convergence of an adaptive solution routine, as well as for a posteriori error estimation, we shall make the following assumption.

Assumption 6.3.1 (Saturation). *There exists a collection of subspaces $({}^\delta G\times {}^\delta U_0)_{\delta\in\Delta} \subseteq Y'\times H$, a mapping $\underline{\cdot}\colon \Delta\to\Delta\colon \delta\mapsto\underline{\delta}$ where $\underline{\delta}\succeq\delta$, and some fixed constant $\zeta<1$ such that for all $\delta\in\Delta$, assuming that $(g,u_0)\in {}^\delta G\times {}^\delta U_0$,*

$$\|u-u_{\underline{\delta}}\|_X \le \zeta\|u-u_\delta\|_X. \tag{6.12}$$

Remark 6.3.2. Notice that above assumption cannot be valid without a restriction on the right-hand side $f = B'A^{-1}g+\gamma_0'u_0 \in X'$. Indeed given any $X^\delta \subset X^{\underline{\delta}} \subsetneq X$, consider a non-zero $f\in X'$ that vanishes on $X^{\underline{\delta}}$. Then $u_\delta = u_{\underline{\delta}} = 0 \neq u$, meaning that (6.12) does not hold.

> *For the time being we will operate under the restrictive assumption that whenever we apply* (6.12) *(visible by the appearance of the constant ζ) we simply assume that $(g,u_0)\in {}^\delta G\times {}^\delta U_0$. Later, in Sect. 6.4.3, we will remove this assumption.*

The discretized problem from (6.11) only serves theoretical purposes. Indeed, since the Schur complement operator S contains the inverse of A, there is no way to determine u_δ exactly. The reason to introduce (6.11) is that S is an *elliptic* operator, so that for $\delta\preceq\tilde{\delta}$ we can make use of $\|u-u_{\tilde{\delta}}\|_X^2 = \|u-u_\delta\|_X^2 - \|u_{\tilde{\delta}}-u_\delta\|_X^2$, being a crucial tool for proving convergence of adaptive algorithms.

6.3.2 Uniformly stable Galerkin discretization of the saddle-point formulation

Our numerical approximations will be based on Galerkin discretizations of the saddle-point formulation (6.9). Let $(Y^\delta)_{\delta\in\Delta}$ be a collection of closed subspaces

of Y, so equipped with $\|\cdot\|_Y$, such that

$$X^\delta \subseteq Y^\delta \quad (\delta \in \Delta), \tag{6.13}$$

and

$$1 \geq \gamma_\Delta := \inf_{\delta\in\Delta} \inf_{0\neq w\in X^\delta} \sup_{0\neq v\in Y^\delta} \frac{(\partial_t w)(v)}{\|\partial_t w\|_{Y'}\|v\|_Y} > 0. \tag{6.14}$$

Notice that $1-\gamma_\Delta$ can be made arbitrarily small by selecting, for each $\delta\in\Delta$, Y^δ sufficiently large in relation to X^δ.

For $\delta, \hat\delta \in \Delta$ with $Y^{\hat\delta} \supseteq Y^\delta$, and $E_Y^{\hat\delta}$, E_X^δ denoting the embeddings $Y^{\hat\delta}\to Y$, $X^\delta \to X$, let $(\mu^{\hat\delta\delta}, u^{\hat\delta\delta}) \in Y^{\hat\delta}\times X^\delta$ be the solution of

$$\begin{bmatrix} {E_Y^{\hat\delta}}' A E_Y^{\hat\delta} & {E_Y^{\hat\delta}}' B E_X^\delta \\ {E_X^\delta}' B' E_Y^{\hat\delta} & -{E_X^\delta}'\gamma_0'\gamma_0 E_X^\delta \end{bmatrix} \begin{bmatrix} \mu^{\hat\delta\delta} \\ u^{\hat\delta\delta}\end{bmatrix} = \begin{bmatrix} {E_Y^{\hat\delta}}' g \\ -{E_X^\delta}'\gamma_0' u_0\end{bmatrix}, \tag{6.15}$$

or, equivalently,

$$\begin{aligned} &\underbrace{{E_X^\delta}'(B' E_Y^{\hat\delta}({E_Y^{\hat\delta}}' A E_Y^{\hat\delta})^{-1}{E_Y^{\hat\delta}}' B + \gamma_0'\gamma_0) E_X^\delta}_{S^{\hat\delta\delta}:=} u^{\hat\delta\delta} \\ &\qquad = \underbrace{{E_X^\delta}'(B' E_Y^{\hat\delta}({E_Y^{\hat\delta}}' A E_Y^{\hat\delta})^{-1}{E_Y^{\hat\delta}}' g + \gamma_0' u_0)}_{f^{\hat\delta\delta}:=}. \end{aligned} \tag{6.16}$$

Below we will see that (6.15)-(6.16) are uniquely solvable. Formulated in 'operator language', (6.15) is the Galerkin discretization of (6.9) on the closed subspace $Y^{\hat\delta}\times X^\delta \subseteq Y\times X$. Unless $Y^{\hat\delta} = Y$, it holds that $S^{\hat\delta\delta} \neq {E_X^\delta}' S E_X^\delta$ and $f^{\hat\delta\delta} \neq {E_X^\delta}' f$, and so generally $u^{\hat\delta\delta} \neq u_\delta$.

As we will see, however, for Y^δ, and thus $Y^{\hat\delta}$, 'large' in relation to X^δ, $u^{\hat\delta\delta}$ will be 'close' to u_δ. This will allow us to show that (r-linear) convergence of a sequence of Galerkin solutions u_δ of (6.11) implies (r-linear) convergence of the corresponding sequence $u^{\hat\delta\delta}$.

We equip X^δ with a family of 'energy' norms

$$\|w\|_{X^{\hat\delta\delta}}^2 := \|w\|_Y^2 + \sup_{0\neq v\in Y^{\hat\delta}} \frac{(\partial_t w)(v)^2}{\|v\|_Y^2} + \|\gamma_T w\|^2.$$

By definition of γ_Δ it holds that

$$\gamma_\Delta \|\cdot\|_X \leq \|\cdot\|_{X^{\hat\delta\delta}} \leq \|\cdot\|_X \quad \text{on } X^\delta. \tag{6.17}$$

As follows from [SW21b, Lemma 3.3], similar to Lemma 6.2.2 we have the following result.

Lemma 6.3.3. *Thanks to* (6.13) *(and $Y^\delta \subseteq Y^{\hat\delta}$), for $w \in X^\delta$ it holds that*

$$\|w\|^2_{X^{\hat\delta\delta}} = (S^{\hat\delta\delta}w)(w) = \sup_{0\neq v\in Y^{\hat\delta}} \frac{(Bw)(v)^2}{\|v\|^2_Y} + \|\gamma_0 w\|^2_H.$$

By using additionally (6.14) this result shows that $(S^{\hat\delta\delta}\cdot)(\cdot)$ is coercive on $X^\delta \times X^\delta$ so that (6.16), and thus (6.15), has a unique solution.

Moreover, we have the following result.

Theorem 6.3.4 ([SW21b, Thm. 3.7])**.** *Thanks to* (6.13) *(and $Y^\delta \subseteq Y^{\hat\delta}$) and* (6.14)*, it holds that*

$$\|u - u_\delta\|_X \leq \|u - u^{\hat\delta\delta}\|_X \leq \gamma_\Delta^{-1}\|u - u_\delta\|_X. \tag{6.18}$$

Remark 6.3.5. Without the assumption of $a(t;\cdot,\cdot)$ being symmetric, the operator A in (6.8), (6.9), (6.10), (6.15), (6.16), and in the definition of $\|\cdot\|_Y$ should be replaced by $A_s := \frac{1}{2}(A+A')$, whereas ∂_t in the definitions of $\|\cdot\|_X$, γ_Δ in (6.18), and $\|\cdot\|_{X^{\hat\delta\delta}}$ should be replaced by $\partial_t + A_a$, where $A_a := \frac{1}{2}(A - A')$. Then, as shown in [SW21a, Thm. 6.1], it holds that $\|u - u^{\hat\delta\delta}\|_X \leq \gamma_\Delta^{-2}\|u - u_\delta\|_X$.

Still without assuming that $a(t;\cdot,\cdot)$ is symmetric, it is interesting that under the original, easier to demonstrate inf-sup condition (6.14) in terms of ∂_t, a quasi-optimality result similar to (6.18) can be shown, where then the upper bound for $\|u-u^{\hat\delta\delta}\|_X/\|u-u_\delta\|_X$ depends on $\|A_a\|_{\mathcal{L}(Y,Y')}$, and cannot be driven to 1 by taking Y^δ sufficiently large in relation to X^δ. The latter, however, will be essential for the analysis in the current work, being the reason why we consider only symmetric $a(t;\cdot,\cdot)$.

6.3.3 Modified discretized saddle-point

In view of obtaining an efficient implementation, in the definition of $(\mu^{\hat\delta\delta}, u^{\hat\delta\delta})$ in (6.15), and so in that of $S^{\hat\delta\delta}$ and $f^{\hat\delta\delta}$ in (6.16), we replace $(E_Y^{\hat\delta\prime} A E_Y^{\hat\delta})^{-1}$ by some $K_Y^{\hat\delta} = K_Y^{\hat\delta\prime} \in \mathcal{L}\mathrm{is}(Y^{\hat\delta\prime}, Y^{\hat\delta})$ for which both, for some constant $\kappa_\Delta \geq 1$,

$$\frac{((K_Y^{\hat\delta})^{-1}v)(v)}{(Av)(v)} \in [\kappa_\Delta^{-1}, \kappa_\Delta] \quad (\delta\in\Delta,\, v\in Y^{\hat\delta}) \tag{6.19}$$

(i.e. $K_Y^{\hat\delta}$ is an optimal (self-adjoint and coercive) *preconditioner* for $E_Y^{\hat\delta\prime} A E_Y^{\hat\delta}$), and which can be applied at linear cost. The resulting system (6.16) is now amenable to the application of the (preconditioned) conjugate residuals iteration.

Despite this modification, we keep using the old notations for $\mu^{\hat\delta\delta}$, $u^{\hat\delta\delta}$, $S^{\hat\delta\delta}$, $\|\cdot\|_{X^{\hat\delta\delta}} := (S^{\hat\delta\delta}\cdot)(\cdot)^{\frac{1}{2}}$, and $f^{\hat\delta\delta}$.

As shown in [SW21b, Remark 3.8], instead of (6.18) now it holds that

$$\|u-u_\delta\|_X \le \|u-u^{\hat\delta\delta}\|_X \le \frac{\kappa_\Delta}{\gamma_\Delta}\|u-u_\delta\|_X, \tag{6.20}$$

whereas one deduces that (6.17) now should be read as

$$\frac{\gamma_\Delta}{\sqrt{\kappa_\Delta}}\|\cdot\|_X \le \|\cdot\|_{X^{\delta\delta}} \le \sqrt{\kappa_\Delta}\|\cdot\|_X \quad \text{on } X^\delta. \tag{6.21}$$

For our forthcoming analysis, we will need $\frac{\kappa_\Delta}{\gamma_\Delta}-1$ to be sufficiently small.

Remark 6.3.6. Later, in the proof of Proposition 6.4.5, temporarily we will consider the system (6.15) with $\hat\delta=\delta$ (i.e. $Y^{\hat\delta}=Y^\delta$), but with X^δ replaced by X, and, as we do in the current subsection, ${E_Y^\delta}'AE_Y^\delta$ replaced by $(K_Y^\delta)^{-1}$. The resulting Schur operator $B'E_Y^\delta K_Y^\delta {E_Y^\delta}'B+\gamma_0'\gamma_0$ will be denoted as $S^{\delta\infty}$.

Notice that the exact solution u solves $S^{\delta\infty}u=B'E_Y^\delta K_Y^\delta {E_Y^\delta}'g+\gamma_0'u_0$. Observing that $S^{\delta\delta}={E_X^\delta}'S^{\delta\infty}E_X^\delta$, we have the Galerkin orthogonality $(S^{\delta\infty}(u-u^{\delta\delta}))(X^\delta)=0$. It holds that

$$\begin{aligned}\|\cdot\|_{X^{\delta\infty}} := (S^{\delta\infty}\cdot)(\cdot)^{\frac12} &= \sqrt{\sup_{0\ne v\in Y^\delta}\frac{((B\cdot)(v))^2}{((K_Y^\delta)^{-1}v)(v)}+\|\gamma_0\cdot\|_H^2}\\ &= \sqrt{(E_Y^\delta{}'B\cdot)(K_Y^\delta E_Y^\delta{}'B\cdot)+\|\gamma_0\cdot\|_H^2}\end{aligned}$$

is only a *semi*-norm on X, which is equal to $\|\cdot\|_{X^{\delta\delta}}$ on X^δ, and

$$\|\cdot\|_{X^{\delta\infty}} \le \sqrt{\kappa_\Delta}\|\cdot\|_X \quad \text{on } X. \tag{6.22}$$

6.4 Convergent adaptive solution method

6.4.1 Preliminaries

For $\delta\in\Delta$, we consider the modified discretized saddle point problem (i.e. (6.16) with $({E_Y^{\hat\delta}}'AE_Y^{\hat\delta})^{-1}$ replaced by $K_Y^{\hat\delta}$) taking $\hat\delta:=\underline\delta$ from Assumption 6.3.1. So for a given 'trial space' X^δ, we employ $Y^{\underline\delta}$ as 'test space', which is known to be sufficiently large to give stability even when employed with trial space $X^{\underline\delta}\supsetneq X^\delta$. We will use this room to (adaptively) expand X^δ to some $X^{\tilde\delta}\subset X^{\underline\delta}$ while keeping $Y^{\underline\delta}$ fixed. Then in a second step we adapt the test space to the new trial space, i.e., replace $Y^{\underline\delta}$ by $Y^{\underline{\tilde\delta}}$. By doing so will construct a sequence $(\delta_i)\subseteq\Delta$ with $\delta_i\preceq\delta_{i+1}$ such that $(u^{\underline{\delta_i}\delta_i})_i$ converges r-linearly to u.

As a first step, in the next lemma it is shown that if one constructs from $w\in X^\delta$ a $v\in X^{\underline\delta}$ that is closer to the best approximation $u_{\underline\delta}$ to u from $X^{\underline\delta}$, then, thanks to Assumption 6.3.1, v is also closer to u.

Lemma 6.4.1. *Let $w \in X^\delta$, $v \in X^{\underline{\delta}}$ be such that for some $\rho \leq 1$,*

$$\|u_{\underline{\delta}} - v\|_X \leq \rho \|u_{\underline{\delta}} - w\|_X.$$

Then

$$\|u - v\|_X \leq \sqrt{\zeta^2 + \rho^2(1-\zeta^2)}\, \|u - w\|_X.$$

Proof. Using $u - u_{\underline{\delta}} \perp_X X^{\underline{\delta}}$ twice, we obtain

$$\begin{aligned}
\|u - v\|_X^2 &= \|u - u_{\underline{\delta}}\|_X^2 + \|u_{\underline{\delta}} - v\|_X^2 \\
&\leq \|u - u_{\underline{\delta}}\|_X^2 + \rho^2 \|u_{\underline{\delta}} - w\|_X^2 \\
&= \|u - u_{\underline{\delta}}\|_X^2 + \rho^2 (\|u - w\|_X^2 - \|u - u_{\underline{\delta}}\|_X^2) \\
&= (1-\rho^2)\|u - u_{\underline{\delta}}\|_X^2 + \rho^2 \|u - w\|_X^2 \\
&\leq (\zeta^2(1-\rho^2) + \rho^2)\|u - w\|_X^2,
\end{aligned}$$

where we used Assumption 6.3.1 and $\|u - u_\delta\|_X \leq \|u - w\|_X$. □

Notice that $u^{\underline{\delta}\delta}$ is the Galerkin approximation from X^δ to the solution $u^{\underline{\delta}\underline{\delta}} \in X^{\underline{\delta}}$ of the system $S^{\underline{\delta}\underline{\delta}} u^{\underline{\delta}\underline{\delta}} = f^{\underline{\delta}\underline{\delta}}$, i.e., it is its best approximation from X^δ w.r.t. $\|\cdot\|_{X^{\underline{\delta}\underline{\delta}}}$. In the next proposition it is shown that an improved Galerkin approximation from an intermediate space $X^{\underline{\delta}} \supseteq X^{\tilde{\delta}} \supseteq X^\delta$, i.e., the function $u^{\underline{\delta}\tilde{\delta}}$, is, for $\frac{\kappa_\Delta}{\gamma_\Delta} - 1$ sufficiently small, also an improved approximation to u, and furthermore that this holds true also for $u^{\tilde{\underline{\delta}}\tilde{\delta}}$. The latter function will be the successor of $u^{\underline{\delta}\delta}$ in our converging sequence.

Proposition 6.4.2. *Let $\delta \preceq \tilde{\delta} \preceq \underline{\delta}$ be such that*

$$\|u^{\underline{\delta}\underline{\delta}} - u^{\underline{\delta}\tilde{\delta}}\|_{X^{\underline{\delta}\underline{\delta}}} \leq \rho \|u^{\underline{\delta}\underline{\delta}} - u^{\underline{\delta}\delta}\|_{X^{\underline{\delta}\underline{\delta}}}. \tag{6.23}$$

Then it holds that

$$\|u - u^{\tilde{\underline{\delta}}\tilde{\delta}}\|_X \leq \underbrace{\frac{\kappa_\Delta}{\gamma_\Delta}\sqrt{\zeta^2 + \hat{\rho}^2(1-\zeta^2)}}_{\bar{\rho}:=} \|u - u^{\underline{\delta}\delta}\|_X,$$

where $\hat{\rho} := \big(1 + \rho\frac{\sqrt{\kappa_\Delta}}{\gamma_\Delta}\big)\sqrt{\frac{\kappa_\Delta^2}{\gamma_\Delta^2} - 1}\,\sqrt{\frac{\zeta^2}{1-\zeta^2}} + \rho\frac{\sqrt{\kappa_\Delta}}{\gamma_\Delta}$. Notice that $\hat{\rho}$ and $\bar{\rho}$ are < 1 when $\rho < 1$ and $\frac{\kappa_\Delta}{\gamma_\Delta} - 1$ is sufficiently small dependent on ρ with $\frac{\kappa_\Delta}{\gamma_\Delta} - 1 \downarrow 0$ when $\rho \uparrow 1$.

Proof. Using that $u - u_{\underline{\delta}} \perp_X X^{\underline{\delta}}$, it follows that $\|u - u^{\underline{\delta}\delta}\|_X^2 \leq \frac{\kappa_\Delta^2}{\gamma_\Delta^2}\|u - u_\delta\|_X^2$ ((6.20)) is equivalent to $\|u_{\underline{\delta}} - u^{\underline{\delta}\delta}\|_X \leq \sqrt{\frac{\kappa_\Delta^2}{\gamma_\Delta^2} - 1}\,\|u - u_\delta\|_X$. Similarly, Assumption 6.3.1 is equivalent to $\|u - u_\delta\|_X \leq \sqrt{\frac{\zeta^2}{1-\zeta^2}}\|u_{\underline{\delta}} - w\|_X$ for any $w \in X^\delta$.

Additionally using (6.21), we infer that

$$\begin{aligned}\|u_{\underline{\delta}} - u^{\delta\tilde{\delta}}\|_X &\le \|u_{\underline{\delta}} - u^{\underline{\delta}\delta}\|_X + \|u^{\underline{\delta}\delta} - u^{\delta\tilde{\delta}}\|_X \\ &\le \|u_{\underline{\delta}} - u^{\underline{\delta}\delta}\|_X + \frac{\sqrt{\kappa_\Delta}}{\gamma_\Delta}\|u^{\underline{\delta}\delta} - u^{\delta\tilde{\delta}}\|_{X^{\underline{\delta}\delta}} \\ &\le \|u_{\underline{\delta}} - u^{\underline{\delta}\delta}\|_X + \rho\frac{\sqrt{\kappa_\Delta}}{\gamma_\Delta}\|u^{\underline{\delta}\delta} - u^{\delta\delta}\|_X \\ &\le \big(1+\rho\frac{\sqrt{\kappa_\Delta}}{\gamma_\Delta}\big)\|u_{\underline{\delta}} - u^{\underline{\delta}\delta}\|_X + \rho\frac{\sqrt{\kappa_\Delta}}{\gamma_\Delta}\|u_{\underline{\delta}} - u^{\delta\delta}\|_X \\ &\le \Big[\big(1+\rho\tfrac{\sqrt{\kappa_\Delta}}{\gamma_\Delta}\big)\sqrt{\tfrac{\kappa_\Delta^2}{\gamma_\Delta^2}-1}\,\sqrt{\tfrac{\zeta^2}{1-\zeta^2}} + \rho\tfrac{\sqrt{\kappa_\Delta}}{\gamma_\Delta}\Big]\|u_{\underline{\delta}} - u^{\delta\delta}\|_X\end{aligned}$$

From Lemma 6.4.1 we conclude that $\|u-u^{\delta\tilde{\delta}}\|_X \le \sqrt{\zeta^2+\hat{\rho}^2(1-\zeta^2)}\|u-u^{\delta\delta}\|_X$. Thanks to (6.20), it holds that

$$\|u - u^{\tilde{\delta}\tilde{\delta}}\|_X \le \frac{\kappa_\Delta}{\gamma_\Delta}\|u - u_{\tilde{\delta}}\|_X \le \frac{\kappa_\Delta}{\gamma_\Delta}\|u - u^{\delta\tilde{\delta}}\|_X,$$

which completes the proof. □

6.4.2 Bulk chasing and a posteriori error estimation

To realize (6.23), i.e., to construct from the Galerkin approximation $u^{\underline{\delta}\delta}$ to $u^{\underline{\delta}\underline{\delta}}$ an improved Galerkin approximation $u^{\underline{\delta}\tilde{\delta}}$, we apply the concept of bulk chasing, also known as Dörfler marking, on a collection of a posteriori error indicators that constitute an efficient and reliable error estimator. We will apply an estimator of 'hierarchical basis' type ([ZMD+11]):

Let $\Theta_\delta = \{\theta_\lambda : \lambda \in J_\delta\} \subseteq X^\delta$ be such that $X^\delta + \operatorname{span}\Theta_\delta = X^{\underline{\delta}}$ and, for some constants $0 < m \le M$, for all $\delta \in \Delta$, $z \in X^\delta$ and $\mathbf{c} := (c_\lambda)_{\lambda\in J_\delta} \subset \mathbb{R}$.

$$m^2\|z + \mathbf{c}^\top\Theta_\delta\|_X^2 \le \|z\|_X^2 + \|\mathbf{c}\|^2 \le M^2\|z + \mathbf{c}^\top\Theta_\delta\|_X^2. \tag{6.24}$$

A suitable collection Θ_δ will be constructed in Sect. 6.6.1.

Proposition 6.4.3. *Assume* (6.24). *Let* $\mathbf{r}_\delta^{\underline{\delta}} := (f^{\underline{\delta}\underline{\delta}} - S^{\underline{\delta}\underline{\delta}}u^{\underline{\delta}\delta})(\Theta_\delta)$, *being the residual vector of* $u^{\underline{\delta}\delta}$. *Let* $J \subseteq J_\delta$ *be such that for some constant* $\vartheta \in (0,1]$,

$$\|\mathbf{r}_\delta^{\underline{\delta}}|_J\| \ge \vartheta\|\mathbf{r}_\delta^{\underline{\delta}}\|,$$

and, for some $\tilde{\delta} \preceq \underline{\delta}$, *let* $X^\delta + \operatorname{span}\Theta_\delta|_J \subseteq X^{\tilde{\delta}}$. *Then with* $\rho := \sqrt{1-\big(\frac{m}{M}\frac{\gamma_\Delta}{\kappa_\Delta^2}\vartheta\big)^2}$, (6.23) *is valid, i.e.,*

$$\|u^{\underline{\delta}\delta} - u^{\delta\tilde{\delta}}\|_{X^{\underline{\delta}\underline{\delta}}} \le \rho\|u^{\underline{\delta}\delta} - u^{\underline{\delta}\delta}\|_{X^{\underline{\delta}\underline{\delta}}}; \tag{6.25}$$

and so, when $\frac{\kappa_\Delta}{\gamma_\Delta} - 1$ *is sufficiently small dependent on* ϑ*, with* $\frac{\kappa_\Delta}{\gamma_\Delta} - 1 \downarrow 0$ *when* $\vartheta \downarrow 0$*, for some constant* $\rho < \bar{\rho} < 1$*,*

$$\|u - u^{\tilde{\delta}\tilde{\delta}}\|_X \leq \bar{\rho}\|u - u^{\delta\delta}\|_X.$$

Proof. As a consequence of (6.24) and (6.21), we have

$$\frac{\gamma_\Delta}{\kappa_\Delta^2} m^2 \|z + \mathbf{c}^\top \Theta_\delta\|^2_{X^{\underline{\delta}\underline{\delta}}} \leq \|z\|^2_{X^{\underline{\delta}\underline{\delta}}} + \|\mathbf{c}\|^2 \leq \frac{\kappa_\Delta^2}{\gamma_\Delta} M^2 \|z + \mathbf{c}^\top \Theta_\delta\|^2_{X^{\underline{\delta}\underline{\delta}}}.$$

We infer that

$$\begin{aligned}
\|u^{\underline{\delta}\tilde{\delta}} - u^{\underline{\delta}\delta}\|_{X^{\underline{\delta}\underline{\delta}}} &= \sup_{0 \neq (z,\mathbf{c}) \in X^\delta \times \mathbb{R}^{\#J_\delta}} \frac{(S^{\underline{\delta}\underline{\delta}}(u^{\underline{\delta}\tilde{\delta}} - u^{\underline{\delta}\delta}))(z + \mathbf{c}^\top \Theta_\delta)}{\|z + \mathbf{c}^\top \Theta_\delta\|_{X^{\underline{\delta}\underline{\delta}}}} \\
&\geq m \frac{\sqrt{\gamma_\Delta}}{\kappa_\Delta} \sup_{0 \neq (z,\mathbf{c}) \in X^\delta \times \mathbb{R}^{\#J_\delta}} \frac{(S^{\underline{\delta}\underline{\delta}}(u^{\underline{\delta}\tilde{\delta}} - u^{\underline{\delta}\delta}))(\mathbf{c}^\top \Theta_\delta)}{\sqrt{\|z\|^2_{X^{\underline{\delta}\underline{\delta}}} + \|\mathbf{c}\|^2}} \\
&\geq m \frac{\sqrt{\gamma_\Delta}}{\kappa_\Delta} \sup_{0 \neq \mathbf{c} \in \mathbb{R}^{\#J_\delta}} \frac{\langle \mathbf{c}|_J, (f^{\underline{\delta}\underline{\delta}} - S^{\underline{\delta}\underline{\delta}} u^{\underline{\delta}\delta})(\Theta_\delta|_J)\rangle}{\|\mathbf{c}|_J\|} \\
&= m \frac{\sqrt{\gamma_\Delta}}{\kappa_\Delta} \|\mathbf{r}_\delta^{\underline{\delta}}|_J\| \geq m \frac{\sqrt{\gamma_\Delta}}{\kappa_\Delta} \vartheta \|\mathbf{r}_{\underline{\delta}}^{\underline{\delta}}\| \\
&= m \frac{\sqrt{\gamma_\Delta}}{\kappa_\Delta} \vartheta \sup_{0 \neq (z,\mathbf{c}) \in X^\delta \times \mathbb{R}^{\#J_\delta}} \frac{(S^{\underline{\delta}\underline{\delta}}(u^{\underline{\delta}\delta} - u^{\underline{\delta}\delta}))(\mathbf{c}^\top \Theta_\delta)}{\sqrt{\|z\|^2_{X^{\underline{\delta}\underline{\delta}}} + \|\mathbf{c}\|^2}} \\
&\geq \frac{m}{M} \frac{\gamma_\Delta}{\kappa_\Delta^2} \vartheta \|u^{\underline{\delta}\delta} - u^{\underline{\delta}\delta}\|_{X^{\underline{\delta}\underline{\delta}}}.
\end{aligned} \tag{6.26}$$

so that

$$\begin{aligned}
\|u^{\underline{\delta}\delta} - u^{\underline{\delta}\tilde{\delta}}\|^2_{X^{\underline{\delta}\underline{\delta}}} &= \|u^{\underline{\delta}\delta} - u^{\underline{\delta}\delta}\|^2_{X^{\underline{\delta}\underline{\delta}}} - \|u^{\underline{\delta}\tilde{\delta}} - u^{\underline{\delta}\delta}\|^2_{X^{\underline{\delta}\underline{\delta}}} \\
&\leq \Big(1 - \big(\frac{m}{M} \frac{\gamma_\Delta}{\kappa_\Delta^2} \vartheta\big)^2\Big) \|u^{\underline{\delta}\delta} - u^{\underline{\delta}\delta}\|^2_{X^{\underline{\delta}\underline{\delta}}},
\end{aligned}$$

which completes the proof of (6.25).

The final statement follows from an application of Proposition 6.4.2. □

Additionally we have that $\|\mathbf{r}_\delta^{\underline{\delta}}\|$ provides an efficient and reliable a posteriori estimator for $\|u - u^{\underline{\delta}\delta}\|_X$:

Proposition 6.4.4. *Assume* (6.24)*. Recalling that* $\zeta < 1$*, let* $\frac{\kappa_\Delta}{\gamma_\Delta} < \frac{1}{\zeta}$*. Then for* $\delta \in \Delta$*,*

$$\frac{m \frac{\sqrt{\gamma_\Delta}}{\kappa_\Delta^{3/2}}}{1 + \zeta \sqrt{\frac{\kappa_\Delta^2}{\gamma_\Delta^2} - 1}} \|\mathbf{r}_\delta^{\underline{\delta}}\| \leq \|u - u^{\underline{\delta}\delta}\|_X \leq \frac{M \frac{\kappa_\Delta^{3/2}}{\gamma_\Delta^{3/2}}}{\sqrt{1 - \zeta^2} - \zeta \sqrt{\frac{\kappa_\Delta^2}{\gamma_\Delta^2} - 1}} \|\mathbf{r}_\delta^{\underline{\delta}}\|.$$

Proof. Assumption 6.3.1 gives $\|u-u_{\underline{\delta}}\|_X \leq \zeta\|u-u^{\delta\delta}\|_X$, which by $u-u_{\underline{\delta}} \perp_X X^{\underline{\delta}}$ yields

$$\|u_{\underline{\delta}} - u^{\delta\delta}\|_X \leq \|u - u^{\delta\delta}\|_X \leq \tfrac{1}{\sqrt{1-\zeta^2}}\|u_{\underline{\delta}} - u^{\delta\delta}\|_X. \tag{6.27}$$

As we already have noted in the proof of Proposition 6.4.2, (6.20) is equivalent to $\|u_{\underline{\delta}}-u^{\underline{\delta}\delta}\|_X \leq \sqrt{\frac{\kappa_\Delta^2}{\gamma_\Delta^2}-1}\,\|u-u_{\underline{\delta}}\|_X$. Together with Assumption 6.3.1, it gives

$$\left|\|u_{\underline{\delta}} - u^{\delta\delta}\|_X - \|u^{\underline{\delta}\delta} - u^{\delta\delta}\|_X\right| \leq \zeta\sqrt{\frac{\kappa_\Delta^2}{\gamma_\Delta^2}-1}\,\|u - u^{\delta\delta}\|_X$$

which in combination with (6.27) and $\zeta\frac{\kappa_\Delta}{\gamma_\Delta} < 1$ yields

$$\tfrac{1}{1+\zeta\sqrt{\frac{\kappa_\Delta^2}{\gamma_\Delta^2}-1}}\|u^{\underline{\delta}\delta} - u^{\delta\delta}\|_X \leq \|u - u^{\underline{\delta}\delta}\|_X \leq \tfrac{1}{\sqrt{1-\zeta^2}-\zeta\sqrt{\frac{\kappa_\Delta^2}{\gamma_\Delta^2}-1}}\|u^{\underline{\delta}\delta} - u^{\delta\delta}\|_X.$$

The proof is completed by (6.21), and $m\frac{\sqrt{\gamma_\Delta}}{\kappa_\Delta}\|\mathbf{r}_\delta^{\underline{\delta}}\| \leq \|u^{\underline{\delta}\delta}-u^{\delta\delta}\|_{X^{\underline{\delta}\delta}} \leq M\frac{\kappa_\Delta}{\sqrt{\gamma_\Delta}}\|\mathbf{r}_\delta^{\underline{\delta}}\|$ where the latter inequalities were shown in (6.26) when reading $(J,\vartheta,\tilde{\delta})$ as $(J_\delta, 1, \underline{\delta})$. □

Next we present an alternative a posteriori error estimator that does not rely on (6.24), that we expect to be more accurate, and that can be a computed at the cost of one additional inner product.

Proposition 6.4.5. *Let* $\frac{\kappa_\Delta}{\gamma_\Delta} < \frac{1}{\zeta}$, *and for* $v \in X^\delta$, *define*

$$\mathcal{E}^\delta(v) = \mathcal{E}^\delta(v; g, u_0) := \sqrt{(E_Y^{\delta\,\prime}(g - Bv))(K_Y^\delta E_Y^{\delta\,\prime}(g - Bv)) + \|u_0 - \gamma_0 v\|_H^2}.$$

Then

$$\frac{\gamma_\Delta - \zeta\kappa_\Delta}{\sqrt{\kappa_\Delta}}\|u - v\|_X \leq \mathcal{E}^\delta(v) \leq \sqrt{\kappa_\Delta(\zeta^2 + (1+\zeta\frac{\kappa_\Delta}{\gamma_\Delta})^2)}\|u - v\|_X.$$

Proof. From Remark 6.3.6, recall that the semi-norm $\|\cdot\|_{X^{\underline{\delta}\infty}}$ on X equals $\|\cdot\|_{X^{\underline{\delta}\underline{\delta}}}$ on $X^{\underline{\delta}}$, and $(S^{\delta\infty}(u - u^{\underline{\delta}\underline{\delta}}))(X^{\underline{\delta}}) = 0$, which implies

$$\|u - w\|_{X^{\delta\infty}}^2 = \|u - u^{\underline{\delta}\underline{\delta}}\|_{X^{\delta\infty}}^2 + \|u^{\underline{\delta}\underline{\delta}} - w\|_{X^{\underline{\delta}\infty}}^2 \quad (w \in X^{\underline{\delta}}), \tag{6.28}$$

and $\|\cdot\|_{X^{\underline{\delta}\infty}} \leq \sqrt{\kappa_\Delta}\|\cdot\|_X$ ((6.22)).

From (6.20) and Assumption 6.3.1, we have for $v \in X^\delta$

$$\|u - u^{\underline{\delta}\underline{\delta}}\|_X \leq \frac{\kappa_\Delta}{\gamma_\Delta}\|u - u_{\underline{\delta}}\|_X \leq \zeta\frac{\kappa_\Delta}{\gamma_\Delta}\|u - v\|_X \tag{6.29}$$

From (6.29), the triangle-inequality, (6.21), and (6.28) we obtain for $v \in X^\delta$

$$\begin{aligned}\|u-v\|_X &\leq \frac{1}{1-\zeta\frac{\kappa_\Delta}{\gamma_\Delta}}\|u^{\delta\underline{\delta}}-v\|_X \\ &\leq \frac{\sqrt{\kappa_\Delta}}{\gamma_\Delta-\zeta\kappa_\Delta}\|u^{\delta\underline{\delta}}-v\|_{X^{\underline{\delta}\underline{\delta}}} \leq \frac{\sqrt{\kappa_\Delta}}{\gamma_\Delta-\zeta\kappa_\Delta}\|u-v\|_{X^{\underline{\delta}\infty}}.\end{aligned}$$

Conversely, we have

$$\begin{aligned}\|u-v\|^2_{X^{\underline{\delta}\infty}} &\overset{(6.28)}{=} \|u-u^{\delta\underline{\delta}}\|^2_{X^{\underline{\delta}\infty}} + \|u^{\delta\underline{\delta}}-v\|^2_{X^{\underline{\delta}\infty}} \\ &\overset{(6.28),(6.22)}{\leq} \|u-u_{\underline{\delta}}\|^2_{X^{\underline{\delta}\infty}} + \kappa_\Delta\|u^{\delta\underline{\delta}}-v\|^2_X \\ &\overset{(6.22),(6.29)}{\leq} \kappa_\Delta\|u-u_{\underline{\delta}}\|^2_X + \kappa_\Delta(1+\zeta\frac{\kappa_\Delta}{\gamma_\Delta})^2\|u-v\|^2_X \\ &\leq \kappa_\Delta(\zeta^2+(1+\zeta\frac{\kappa_\Delta}{\gamma_\Delta})^2)\|u-v\|^2_X.\end{aligned}$$

by again applying Assumption 6.3.1.

Noting that $\|u-v\|^2_{X^{\underline{\delta}\infty}} := {E^{\underline{\delta}}_Y}'(g-Bv)(K^{\underline{\delta}}_Y E^{\underline{\delta}}_Y{}'(g-Bv)) + \|u_0-\gamma_0 v\|^2_H$, the proof is completed. □

Notice that the estimator of $\|u-v\|_X$ from Proposition 6.4.5 is exact when $\zeta=0$ and $\kappa_\Delta = 1 = \gamma_\Delta$, where the one from Proposition 6.4.3, for $v = u^{\delta\underline{\delta}}$, is exact only when additionally $m = 1 = M$.

6.4.3 Data oscillation

In view of the discussion following Assumption 6.3.1, notice that *all* results obtained so far that depend on the 'saturation constant' ζ, i.e., Lemma 6.4.1, and Propositions 6.4.2, 6.4.3, and 6.4.4, are only valid under the condition that $(g,u_0) \in {}^\delta G \times {}^\delta U_0$.

Let us now consider the situation that the solutions and residuals in these statements refer to solutions and residuals with the true data $(g,u_0) \in Y' \times H$ being *replaced* by an approximation $({}^\delta g, {}^\delta u_0) \in {}^\delta G \times {}^\delta U_0$. In the following we denote such solutions and residuals with an *additional left superscript* δ, or more generally $\tilde{\delta}$ when a right-hand side $({}^{\tilde{\delta}} g, {}^{\tilde{\delta}} u_0) \in {}^{\tilde{\delta}} G \times {}^{\tilde{\delta}} U_0$ has been used for their computation.

Proposition 6.4.6. *Assume* (6.24)*, and let* $\vartheta \in (0,1]$ *be a constant. Then for* $\frac{\kappa_\Delta}{\gamma_\Delta}-1$ *and a constant* $\hat{\omega} > 0$ *both being sufficiently small dependent on* ϑ*, with* $\max(\frac{\kappa_\Delta}{\gamma_\Delta}-1,\hat{\omega}) \downarrow 0$ *when* $\vartheta \downarrow 0$*, there exists a constant* $\check{\rho} < 1$ *such that for* $J \subseteq J_\delta$ *with* $\|{}^\delta\mathbf{r}_{\tilde{\delta}}|_J\| \geq \vartheta\|{}^\delta\mathbf{r}_{\tilde{\delta}}\|$*, and* $X^\delta + \operatorname{span}\Theta_\delta|_J \subseteq X^{\tilde{\delta}}$*, and*

$$\max\left(\|g-{}^\delta g\|_{Y'}+\|u_0-{}^\delta u_0\|_H, \|g-{}^{\tilde{\delta}} g\|_{Y'}+\|u_0-{}^{\tilde{\delta}} u_0\|_H\right) \leq \hat{\omega}\|{}^\delta\mathbf{r}_{\tilde{\delta}}\|,$$

it holds that

$$\|{}^{\tilde\delta}u - {}^{\tilde\delta}u^{\tilde\delta\tilde\delta}\|_X \le \check\rho \|{}^{\delta}u - {}^{\delta}u^{\delta\delta}\|_X.$$

Proof. In the newly introduced notations, the statements of Propositions 6.4.3 and 6.4.4 read as

$$\|{}^{\delta}u - {}^{\delta}u^{\tilde\delta\tilde\delta}\|_X \le \bar\rho \|{}^{\delta}u - {}^{\delta}u^{\delta\delta}\|_X,$$

and

$$\|{}^{\delta}\mathbf{r}_{\delta}^{\tilde\delta}\| \eqsim \|{}^{\delta}u - {}^{\delta}u^{\delta\delta}\|_X \tag{6.30}$$

The proof is easily completed by

$$\left.\begin{matrix}\|{}^{\tilde\delta}u - {}^{\delta}u\|_X \\ \|{}^{\tilde\delta}u^{\tilde\delta\tilde\delta} - {}^{\delta}u^{\tilde\delta\tilde\delta}\|_X\end{matrix}\right\} \lesssim \|{}^{\tilde\delta}g - {}^{\delta}g\|_{Y'} + \|{}^{\tilde\delta}u_0 - {}^{\delta}u_0\|_H \le 2\hat\omega \|{}^{\delta}\mathbf{r}_{\tilde\delta}^{\tilde\delta}\| \eqsim \hat\omega \|{}^{\delta}u - {}^{\delta}u^{\delta\delta}\|_X. \ \square$$

In view of the latter proposition, we make the following assumption.

Assumption 6.4.7. *We assume to have maps of the following types available:*

$\Delta \to Y' \times H : \delta \mapsto ({}^{\delta}g, {}^{\delta}u_0) \in {}^{\delta}G \times {}^{\delta}U_0$,

$\eta : \Delta \to \mathbb{R}$ *such that* $\|g - {}^{\delta}g\|_{Y'} + \|u_0 - {}^{\delta}u_0\|_H \le \eta(\delta)$, *and* $\eta(\tilde\delta) \le \eta(\delta)$ *when* $\tilde\delta \succeq \delta$,

$\mathbb{R}_{>0} \to \Delta : \varepsilon \mapsto \delta(\varepsilon)$ *such that* $\eta(\delta(\varepsilon)) \le \varepsilon$.

Notice that this in particular means that for any $\varepsilon > 0$ we are able to find a $\delta \in \Delta$ and $({}^{\delta}g, {}^{\delta}u_0) \in {}^{\delta}G \times {}^{\delta}U_0$ with $\|g - {}^{\delta}g\|_{Y'} + \|u_0 - {}^{\delta}u_0\|_H \le \varepsilon$. A specification of a suitable family $({}^{\delta}G, {}^{\delta}U_0)_{\delta\in\Delta}$ will be given in Sect. 6.6.4.

Given a $\delta \in \Delta$, and thinking of $({}^{\delta}g, {}^{\delta}u_0)$ being a quasi-best approximation to (g, u_0) from ${}^{\delta}G \times {}^{\delta}U_0$, the difference $(g, u_0) - ({}^{\delta}g, {}^{\delta}u_0)$ is often referred to as *data-oscillation*.

6.4.4 A convergent algorithm

In view of the statement from Proposition 6.4.6, in the following we will use the short-hand notations

$$u^{\delta} = {}^{\delta}u^{\delta\delta}, \quad \mathbf{r}^{\delta} = {}^{\delta}\mathbf{r}_{\delta}^{\delta},$$

i.e., u^{δ} is the solution of

$$\underbrace{{E_X^{\delta}}'(B'E_Y^{\delta}K_Y^{\delta}{E_Y^{\delta}}'B + \gamma_0'\gamma_0)E_X^{\delta}}_{S^{\delta\delta}=} u^{\delta} = \underbrace{{E_X^{\delta}}'(B'E_Y^{\delta}K_Y^{\delta}{E_Y^{\delta}}'{}^{\delta}g + \gamma_0'{}^{\delta}u_0)}_{f^{\delta}={}^{\delta}f_{\delta}^{\delta}:=}. \tag{6.31}$$

(cf. (6.16) and Sect. 6.3.3), and

$$\mathbf{r}^{\delta} = {E_X^{\delta}}'\big[B'E_Y^{\delta}K_Y^{\delta}{E_Y^{\delta}}'({}^{\delta}g - Bu^{\delta}) + \gamma_0'({}^{\delta}u_0 - \gamma_0 u^{\delta})\big](\Theta_{\delta}) \tag{6.32}$$

Instead of solving (6.31) exactly, we will allow it to be solved approximately with a sufficiently small relative tolerance by the application of an iterative method. To that end, we assume to have available a $K_X^\delta = {K_X^\delta}' \in \mathcal{L}\mathrm{is}(X^{\delta\prime}, X^\delta)$ for which both

$$((K_X^\delta)^{-1}w)(w) \eqsim \|w\|_X^2 \quad (w \in X^\delta) \tag{6.33}$$

(i.e. K_X^δ is an optimal (self-adjoint and coercive) *preconditioner* for $S^{\underline{\delta}\delta}$), and which can be applied at linear cost. Besides for an efficient iterative solving of (6.31), we will use this preconditioner to compute a quantity that is equivalent to the X-norm of the (algebraic) error in any approximation from X^δ to u^δ.

We denote such an *approximate* solution of (6.31) by $\tilde{u}^\delta \in X^\delta$, with corresponding residual vector $\tilde{\mathbf{r}}^\delta$ defined as in (6.32) by replacing u^δ by $\tilde{u}^\delta$.

Algorithm 6.4.8.
Let $\omega > 0$, $\vartheta \in (0,1]$, $0 < \xi < 1$ be constants, and let $\varepsilon > 0$.
$\delta := \delta_{\mathrm{init}} \in \Delta$, $t_\delta \eqsim \|g\|_{Y'} + \|u_0\|_H$.
do
  do compute $\tilde{u}^\delta \in X^\delta$ with $\tilde{t}_\delta := \sqrt{(f^\delta - S^{\underline{\delta}\delta}\tilde{u}^\delta)(K_X^\delta(f^\delta - S^{\underline{\delta}\delta}\tilde{u}^\delta))} \leq \frac{t_\delta}{2}$; $t_\delta := \tilde{t}_\delta$
    `if` $e_\delta := \|\tilde{\mathbf{r}}^\delta\| + \eta(\delta) + t_\delta \leq \varepsilon$ `then stop endif`
  `until` $t_\delta \leq \xi e_\delta$
  `if` $\eta(\delta) > \omega\|\tilde{\mathbf{r}}^\delta\|$
  `then` select $\tilde{\delta} \in \Delta$ s.t. $X^{\tilde{\delta}} \supseteq X^\delta$ is (a near-smallest) space with $\eta(\tilde{\delta}) \leq \eta(\delta)/2$.
  `else` determine $\delta \preceq \tilde{\delta} \preceq \underline{\delta}$ s.t. $X^{\tilde{\delta}}$ is (a near- smallest) space that for a $J \subseteq I_{\tilde{\delta}}^{\delta}$
      contains $X^\delta + \operatorname{span}\Theta_\delta|_J$ where $\|\tilde{\mathbf{r}}^\delta|_J\| \geq \vartheta\|\tilde{\mathbf{r}}^\delta\|$.
  `endif`
  $t_{\tilde{\delta}} := e_\delta$, $\delta := \tilde{\delta}$
`enddo`

Theorem 6.4.9. *Assume* (6.24)*, and let the constants* γ_Δ *and* κ_Δ *be as defined in* (6.14) *and* (6.19)*, respectively. For constants* ϑ*,* ω/ϑ*,* ξ/ϑ*,* $(\frac{\kappa_\Delta}{\gamma_\Delta} - 1)/\omega$ *that are sufficiently small, with additionally* ω *and* $\frac{\kappa_\Delta}{\gamma_\Delta} - 1$ *sufficiently small dependent on* ϑ *with* $\max(\frac{\kappa_\Delta}{\gamma_\Delta} - 1, \omega) \downarrow 0$ *when* $\vartheta \downarrow 0$*, there exists a constant* $\breve{\rho} < 1$ *such that between any two successive passings of the until-clause the value of* $\vartheta\|u - \tilde{u}^\delta\|_X + \eta(\delta)$ *decreases with at least a factor* $\breve{\rho}$*. For any* $\varepsilon > 0$*, Algorithm 6.4.8 terminates, and at termination it holds that* $\|u - \tilde{u}^\delta\|_X + \eta(\delta) \lesssim \varepsilon$*.*

Remark 6.4.10. Minor adaptations to the proof show that the statement remains true when one takes $e_\delta := \mathcal{E}^\delta(\tilde{u}^\delta; {}^\delta g, {}^\delta u_0) + \eta(\delta)$ in Algorithm 6.4.8. Having to compute $\tilde{\mathbf{r}}^\delta$ anyway, the additional cost of computing this e_δ is small, and it can be expected to be closer to $\|u - \tilde{u}^\delta\|_X + \eta(\delta)$.

Proof. By replacing w by $K_X^\delta S^{\underline{\delta}\delta} w$ in (6.33), one infers that $(S^{\underline{\delta}\delta} v)(K_X^\delta S^{\underline{\delta}\delta} v) \eqsim \|v\|_X^2$ for $v \in X^\delta$, and so

$$\|u^\delta - \tilde{u}^\delta\|_X^2 \eqsim (f^\delta - S^{\underline{\delta}\delta}\tilde{u}^\delta)(K_X^\delta(f^\delta - S^{\underline{\delta}\delta}\tilde{u}^\delta)). \tag{6.34}$$

From (6.21) and (6.24), we have

$$\|\mathbf{r}^\delta - \tilde{\mathbf{r}}^\delta\| = \sup_{0\neq \mathbf{c}\in\mathbb{R}^{\#J^\delta}} \frac{(S^{\underline{\delta}\delta}(\tilde{u}^\delta - u^\delta))(\mathbf{c}^\top \Theta_\delta)}{\|\mathbf{c}\|} \leq \tfrac{\kappa_\Delta^{\frac{1}{2}}}{m}\|u^\delta - \tilde{u}^\delta\|_{X^{\underline{\delta}\delta}} \leq \tfrac{\kappa_\Delta}{m}\|u^\delta - \tilde{u}^\delta\|_X. \tag{6.35}$$

If the algorithm stops, then

$$\begin{aligned}
\|u - \tilde{u}^\delta\|_X &\leq \|u - {}^\delta u\|_X + \|{}^\delta u - u^\delta\|_X + \|u^\delta - \tilde{u}^\delta\|_X \\
&\lesssim \eta(\delta) + \|{}^\delta u - u^\delta\|_X + \|u^\delta - \tilde{u}^\delta\|_X \quad \text{(by Thm. 6.2.1 \& Ass. 6.4.7)} \\
&\overset{(6.30)}{\eqsim} \eta(\delta) + \|\mathbf{r}^\delta\| + \|u^\delta - \tilde{u}^\delta\|_X \leq \eta(\delta) + \|\tilde{\mathbf{r}}^\delta\| + \|\mathbf{r}^\delta - \tilde{\mathbf{r}}^\delta\| + \|u^\delta - \tilde{u}^\delta\|_X \\
&\overset{(6.35)}{\lesssim} \eta(\delta) + \|\tilde{\mathbf{r}}^\delta\| + \|u^\delta - \tilde{u}^\delta\|_X \overset{(6.34)}{\lesssim} \eta(\delta) + \|\tilde{\mathbf{r}}^\delta\| + t_\delta \leq \varepsilon.
\end{aligned}$$

The inner do-loop always terminates either by passing the until-clause or by the stop-statement. Indeed, inside this loop the value of t_δ is driven to 0, so that $\|\tilde{\mathbf{r}}^\delta\| + \eta(\delta)$ tends to $\|\mathbf{r}^\delta\| + \eta(\delta)$. So if $\|\mathbf{r}^\delta\| + \eta(\delta) \neq 0$, then at some moment $t_\delta \leq \xi(\|\tilde{\mathbf{r}}^\delta\| + \eta(\delta) + t_\delta)$, whereas if $\|\mathbf{r}^\delta\| + \eta(\delta) = 0$ then at some moment $e_\delta \leq \varepsilon$.

When passing the until-clause, it holds that $t_\delta \leq \xi(\|\tilde{\mathbf{r}}^\delta\| + \eta(\delta) + t_\delta)$, and so by using $\xi < 1$ kicking back t_δ,

$$\begin{aligned}
t_\delta &\lesssim \xi(\|\tilde{\mathbf{r}}^\delta\| + \eta(\delta)) \\
&\leq \xi(\|\mathbf{r}^\delta\| + \|\tilde{\mathbf{r}}^\delta - \mathbf{r}^\delta\| + \eta(\delta)) \\
&\lesssim \xi(\|{}^\delta u - u^\delta\|_X + \|\tilde{u}^\delta - u^\delta\|_X + \eta(\delta)) \\
&\lesssim \xi(\|u - u^\delta\|_X + t_\delta + \eta(\delta))
\end{aligned} \tag{6.36}$$

Taking ξ small enough and kicking back t_δ, we obtain $t_\delta \lesssim \xi(\|u - u^\delta\|_X + \eta(\delta))$, and similarly

$$t_\delta \lesssim \xi(\|u - \tilde{u}^\delta\|_X + \eta(\delta)), \tag{6.37}$$

$$t_\delta \lesssim \xi(\|{}^\delta u - u^\delta\|_X + \eta(\delta)). \tag{6.38}$$

When passing the until-clause, furthermore we have

$$\begin{aligned}
\|u - \tilde{u}^\delta\|_X &\lesssim t_\delta + \|u - u^\delta\|_X \lesssim t_\delta + \|{}^\delta u - u^\delta\|_X + \eta(\delta) \eqsim t_\delta + \|\mathbf{r}^\delta\| + \eta(\delta) \\
&\leq t_\delta + \|\mathbf{r}^\delta - \tilde{\mathbf{r}}^\delta\| + \|\tilde{\mathbf{r}}^\delta\| + \eta(\delta) \lesssim t_\delta + \|u^\delta - \tilde{u}^\delta\|_X + \|\tilde{\mathbf{r}}^\delta\| + \eta(\delta) \\
&\lesssim t_\delta + \|\tilde{\mathbf{r}}^\delta\| + \eta(\delta) \overset{(6.36)}{\lesssim} \|\tilde{\mathbf{r}}^\delta\| + \eta(\delta).
\end{aligned} \tag{6.39}$$

Denoting δ at the subsequent passing of the until-clause as $\tilde{\delta}$, we have

$$\begin{aligned}\|u-\tilde{u}^{\tilde{\delta}}\|_X &\lesssim t_{\tilde{\delta}} + \|^{\tilde{\delta}}u - u^{\tilde{\delta}}\|_X + \eta(\tilde{\delta})\\ &\overset{(6.20)}{\leq} t_{\tilde{\delta}} + \frac{\kappa_\Delta}{\gamma_\Delta}\|^{\tilde{\delta}}u - {}^{\tilde{\delta}}u_{\tilde{\delta}}\|_X + \eta(\tilde{\delta})\\ &\overset{\delta\prec\tilde{\delta}}{\leq} t_{\tilde{\delta}} + \frac{\kappa_\Delta}{\gamma_\Delta}\|^{\tilde{\delta}}u - \tilde{u}^{\delta}\|_X + \eta(\tilde{\delta})\\ &\leq t_{\tilde{\delta}} + \frac{\kappa_\Delta}{\gamma_\Delta}\|u-\tilde{u}^{\delta}\|_X + \mathcal{O}(\eta(\tilde{\delta})).\end{aligned}$$

By using $t_{\tilde{\delta}} \lesssim \xi(\|u-\tilde{u}^{\tilde{\delta}}\|_X + \eta(\tilde{\delta}))$ ((6.37)) and kicking back $\|u-\tilde{u}^{\tilde{\delta}}\|_X$, we infer that for ξ sufficiently small,

$$\|u-\tilde{u}^{\tilde{\delta}}\|_X \leq (\frac{\kappa_\Delta}{\gamma_\Delta} + \mathcal{O}(\xi))\|u-\tilde{u}^{\delta}\|_X + \mathcal{O}(\eta(\tilde{\delta})). \tag{6.40}$$

In the case that

$$\eta(\delta) > \omega\|\tilde{\mathbf{r}}^{\delta}\|, \tag{6.41}$$

it holds that

$$\eta(\tilde{\delta}) \leq \eta(\delta)/2,$$

and so, thanks to (6.39) and (6.41),

$$\|u-\tilde{u}^{\delta}\|_X \lesssim \eta(\delta)/\omega. \tag{6.42}$$

For any constant $\rho_1 < 1$, using (6.40) and (6.42) we have

$$\begin{aligned}\|u-\tilde{u}^{\tilde{\delta}}\|_X &\leq \rho_1\|u-\tilde{u}^{\delta}\|_X + \tfrac{\kappa_\Delta/\gamma_\Delta+\mathcal{O}(\xi)-\rho_1}{\omega}\omega\|u-\tilde{u}^{\delta}\|_X + \mathcal{O}(\eta(\tilde{\delta}))\\ &\leq \rho_1\|u-\tilde{u}^{\delta}\|_X + \big(\tfrac{\kappa_\Delta/\gamma_\Delta+\mathcal{O}(\xi)-\rho_1}{\omega} + 1\big)C\eta(\delta),\end{aligned}$$

for some constant $C > 0$, and so

$$\vartheta\|u-\tilde{u}^{\tilde{\delta}}\|_X + \eta(\tilde{\delta}) \leq \rho_1\vartheta\|u-\tilde{u}^{\delta}\|_X + \Big(\vartheta\big(\tfrac{\kappa_\Delta/\gamma_\Delta-\rho_1+\mathcal{O}(\xi)}{\omega} + 1\big)C + \tfrac{1}{2}\Big)\eta(\delta)$$

Now let $\vartheta > 0$ be such that $2\vartheta C + \frac{1}{2} < 1$. Given a constant ω (which later will be selected such that ω/ϑ is sufficiently small), let $(\frac{\kappa_\Delta}{\gamma_\Delta} - 1)/\omega$, $(1-\rho_1)/\omega$, ξ/ω be sufficiently small such that the expression $\frac{\kappa_\Delta/\gamma_\Delta-\rho_1+\mathcal{O}(\xi)}{\omega} \leq 1$. We conclude that in this case $\vartheta\|u-\tilde{u}^{\delta}\|_X + \eta(\delta)$ is reduced by at least a factor $\max(\rho_1, 2\vartheta C + \frac{1}{2}) < 1$.

Next we consider the other case that

$$\eta(\delta) \leq \omega\|\tilde{\mathbf{r}}^{\delta}\|, \tag{6.43}$$

so that

$$\|\tilde{\mathbf{r}}^\delta|_J\| \geq \vartheta \|\tilde{\mathbf{r}}^\delta\|. \tag{6.44}$$

We have

$$\begin{aligned}\|\mathbf{r}^\delta - \tilde{\mathbf{r}}^\delta\| \lesssim \|u^\delta - \tilde{u}^\delta\|_X \lesssim t_\delta \overset{(6.36)}{\leq} \xi(\|\tilde{\mathbf{r}}^\delta\| + \eta(\delta)) \overset{(6.43)}{\lesssim} \xi\|\tilde{\mathbf{r}}^\delta\| \\ \leq \xi(\|\mathbf{r}^\delta\| + \|\mathbf{r}^\delta - \tilde{\mathbf{r}}^\delta\|),\end{aligned} \tag{6.45}$$

and so by taking ξ sufficiently small and kicking back $\|\mathbf{r}^\delta - \tilde{\mathbf{r}}^\delta\|$, also

$$\|\mathbf{r}^\delta - \tilde{\mathbf{r}}^\delta\| \lesssim \xi\|\mathbf{r}^\delta\|, \tag{6.46}$$

which together with (6.43) implies that

$$\eta(\delta) \lesssim \omega\|\mathbf{r}^\delta\|. \tag{6.47}$$

From (6.44)-(6.46) we infer that

$$\|\mathbf{r}^\delta\| \overset{(6.45)}{\lesssim} \|\tilde{\mathbf{r}}^\delta\| \overset{(6.44)}{\lesssim} \|\tilde{\mathbf{r}}^\delta|_J\| \leq \|\mathbf{r}^\delta|_J\| + \|\mathbf{r}^\delta - \tilde{\mathbf{r}}^\delta\| \overset{(6.46)}{\lesssim} \|\mathbf{r}^\delta|_J\| + \xi\|\mathbf{r}^\delta\|.$$

By taking ξ sufficiently small and kicking back $\|\mathbf{r}^\delta\|$, we conclude that there exists a constant $\tilde{\vartheta} > 0$ such that

$$\|\mathbf{r}^\delta|_J\| \geq \tilde{\vartheta}\|\mathbf{r}^\delta\|.$$

Assuming that $\frac{\kappa_\Delta}{\gamma_\Delta} - 1$ and ω are small enough, using (6.47) an application of Proposition 6.4.6 shows that there exists a constant $\rho_3 < 1$ such that

$$\|^{\tilde{\delta}}u - u^{\tilde{\delta}}\|_X \leq \rho_3 \|^\delta u - u^\delta\|_X. \tag{6.48}$$

Furthermore we have

$$\begin{aligned}\|\tilde{\mathbf{r}}^\delta\| \leq \|\mathbf{r}^\delta\| + \|\tilde{\mathbf{r}}^\delta - \mathbf{r}^\delta\| \lesssim \|^\delta u - u^\delta\|_X + t_\delta \lesssim \|u - \tilde{u}^\delta\|_X + \eta(\delta) + t_\delta \\ \overset{(6.37)}{\lesssim} \|u - \tilde{u}^\delta\|_X + \eta(\delta),\end{aligned}$$

and so from (6.43) by kicking back $\eta(\delta)$ and taking ω sufficiently small,

$$\eta(\delta) \lesssim \omega\|u - \tilde{u}^\delta\|_X. \tag{6.49}$$

We conclude that

$$\begin{aligned}
\vartheta\|u-\tilde{u}^{\tilde{\delta}}\|_X+\eta(\tilde{\delta}) &\le \vartheta\|^{\tilde{\delta}}u-u^{\tilde{\delta}}\|_X+\mathcal{O}(\eta(\tilde{\delta})+t_{\tilde{\delta}})\\
&\overset{(6.38)}{\le} (\vartheta+\mathcal{O}(\xi))\|^{\tilde{\delta}}u-u^{\tilde{\delta}}\|_X+\mathcal{O}(\eta(\tilde{\delta}))\\
&\overset{(6.48)}{\le} (\vartheta+\mathcal{O}(\xi))\rho_3\|^{\delta}u-u^{\delta}\|_X+\mathcal{O}(\eta(\tilde{\delta}))\\
&\le (\vartheta+\mathcal{O}(\xi))\rho_3\big[\|u-\tilde{u}^{\delta}\|_X+\mathcal{O}(\eta(\delta))+t_\delta\big]+\mathcal{O}(\eta(\tilde{\delta}))\\
&\overset{(6.37)}{\le} (\vartheta+\mathcal{O}(\xi))\rho_3\big[(1+\mathcal{O}(\xi))\|u-\tilde{u}^{\delta}\|_X+\mathcal{O}(\eta(\delta))\big]+\mathcal{O}(\eta(\delta))\\
&\overset{(6.49)}{\le} \big[(\vartheta+\mathcal{O}(\xi))\rho_3(1+\mathcal{O}(\xi))+\mathcal{O}(\omega)\big]\|u-\tilde{u}^{\delta}\|_X\\
&= \big[\rho_3+\mathcal{O}(\tfrac{\omega+\xi}{\vartheta})\big]\vartheta\|u-\tilde{u}^{\delta}\|_X.
\end{aligned}$$

So for ω/ϑ and ξ/ϑ sufficiently small, also in this case we established a reduction of $\vartheta\|u-\tilde{u}^{\tilde{\delta}}\|_X+\eta(\tilde{\delta})$ by at least a constant factor less than 1.

What is left to show is that the algorithm terminates. We have shown that the value of $\vartheta\|u-\tilde{u}^{\delta}\|_X+\eta(\delta)$ at passing the until-clause is r-linearly converging. We consider the corresponding value of $e_\delta=\|\tilde{\mathbf{r}}^{\delta}\|+\eta(\delta)+t_\delta$. Arguments that we have used multiple times show that $\|\tilde{\mathbf{r}}^{\delta}\|\lesssim\|u-\tilde{u}^{\delta}\|_X+\eta(\delta)+t_\delta$, and so $e_\delta\lesssim\|u-\tilde{u}^{\delta}\|_X+\eta(\delta)+t_\delta$. Using that $t_\delta\le\xi(\|\tilde{\mathbf{r}}^{\delta}\|+\eta(\delta))\le\xi e_\delta$, for ξ sufficiently small kicking back e_δ shows that

$$e_\delta\lesssim\|u-\tilde{u}^{\delta}\|_X+\eta(\delta),$$

and so

$$e_\delta\lesssim\vartheta\|u-\tilde{u}^{\delta}\|_X+\eta(\delta).$$

This last statement implies that at some moment $e_\delta\le\varepsilon$, meaning that the algorithm stops. □

6.5 Wavelets-in-time tensorized with finite-elements-in-space

We specify the parabolic problem at hand, as well as the type of families $(X^\delta)_{\delta\in\Delta}$ and $(Y^\delta)_{\delta\in\Delta}$ of 'trial' and 'test' spaces. A likely harmless minor further restriction to these families that will be needed for the construction of an X-stable collection Θ_δ that spans an X-stable complement space of X^δ in $X^{\underline{\delta}}$, specifically condition (6.24), will be postponed to Sect. 6.6.1.

6.5.1 Continuous problem

For some bounded domain $\Omega\subset\mathbb{R}^d$, we take $H=L_2(\Omega)$ and, for some closed $\emptyset\subseteq\Gamma_D\subseteq\partial\Omega$, $V=H^1_{0,\Gamma_D}(\Omega):=\mathrm{clos}_{H^1(\Omega)}\{u\in C^\infty(\Omega)\cap H^1(\Omega)\colon u|_{\Gamma_D}=0\}$,

and

$$a(t;\eta,\zeta) := \int_\Omega \mathbf{K}\nabla\eta\cdot\nabla\zeta + c\eta\zeta\,d\mathbf{x},$$

where $\mathbf{K} = \mathbf{K}^\top \in L_\infty(I\times\Omega)$ with $\mathbf{K}(\cdot) \eqsim \mathrm{Id}$ a.e., and $c \in L_\infty(I\times\Omega)$.

W.l.o.g. we take $T = 1$.

6.5.2 Wavelets in time

We will construct the trial and test spaces as the span of wavelets-in-time tensorized with finite element spaces-in-space. In this subsection we collect some assumptions on the wavelets.

At the *'trial side'* we consider a countable collection $\Sigma = \{\sigma_\lambda\colon \lambda\in\vee_\Sigma\}$ of functions $I\to\mathbb{R}$ known as *wavelets*. To each $\lambda\in\vee_\Sigma$ we associate a value $|\lambda|\in\mathbb{N}_0$, called the *level* of λ. We assume that the wavelets are *locally supported* meaning that $\sup_{n\in\mathbb{N},\,\ell\in\mathbb{N}_0}\#\{\lambda\in\vee_\Sigma\colon |\lambda|=\ell,\,|\operatorname{supp}\sigma_\lambda\cap 2^{-\ell}(n+[0,1])|>0\}<\infty$ and $\operatorname{diam}\operatorname{supp}\sigma_\lambda\lesssim 2^{-|\lambda|}$. To each $\lambda\in\vee_\Sigma$ with $|\lambda|>0$, we associate one or more $\tilde\lambda\in\vee_\Sigma$ with $|\tilde\lambda| = |\lambda|-1$ and $|\operatorname{supp}\sigma_\lambda\cap\operatorname{supp}\sigma_{\tilde\lambda}|>0$ which we call the *parent(s)* of λ. We denote this relation between a parent $\tilde\lambda$ and a child λ by

$$\tilde\lambda\lhd_\Sigma\lambda.$$

The definitions of parents and children give rise to obvious notions of ancestors and descendants.

To each $\lambda\in\vee_\Sigma$ we associate some neighbourhood $\mathcal{S}_\Sigma(\lambda)$ of $\operatorname{supp}\sigma_\lambda$ with $\operatorname{diam}\mathcal{S}_\Sigma(\lambda)\lesssim 2^{-|\lambda|}$ and

$$\tilde\lambda\lhd_\Sigma\lambda\Longrightarrow\mathcal{S}_\Sigma(\tilde\lambda)\supseteq\mathcal{S}_\Sigma(\lambda).$$

For some wavelets bases, e.g. Alpert wavelets ([Alp93]), it suffices to take $\mathcal{S}_\Sigma(\lambda) = \operatorname{supp}\sigma_\lambda$. With $C_\Sigma := \sup_{\lambda\in\vee_\Sigma}2^{|\lambda|}\operatorname{diam}\operatorname{supp}\sigma_\lambda$, a neighbourhood that in any case is sufficiently large is $\{t\in I\colon \operatorname{dist}(t,\operatorname{supp}\sigma_\lambda)\le C_\Sigma 2^{-|\lambda|}\}$. Indeed, if with this definition $t\in\mathcal{S}_\Sigma(\lambda)$ and $\tilde\lambda\lhd_\Sigma\lambda$, then $\operatorname{dist}(t,\operatorname{supp}\sigma_{\tilde\lambda})\le\operatorname{dist}(t,\operatorname{supp}\sigma_\lambda)+\operatorname{diam}(\operatorname{supp}\sigma_\lambda)\le 2C_\Sigma 2^{-|\lambda|}$, i.e. $t\in\mathcal{S}_\Sigma(\tilde\lambda)$.

We assume that Σ is a Riesz basis for $L_2(I)$, and, when renormalized in $H^1(I)$-norm, it is a Riesz basis for $H^1(I)$. Although not essential, thinking of wavelets being (essentially) constructed by means of *dilation*, we assume that

$$\|\sigma_\lambda\|_{H^1(I)}\eqsim 2^{|\lambda|}.$$

At the *'test side'* we consider a similar collection $\Psi = \{\psi_\mu\colon \mu\in\vee_\Psi\}$ of wavelets, with the difference though that this one has to be an even *orthonormal* basis for $L_2(\Omega)$, whilst, renormalized in $H^1(I)$-norm, it does not need to be a Riesz basis for $H^1(I)$.

We will assume that Σ and Ψ are selected such that for any $\ell \in N_0$,

$$\operatorname{span}\{\sigma_\lambda \colon |\lambda| \le \ell\} \cup \operatorname{span}\{\sigma'_\lambda \colon |\lambda| \le \ell\} \subseteq \operatorname{span}\{\psi_\mu \colon |\mu| \le \ell\},$$

so that in particular

$$|\mu| > |\lambda| \Longrightarrow \langle \sigma_\lambda, \psi_\mu \rangle_{L_2(I)} = 0 = \langle \sigma'_\lambda, \psi_\mu \rangle_{L_2(I)}. \tag{6.50}$$

6.5.3 Uniform stability

In the following proposition, we further specify the type of families of trial and test spaces that we consider, and formulate sufficient conditions for the requirements (6.13)-(6.14), which implied uniform stability of the Galerkin discretizations of our saddle-point problem (6.9).

Proposition 6.5.1. *For $\delta \in \Delta$, let*

$$X^\delta = \sum_{\lambda \in \vee_\Sigma} \sigma_\lambda \otimes W_\lambda^\delta, \quad Y^\delta = \sum_{\mu \in \vee_\Psi} \psi_\mu \otimes V_\mu^\delta$$

for subspaces $W_\lambda^\delta, V_\mu^\delta \subseteq V$ of which finitely many are non-zero. Let

$$\langle \sigma_\lambda, \psi_\mu \rangle_{L_2(I)} \neq 0 \Longrightarrow V_\mu^\delta \supseteq W_\lambda^\delta, \tag{6.51}$$

and, for some constant $\gamma_\Delta > 0$, for any $\mu \in \vee_\Psi$,

$$\inf_{0 \neq w \in \sum_{\{\lambda \in \vee_\Sigma \colon \langle \sigma'_\lambda, \psi_\mu \rangle_{L_2(I)} \neq 0\}} W_\lambda^\delta} \sup_{0 \neq v \in V_\mu^\delta} \frac{w(v)}{\|w\|_{V'} \|v\|_V} \ge \gamma_\Delta. \tag{6.52}$$

Then $X^\delta \subseteq Y^\delta$ and

$$\inf_{0 \neq w \in X^\delta} \sup_{0 \neq v \in Y^\delta} \frac{(\partial_t w)(v)}{\|\partial_t w\|_{Y'} \|v\|_Y} \ge \gamma_\Delta,$$

i.e., the conditions (6.13)-(6.14) *for uniform stability are satisfied.*

Proof. For $w_\lambda \in W_\lambda^\delta$ and $w := \sum_{\lambda \in \vee_\Sigma} \sigma_\lambda \otimes w_\lambda \in X^\delta$, $\sigma_\lambda = \sum_{\mu \in \vee_\Psi} \langle \sigma_\lambda, \psi_\mu \rangle_{L_2(I)} \psi_\mu$ shows that $w = \sum_{\mu \in \vee_\Psi} \psi_\mu \otimes \sum_{\lambda \in \vee_\Sigma} \langle \sigma_\lambda, \psi_\mu \rangle_{L_2(I)} w_\lambda \in Y^\delta$ by the first assumption. Similarly $\partial_t w = \sum_{\mu \in \vee_\Psi} \psi_\mu \otimes \tilde{v}_\mu$ where $\tilde{v}_\mu := \sum_{\lambda \in \vee_\Sigma} \langle \sigma'_\lambda, \psi_\mu \rangle_{L_2(I)} w_\lambda$. For any $\varepsilon > 0$, the second assumption shows that for any $\mu \in \vee_\Psi$, there exists a $v_\mu \in V_\mu^\delta$ with $\tilde{v}_\mu(v_\mu) \ge (\gamma_\Delta - \varepsilon) \|\tilde{v}_\mu\|_{V'} \|v_\mu\|_V$ and $\|\tilde{v}_\mu\|_{V'} = \|v_\mu\|_V$. With $v := \sum_{\mu \in \vee_\Psi} \psi_\mu \otimes v_\mu \in Y^\delta$, we infer that $(\partial_t w)(v) = \sum_{\mu \in \vee_\Psi} \tilde{v}_\mu(v_\mu) \ge (\gamma_\Delta - \varepsilon) \|\partial_t w\|_{Y'} \|v\|_Y$. □

In order to be able to apply at linear cost the arising linear operators in (6.31)-(6.32), we will restrict the type of trial spaces $X^\delta = \sum_{\lambda\in\vee_\Sigma} \sigma_\lambda \otimes W_\lambda^\delta$ by imposing the following *tree condition*

$$\tilde\lambda \lhd_\Sigma \lambda \Longrightarrow W_{\tilde\lambda}^\delta \supseteq W_\lambda^\delta. \tag{6.53}$$

For the same reason the analogous condition will be needed for Y^δ. For X^δ that satisfies (6.53), below the latter will be verified, and sufficient, more easily verifiable conditions for (6.51)-(6.52) are derived.

Proposition 6.5.2. *Let $X^\delta = \sum_{\lambda\in\vee_\Sigma} \sigma_\lambda \otimes W_\lambda^\delta$ satisfy* (6.53)*. For $\mu \in \vee_\Psi$, set*

$$\hat W_\mu^\delta := \sum_{\{\lambda\in\vee_\Sigma:\ |\lambda|=|\mu|,\, |\mathcal{S}_\Psi(\mu)\cap\mathcal{S}_\Sigma(\lambda)|>0\}} W_\lambda^\delta.$$

Build $Y^\delta = \sum_{\mu\in\vee_\Psi} \psi_\mu \otimes V_\mu^\delta$ by taking $V_\mu^\delta = \{0\}$ when $\hat W_\mu^\delta = \{0\}$, and otherwise

$$V_\mu^\delta \supseteq \hat W_\mu^\delta$$

where

$$\inf_{0\neq w\in\hat W_\mu^\delta} \sup_{0\neq v\in V_\mu^\delta} \frac{w(v)}{\|w\|_{V'}\|v\|_V} \geq \gamma_\Delta \tag{6.54}$$

for some constant $\gamma_\Delta > 0$. Then the conditions (6.51) *and* (6.52) *from Proposition 6.5.1 for uniform stability are satisfied.*

When $\dim V_\mu^\delta \lesssim \dim \hat W_\mu^\delta$, then $\dim Y^\delta \lesssim \dim X^\delta$, and under the natural condition that a larger $\hat W_\mu^\delta$ gives rise to a larger (more precisely, not smaller) V_μ^δ, the constructed Y^δ satisfies the tree condition

$$\tilde\mu \lhd_\Psi \mu \Longrightarrow V_{\tilde\mu} \supseteq V_\mu. \tag{6.55}$$

Proof. Let $\langle\sigma_\lambda, \psi_\mu\rangle_{L_2(I)} \neq 0$ or $\langle\sigma'_\lambda, \psi_\mu\rangle_{L_2(I)} \neq 0$. Then $|\mathcal{S}_\Sigma(\lambda)\cap\mathcal{S}_\Psi(\mu)| > 0$ and $|\lambda| \geq |\mu|$ by (6.50). When $|\lambda| > |\mu|$, λ has an ancestor $\tilde\lambda$ with $|\tilde\lambda| = |\mu|$, $W_{\tilde\lambda}^\delta \supseteq W_\lambda^\delta$, and $\mathcal{S}_\Sigma(\tilde\lambda) \supseteq \mathcal{S}_\Sigma(\lambda)$, and thus $|\mathcal{S}_\Sigma(\tilde\lambda)\cap\mathcal{S}_\Psi(\mu)| > 0$. We conclude that both $\sum_{\{\lambda\in\vee_\Sigma:\ \langle\sigma_\lambda,\psi_\mu\rangle_{L_2(I)}\neq 0\}} W_\lambda^\delta$ and $\sum_{\{\lambda\in\vee_\Sigma:\ \langle\sigma'_\lambda,\psi_\mu\rangle_{L_2(I)}\neq 0\}} W_\lambda^\delta$ are included in $\hat W_\mu^\delta$, so that (6.51) and (6.52) are guaranteed by the selection of V_μ^δ.

The statement $\dim Y^\delta \lesssim \dim X^\delta$ when $\dim V_\mu^\delta \lesssim \dim \hat W_\mu^\delta$ follows from $\dim \hat W_\mu^\delta \leq \sum_{\{\lambda\in\vee_\Sigma:\ |\lambda|=|\mu|,\, |\mathcal{S}_\Psi(\mu)\cap\mathcal{S}_\Sigma(\lambda)|>0\}} \dim W_\lambda^\delta$, and the fact that for any $\lambda \in \vee_\Sigma$, the number of $\mu \in \vee_\Psi$ with $|\mu| = |\lambda|$ and $|\mathcal{S}_\Psi(\mu)\cap\mathcal{S}_\Sigma(\lambda)| > 0$ is uniformly bounded.

Let $\tilde\mu \lhd_\Psi \mu$, and so $\mathcal{S}_\Psi(\tilde\mu) \supseteq \mathcal{S}_\Psi(\mu)$. For each $\lambda \in \vee_\Sigma$ with $|\lambda| = |\mu|$ and $|\mathcal{S}_\Psi(\mu)\cap\mathcal{S}_\Sigma(\lambda)| > 0$, there exists a $\tilde\lambda \lhd_\Sigma \lambda$, thus with $\mathcal{S}_\Sigma(\tilde\lambda) \supseteq \mathcal{S}_\Sigma(\lambda)$, and $W_{\tilde\lambda}^\delta \supseteq W_\lambda^\delta$ by (6.53). We conclude that $\hat W_{\tilde\mu}^\delta \supseteq \hat W_\mu^\delta$, which completes the proof of (6.55). □

As follows from [DSW21, Thm. 3.10] (taking B to be the Riesz map $H \to H'$) condition (6.54) has the following equivalent formulation.

Proposition 6.5.3. *Condition* (6.54) *is equivalent to existence of a projector* $Q \in \mathcal{L}(V,V)$ *with* $\operatorname{ran} Q \subseteq V_\mu^\delta$, $\operatorname{ran} Q^* \supseteq \hat{W}_\lambda^\delta$, *and* $\|Q\|_{\mathcal{L}(V,V)} \leq \frac{1}{\gamma_\Delta}$.

6.5.4 Selection of the spatial approximation spaces as finite element spaces

We will select the spaces W_λ^δ from a collection $\mathcal{O}$ of finite element spaces in V, which collection is closed under taking (finite) sums, and for which

$$\inf_{W\in\mathcal{O}} \inf_{0\neq w\in W} \sup_{0\neq v\in W} \frac{w(v)}{\|w\|_{V'}\|v\|_V} > 0. \tag{6.56}$$

Consequently, the stability conditions (6.13)-(6.14) are satisfied for some $\gamma_\Delta > 0$ by simply taking in Proposition 6.5.2.

$$V_\mu^\delta := \hat{W}_\mu^\delta \in \mathcal{O}. \tag{6.57}$$

As follows from Proposition 6.5.3, (6.56) is equivalent to uniform boundedness w.r.t. the norm on V of the H-orthogonal projector onto $W \in \mathcal{O}$. It is well-known that an example of such a collection $\mathcal{O}$ is given by the set of all finite element spaces W_λ^δ w.r.t. quasi-uniform, uniformly shape regular conforming partitions of Ω into, say, d-simplices.

It is known that the uniform boundedness w.r.t. the V-norm of the H-orthogonal projector holds also true for finite element spaces w.r.t. *locally refined* partitions as long as the grading of the partitions is sufficiently mild. In [GHS16] it has been shown that for $d = 2$ spatial dimensions, and polynomial orders up to 12, the collection of all conforming partitions that can be generated by *newest vertex bisection* (NVB), starting from a fixed conforming initial partition $\mathcal{T}_\perp$ with an assignment of the newest vertices that satisfies a so-called matching condition, is sufficiently mildly graded in the above sense. Since the overlay of two conforming NVB partitions is a conforming NVB partition, this collection is closed under taking (finite) sums. In other words, with this collection of finite element spaces, which we will employ in our experiments, again the choice (6.57) guarantees uniform stability.

In [Car04] a result similar to that from [GHS16] has been shown for red-blue-green refinement and lowest order finite elements again for $d = 2$. Unfortunately, for $d > 2$ such results seem not yet to be available.

Remark 6.5.4 (Getting γ_Δ close to 1). We discussed uniform boundedness w.r.t. the V-norm of the H-orthogonal projectors onto a family of finite element spaces, which, by taking $V_\mu^\delta := \hat{W}_\mu^\delta$ in Proposition 6.5.2, yields the uniform inf-sup condition (6.14) *some* value $\gamma_\Delta > 0$, and so uniform stability of the Galerkin discretizations of the saddle-point (6.9).

For proving convergence of our adaptive routine Algorithm 6.4.8, however, we needed a value of $\gamma_\Delta > 0$ that is sufficiently close to 1. Although in our numerical experiments, reported on in Sect. 6.7, with continuous piecewise linear finite element spaces generated by conforming NVB and $V_\mu^\delta = \hat{W}_\mu^\delta$, the adaptive routine is r-linearly converging, there is no guarantee that $1 - \gamma_\Delta$ is sufficiently small.

Restricting to quasi-uniform partitions, below we show that $1 - \gamma_\Delta$ can be made arbitrarily small by taking the mesh underlying V_μ^δ to be a sufficiently deep, but fixed *refinement* of the mesh underlying $\hat{W}_\mu^\delta$. One may conjecture that the same result holds true for sufficiently mildly graded locally refined meshes.

Let the diameters of any d-simplex in the partitions underlying $\hat{W}_\mu^\delta$ and V_μ^δ be proportional to h_c and h_f, respectively. For $s \in [0,1]$, let $\mathcal{H}^s := [H, V]_{s,2}$. In any case when Ω is a Lipschitz domain, it is known that there exists an $s \in (0,1]$ such that the solution $u \in V$ of $\langle u, v\rangle_V = f(v)$ $(v \in V)$ satisfies $\|u\|_{H^{1+s}(\Omega)} \lesssim \|f\|_{(\mathcal{H}^{1-s})'}$, assuming the right-hand side is bounded. From this, the Aubin-Nitsche duality argument shows that the V-orthogonal projector P_μ onto V_μ^δ satisfies $\|\mathrm{Id} - P_\mu\|_{\mathcal{L}(V,\mathcal{H}^{1-s})} \lesssim h_f^s$. On the other hand, on $\hat{W}_\mu^\delta$ we have the following inverse inequality $\|\cdot\|_{(\mathcal{H}^{1-s})'} \lesssim h_c^{-s}\|\cdot\|_{V'}$ (e.g. (2.44)).

Given $w \in \hat{W}_\mu^\delta$, for any $\varepsilon > 0$, there exists a $v \in V$ with $w(v) \geq (1-\varepsilon)\|w\|_{V'}\|v\|_V$. We infer that, for some constant $C > 0$,

$$\begin{aligned} w(P_\mu v) &= w(v) + w((\mathrm{Id} - P_\mu)v) \\ &\geq (1-\varepsilon)\|w\|_{V'}\|v\|_V - \|w\|_{(\mathcal{H}^{1-s})'}\|(\mathrm{Id} - P_\mu)v\|_{\mathcal{H}^{1-s}} \\ &\geq (1 - (\varepsilon + C(h_f/h_c)^s)\|w\|_{V'}\|v\|_V \\ &\geq (1 - (\varepsilon + C(h_f/h_c)^s)\|w\|_{V'}\|P_\mu v\|_V. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, we conclude that $\gamma_\Delta \geq (1 - C(h_f/h_c)^s)$ which proves our assertion.

6.5.5 Best possible rates

Although so far we have not proved it, we expect that the sequence of approximations generated by our adaptive Algorithm 6.4.8 is not only r-linearly converging, but, ignoring data oscillation, that it is a sequence of approximations from a sequence of spaces from the family $(X^\delta)_{\delta\in\Delta}$ that converges with the best possible rate. In this subsection, we show that with our selection of the $(X^\delta)_{\delta\in\Delta}$, under some (mild) smoothness conditions on the solution u this best possible rate *equals* the rate of best approximation to the solution of the corresponding stationary problem from the spatial finite element spaces w.r.t. the V-norm.

Consider a family of spaces $X^\delta = \sum_{\lambda\in\vee_\Sigma} \sigma_\lambda \otimes W_\lambda^\delta$ that satisfies (6.53), with the W_λ^δ selected from a collection of finite element spaces $\mathcal{O}$ that in any

case contains all such spaces that correspond to uniform refinements of some initial partition of Ω. Let Σ a collection of wavelets of order d_t, and assume that the finite element spaces are of order d_x. When for each X^δ, the space Y^δ is selected as in Proposition 6.5.2, then the combination of (6.20) and the analysis from [SS09, Sect. 7.1] shows that if the exact solution u of our parabolic problem satisfies the mixed regularity condition $u \in H^{d_t}(I) \otimes H^{d_x}(\Omega)$, then a suitable (non-adaptive) choice of the spaces W_λ^δ yields a sequence of solutions $u^{\hat{\delta}\delta} \in X^\delta$ (for arbitrary $Y^{\hat{\delta}} \supset Y^\delta$) of the modified discretized saddle-point from Sect 6.3.3, for which

$$\|u - u^{\hat{\delta}\delta}\|_X \lesssim (\dim X^\delta)^{-\min(d_t - 1, \frac{d_x-1}{d})}.$$

Note that for $d_t - 1 \geq \frac{d_x-1}{d}$, the rate $\frac{d_x-1}{d}$ equals the best rate in the V-norm that can be expected when the finite element spaces are employed for solving the corresponding *stationary* problem, which is posed on a d-dimensional domain instead over the $d+1$-dimensional space-time cylinder.

For an optimal *adaptive* choice of the W_λ^δ as finite element spaces w.r.t. a sufficiently 'rich' collection of locally refined partitions, as the collection of all conforming NVB partitions, it can be expected that the rate $\min(d_t - 1, \frac{d_x-1}{d})$ is realized under *much* milder regularity conditions on u.

When instead of being finite element spaces, the spaces W_λ^δ can be selected as the spans of some wavelets from a Riesz basis for V of order d_x, and additionally the tree condition (6.53) is dropped, a precise characterization of those u that can be approximated at a rate $s < \min(d_t - 1, \frac{d_x-1}{d})$ in terms of tensor products of Besov spaces can be deduced from [Nit06, SU09]. The collection of finite element spaces w.r.t. locally refined meshes, as those generated by NVB, is very resemblant to the collection of spans of sets of such wavelets when on these sets a tree condition is imposed similar to the tree constraint (6.53) that we imposed in the temporal direction. In other words, the collection of spaces X^δ that we consider is similar to the collection of spans of sets of tensor products of temporal and spatial wavelets when these sets satisfy a 'double-tree' constraint. In view of results from [BDDP02], we do not expect that this constraint makes the resulting approximation classes much smaller.

6.5.6 Preconditioners

Our adaptive solution method of the parabolic problem requires optimal preconditioners for ${E_Y^\delta}' A E_Y^\delta$ and $S^{\delta\delta}$, i.e., for both $Z = Y$ and $Z = X$ and $\delta \in \Delta$, we need operators $K_Z^\delta = {K_Z^\delta}' \in \mathcal{L}(Z^{\delta'}, Z^\delta)$ with $h(K_Z^\delta h) \eqsim \|h\|_{Z^{\delta'}}^2$ ($h \in Z^{\delta'}$), moreover which should be applicable at linear cost.

To construct these preconditioners, for $Z \in \{Y, X\}$ we will select a symmetric, bounded, and coercive bilinear form on $Z \times Z$, and after selecting *some* basis for Z^δ, we will construct a matrix $\mathbf{K}_Z^\delta = {\mathbf{K}_Z^\delta}^\top$ that can be applied in linear complexity, and that is *uniformly* spectrally equivalent to the inverse

of the *stiffness matrix* corresponding to this bilinear form (being the matrix representation of the linear mapping $Z^\delta \to Z^{\delta\prime}$ defined by the bilinear form w.r.t. the basis for Z^δ being chosen and the corresponding dual basis for $Z^{\delta\prime}$). Then $K_Z^\delta \in \mathcal{L}(Z^{\delta\prime}, Z^\delta)$, defined as the operator whose matrix representation is $\mathbf{K}_Z^\delta$ w.r.t. the aforementioned bases of $Z^{\delta\prime}$ and Z^δ, is the preconditioner that satisfies our needs.

Notice that the choice of the basis for Z^δ is irrelevant. Indeed, denoting the aforementioned stiffness matrix as $\mathbf{C}_Z^\delta$ with corresponding operator $C_Z^\delta = C_Z^{\delta\prime} \in \mathcal{L}\text{is}(Z^\delta, Z^{\delta\prime})$, one may verify that

$$\begin{aligned}\|K_Z^\delta\|_{\mathcal{L}(Z^{\delta\prime},Z^\delta)}\|(K_Z^\delta)^{-1}\|_{\mathcal{L}(Z^\delta,Z^{\delta\prime})} &\eqsim \|K_Z^\delta C_Z^\delta\|_{\mathcal{L}(Z^\delta,Z^\delta)}\|(K_Z^\delta C_Z^\delta)^{-1}\|_{\mathcal{L}(Z^\delta,Z^\delta)}\\ &= \frac{\lambda_{\max}(\mathbf{K}_Z^\delta \mathbf{C}_Z^\delta)}{\lambda_{\min}(\mathbf{K}_Z^\delta \mathbf{C}_Z^\delta)}.\end{aligned}$$

Preconditioner at the 'test side'

Let $Y = Z$. Since Ψ is an orthonormal basis for $L_2(I)$, any $y \in Y$ is of the form $\sum_{\mu\in\vee_\Psi} \psi_\mu \otimes v_\mu$ for some $v_\mu \in V$ with $\sum_{\mu\in\vee_\Psi} \|v_\mu\|_V^2 < \infty$. Taking as bilinear form on $Y \times Y$ simply the scalar product on $Y \times Y$, we have

$$\langle \sum_{\mu_1\in\vee_\Psi} \psi_{\mu_1} \otimes v_{\mu_1}^{(1)}, \sum_{\mu_2\in\vee_\Psi} \psi_{\mu_2} \otimes v_{\mu_2}^{(2)}\rangle_Y = \sum_{\mu\in\vee_\Psi} \langle v_\mu^{(1)}, v_\mu^{(2)}\rangle_V.$$

Equipping $Y^\delta = \sum_{\mu\in\vee_\Psi} \psi_\mu \otimes V_\mu^\delta$ with a basis of type $\cup_{\mu\in\vee_\Psi}\psi_\mu \otimes \Phi_\mu^\delta$, the resulting stiffness matrix reads as $\mathrm{blockdiag}[\mathbf{A}_\mu^\delta]_{\mu\in\vee_\Psi}$, where $\mathbf{A}_\mu^\delta = \langle\Phi_\mu^\delta, \Phi_\mu^\delta\rangle_V$ is the stiffness matrix of $\langle\cdot,\cdot\rangle_V$ w.r.t. Φ_μ^δ. Selecting $\mathbf{K}_\mu^\delta \eqsim (\mathbf{A}_\mu^\delta)^{-1}$, the matrix representation of the optimal preconditioner reads as

$$\mathbf{K}_Y^\delta = \mathrm{blockdiag}[\mathbf{K}_\mu^\delta]_{\mu\in\vee_\Psi}.$$

It is well-known that when V_μ^δ is a finite element space, possibly w.r.t. a locally refined partition, suitable $\mathbf{K}_\mu^\delta$ of *multi-grid* type are available. These $\mathbf{K}_\mu^\delta$ can be applied in linear complexity, and so can $\mathbf{K}_Y^\delta$.

To show, in Theorem 6.4.9, that our adaptive Algorithm 6.4.8 is r-linearly converging we required $C_\Delta - 1$ to be sufficiently small, which requires that $\|(E_Y^{\delta\prime} A E_Y^\delta)^{-1} - K_Y^\delta\|_{\mathcal{L}(Y^{\delta\prime},Y^\delta)}$ or, equivalently, $\|\mathrm{Id} - K_Y^\delta E_Y^{\delta\prime} A E_Y^\delta\|_{\mathcal{L}(Y^\delta,Y^\delta)}$ is sufficiently small, i.e. the eigenvalues of $K_Y^\delta E_Y^{\delta\prime} A E_Y^\delta$ are sufficiently close to 1. Given an initial optimal, self-adjoint, and coercive preconditioner K_Y^δ, and some upper and lower bounds on the spectrum of the preconditioned system, one can satisfy the latter condition by polynomial acceleration using Chebychev polynomials of sufficiently high degree. In our numerical experiments, it turned out that it was not needed to apply this 'acceleration'.

Preconditioner at the 'trial side'

The preconditioner presented in this section is inspired by constructions of preconditioners in [And16, NS19] for parabolic problems discretized on a tensor product of temporal and spatial spaces.

Thanks to Σ and $\{2^{-|\lambda|}\sigma_\lambda\colon \lambda\in \vee_\Sigma\}$ being Riesz bases for $L_2(I)$ and $H^1(I)$, any $x\in X$ is of the form $\sum_{\lambda\in\vee_\Sigma}\sigma_\lambda\otimes w_\lambda$ for some $w_\lambda\in V$ with $\sum_{\lambda\in\vee_\Sigma}\|w_\lambda\|_V^2+4^{|\lambda|}\|w_\lambda\|_{V'}^2<\infty$, and

$$\langle \sum_{\lambda_1\in\vee_\Sigma}\sigma_{\lambda_1}\otimes w_{\lambda_1}^{(1)},\sum_{\lambda_2\in\vee_\Sigma}\sigma_{\lambda_2}\otimes w_{\lambda_2}^{(2)}\rangle := \sum_{\lambda\in\vee_\Sigma}\langle w_\lambda^{(1)},w_\lambda^{(2)}\rangle_V+4^{|\lambda|}\langle w_\lambda^{(1)},w_\lambda^{(2)}\rangle_{V'}$$

is a symmetric, bounded, and coercive bilinear form on $X\times X$. Equipping $X^\delta=\sum_{\lambda\in\vee_\Sigma}\sigma_\lambda\otimes W_\lambda^\delta$ with a basis of type $\cup_{\lambda\in\vee_\Sigma}\sigma_\lambda\otimes\Phi_\lambda^\delta$, the resulting stiffness matrix reads as

$$\mathrm{blockdiag}[\mathbf{A}_\lambda^\delta+4^{|\lambda|}\langle\Phi_\lambda^\delta,\Phi_\lambda^\delta\rangle_{V'}]_{\lambda\in\vee_\Sigma}$$

where $\mathbf{A}_\lambda^\delta=\langle\Phi_\lambda^\delta,\Phi_\lambda^\delta\rangle_V$.

Thanks to our assumption (6.56), for $u\in W_\lambda^\delta$ it holds that $\|u\|_{V'}\lesssim\sup_{0\neq w\in W_\lambda^\delta}\frac{\langle u,w\rangle}{\|w\|_V}\le\|u\|_{V'}$. With $\mathbf{u}$ denoting the representation of u w.r.t. Φ_λ^δ, we have

$$\sup_{0\neq w\in W_\lambda^\delta}\frac{\langle u,w\rangle}{\|w\|_V}=\|(\mathbf{A}_\lambda^\delta)^{-\frac12}\mathbf{M}_\lambda^\delta\mathbf{u}\|,$$

where $\mathbf{M}_\lambda^\delta=\langle\Phi_\lambda^\delta,\Phi_\lambda^\delta\rangle$, so that

$$\langle\Phi_\lambda^\delta,\Phi_\lambda^\delta\rangle_{V'}\lesssim\mathbf{M}_\lambda^\delta(\mathbf{A}_\lambda^\delta)^{-1}\mathbf{M}_\lambda^\delta\le\langle\Phi_\lambda^\delta,\Phi_\lambda^\delta\rangle_{V'}.$$

Since both $\mathbf{A}_\lambda^\delta$ and $\mathbf{M}_\lambda^\delta$ are symmetric and positive definite, [PW12, Thm. 4] shows that

$$\begin{aligned}\tfrac12\big(\mathbf{A}_\lambda^\delta+4^{|\lambda|}\mathbf{M}_\lambda^\delta(\mathbf{A}_\lambda^\delta)^{-1}\mathbf{M}_\lambda^\delta\big)&\le(\mathbf{A}_\lambda^\delta+2^{|\lambda|}\mathbf{M}_\lambda^\delta)(\mathbf{A}_\lambda^\delta)^{-1}(\mathbf{A}_\lambda^\delta+2^{|\lambda|}\mathbf{M}_\lambda^\delta)\\&\le\mathbf{A}_\lambda^\delta+4^{|\lambda|}\mathbf{M}_\lambda^\delta(\mathbf{A}_\lambda^\delta)^{-1}\mathbf{M}_\lambda^\delta.\end{aligned}$$

Now assuming that

$$\mathbf{K}_\lambda^\delta\eqsim(\mathbf{A}_\lambda^\delta+2^{|\lambda|}\mathbf{M}_\lambda^\delta)^{-1},\tag{6.58}$$

we infer that

$$\mathbf{K}_X^\delta=\mathrm{blockdiag}\left[\mathbf{K}_\lambda^\delta\mathbf{A}_\lambda^\delta\mathbf{K}_\lambda^\delta\right]_\lambda$$

is the matrix representation of an optimal preconditioner.

Notice that (6.58) requires an optimal preconditioner of a discretized reaction-diffusion equation that is robust w.r.t. to the size of the (constant) reaction term. In [OR00] it was shown that, under a 'full-regularity' assumption, for quasi-uniform meshes multiplicative multi-grid yields such a preconditioner,

moreover whose application can be performed at linear cost. Although we expect that using the theory of subspace correction methods the full regularity assumption can be avoided, and furthermore that the optimality, robustness and linear complexity result extends to locally refined meshes, proofs of such extensions seem not to be available.

6.6 A concrete realization

6.6.1 The collection $\mathcal{O}$ of finite element spaces, and the map $\delta \to \underline{\delta}$

We further specify the collection $\mathcal{O}$ of finite element spaces, construct a linearly independent set in $H^1_{0,\Gamma_D}(\Omega)$, known as the hierarchical basis, and equip it with a tree structure such that there exists a 1-1 correspondence between the finite element spaces in $\mathcal{O}$, and the spans of subsets of the hierarchical basis that form trees.

With this specification of $\mathcal{O}$, there will be a 1-1 correspondence between the spaces $X^\delta = \sum_{\lambda\in\sigma_\lambda} \sigma_\lambda \otimes W^\delta_\lambda$ with $W^\delta_\lambda \in \mathcal{O}$ that satisfy (6.53), and the spans of collections of tensor products of wavelets σ_λ and hierarchical basis functions whose sets of index pairs are *lower*, also known as *downward closed.* Given such a X^δ, we will define $X^{\underline{\delta}}$ by a certain enlargement the lower set.

For $d \geq 2$, let $\mathbb{T}$ be the family of all *conforming* partitions of a polytope $\Omega \subset \mathbb{R}^d$ into (closed) d-simplices that can be created by NVB starting from some given conforming initial partition $\mathcal{T}_\perp$ with an assignment of the newest vertices that satisfies the matching condition, see [Ste08b]. We define a *partial order* on $\mathbb{T}$ by writing $\mathcal{T} \preceq \tilde{\mathcal{T}}$ when $\tilde{\mathcal{T}}$ is a refinement of $\mathcal{T}$.

With some small adaptations that we leave to the reader, in the following the case $d = 1$ can be included by letting $\mathbb{T}$ to be the family of a partitions of Ω into (closed) subintervals that can be constructed by bisections from $\mathcal{T}_\perp = \{\Omega\}$ such that the generations of any two neighbouring subintervals in any $\mathcal{T} \in \mathbb{T}$ differ by not more than one.

The collection $\mathcal{O}$ that we will consider is formed by the spaces $W = W_\mathcal{T}$ of *continuous piecewise linears* w.r.t. $\mathcal{T} \in \mathbb{T}$, zero on a possible Dirichlet boundary Γ_D being the union of $\partial T \cap \partial\Omega$ for some $T \in \mathcal{T}_\perp$. We expect that generalizations to finite element spaces of higher order do not impose essential difficulties.

For $T \in \mathfrak{T} := \cup_{\mathcal{T}\in\mathbb{T}}\{T\colon T \in \mathcal{T}\}$, we set $\mathrm{gen}(T)$ to be the number of bisections needed to create T from its 'ancestor' $T' \in \mathcal{T}_\perp$. With $\mathfrak{N}$ being the set of all vertices (or nodes) of all $T \in \mathfrak{T}$, for $\nu \in \mathfrak{N}$ we set $\mathrm{gen}(\nu) := \min\{\mathrm{gen}(T)\colon T \in \mathfrak{T},\, \nu \in T\}$.

Any $\nu \in \mathfrak{N}$ with $\mathrm{gen}(\nu) > 0$ is the midpoint of an edge of one or more $T \in \mathfrak{T}$ with $\mathrm{gen}(T) = \mathrm{gen}(\nu) - 1$. The vertices $\tilde{\nu}$ of these T with $\mathrm{gen}(\tilde{\nu}) = \mathrm{gen}(\nu) - 1$ are defined as the parents of ν. We denote this relation between a parent $\tilde{\nu}$ and a child ν by $\tilde{\nu} \lhd_\mathfrak{N} \nu$, see Figure 6.1. Vertices $\nu \in \mathfrak{N}$ with $\mathrm{gen}(\nu) = 0$ have no parents.

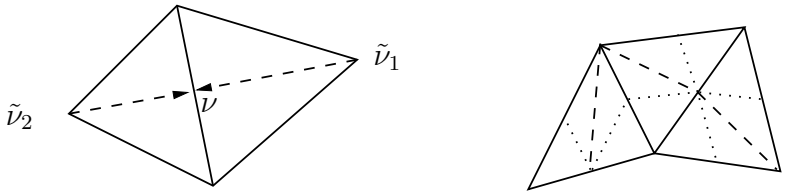


FIGURE 6.1. $\tilde\nu_1, \tilde\nu_2 \lhd_{\mathfrak{N}} \nu$, and $\mathcal{T}$ and its refinement $\mathcal{T}^{d+}$ (for $d = 2$).

An (essentially) non-overlapping partition $\mathcal{T}$ of $\overline{\Omega}$ into $T \in \mathfrak{T}$ is in $\mathbb{T}$ *if and only if* the set $N_{\mathcal{T}}$ of vertices of all $T \in \mathcal{T}$ forms a *tree*, meaning that it contains all $\nu \in \mathfrak{N}$ with $\mathrm{gen}(\nu) = 0$ as well as all parents of any $\nu \in N_{\mathcal{T}}$ with $\mathrm{gen}(\nu) > 0$, cf. [DKS16] for the $d = 2$ case.

Definition 6.6.1. For any $\mathcal{T} \in \mathbb{T}$, we define $\mathcal{T}^{d+} \in \mathbb{T}$ (denoted as $\mathcal{T}^{++}$ in [DKS16] for the $d = 2$ case) by replacing any $T \in \mathcal{T}$ by its 2^d 'descendants' of the dth generation, see Figure 6.1.

Since this refinement adds exactly one vertex at the midpoint on any edge of all $T \in \mathcal{T}$, one infers that indeed $\mathcal{T}^{d+} \in \mathbb{T}$. The corresponding tree $N_{\mathcal{T}^{d+}}$ is created from $N_{\mathcal{T}}$ by the addition of all descendants up to generation d of all $\nu \in N_{\mathcal{T}}$.[1]

For $\nu \in \mathfrak{N}$, we set ϕ_ν as the continuous piecewise linear function w.r.t. the *uniform partition* $\{T \in \mathfrak{T}\colon\ \mathrm{gen}(T) = \mathrm{gen}(\nu)\} \in \mathbb{T}$, which function is 1 at ν and 0 at all other vertices of this partition. Setting $\mathfrak{N}_0 := \mathfrak{N} \setminus \Gamma_D$ and, for any $\mathcal{T} \in \mathbb{T}$, $N_{\mathcal{T},0} := N_{\mathcal{T}} \setminus \Gamma_D$, the collection $\{\phi_\nu\colon \nu \in \mathfrak{N}_0\}$ is known as the *hierarchical basis*, and for any $\mathcal{T} \in \mathbb{T}$, it holds that $W_{\mathcal{T}} = \mathrm{span}\{\phi_\nu\colon \nu \in N_{\mathcal{T},0}\}$.

With above specification of the collection $\mathcal{O}$ of finite element spaces, there exists a 1-1 correspondence between the spaces $\sum_{\lambda\in\vee_\Sigma} \sigma_\lambda \otimes W_\lambda^\delta$ with $W_\lambda^\delta \in \mathcal{O}$ that satisfy (6.53), and the spaces of the form

$$X^\delta = \mathrm{span}\{\sigma_\lambda \otimes \phi_\nu\colon (\lambda,\nu) \in I_{\delta,0} := I_\delta \setminus (\vee_\Sigma \times \Gamma_D)\} \tag{6.59}$$

for some finite $I_\delta \subset \vee_\Sigma \times \mathfrak{N}$ being a *lower set* in the sense

$$(\lambda,\nu) \in I_\delta \text{ and } \begin{cases} \tilde\lambda \lhd_\Sigma \lambda & \Longrightarrow\ (\tilde\lambda,\nu) \in I_\delta, \\ \tilde\nu \lhd_{\mathfrak{N}} \nu \text{ or } \mathrm{gen}(\tilde\nu) = 0 & \Longrightarrow\ (\lambda,\tilde\nu) \in I_\delta. \end{cases} \tag{6.60}$$

For above specification of X^δ, from Proposition 6.5.2 with the specification (6.57) one infers that the corresponding space

$$Y^\delta = \mathrm{span}\{\psi_\mu \otimes \phi_\nu\colon (\mu,\nu) \in I_{\delta,0}^Y\}, \tag{6.61}$$

[1]The addition of only all children of all $\nu \in N_{\mathcal{T}}$ yields a tree only if $\mathcal{T}$ is a uniform partition.

where

$$I^Y_{\delta,0} := \{(\mu,\nu)\colon \exists(\lambda,\nu)\in I_{\delta,0},\ \mu\in\vee_\Psi,\ |\mu|=|\lambda|,\ |\mathcal{S}_\Psi(\mu)\cap\mathcal{S}_\Sigma(\lambda)|>0\} \tag{6.62}$$

which index set is a lower set.

Remark 6.6.2 (Complexity of matrix-vector multiplications). The fact that the index sets of the bases for X^δ and Y^δ are lower sets is the key why it is possible to compute residuals of the system $S^{\underline{\delta}\delta}u^\delta = f^\delta$ ((6.31)) in $\mathcal{O}(\dim X^\delta)$ operations. Indeed, when one has a bilinear form that is 'local' and equals the tensor product of bilinear forms in time and space, and two spaces spanned by tensor product multi-level bases corresponding to lower sets, then the resulting generalized system matrix w.r.t. both bases can be applied in a number of operations that is proportional to the sum of the dimensions of both spaces. The algorithm that realizes this complexity makes a clever use of multi- to single-scale transformations alternately in time and space. In a 'uniform' sparse-grid setting, i.e., without 'local refinements', this algorithm was introduced in [BZ96], and it was later extended to general lower sets in [KS14]. The definition of a lower set in [KS14], there called multi-tree, is more restrictive than our current definition that allows more localized refinements. Details about the matrix-vector multiplication and a proof of its optimal computational complexity is given in Chapter 7.

Definition 6.6.3. Given $X^\delta = \operatorname{span}\{\sigma_\lambda\otimes\hat\phi_\nu\colon(\lambda,\nu)\in I_{\delta,0}\}$ for some lower set $I_\delta\subset\vee_\Sigma\times\mathfrak{N}$, we define the lower set $I_{\underline{\delta}}$, and with that $X^{\underline{\delta}}$, by adding, for each $(\lambda,\nu)\in I_\delta$ and any child $\tilde\lambda$ of λ and any descendant $\tilde\nu$ of ν up to generation d, all pairs $(\tilde\lambda,\nu)$ and $(\lambda,\tilde\nu)$ to I_δ.

6.6.2 The collection Θ_δ such that $X^{\underline{\delta}} = X^\delta\oplus\Theta_\delta$

Recall that for the bulk chasing process we need an 'X-stable' basis Θ_δ that spans an 'X-stable' complement space of X^δ in $X^{\underline{\delta}}$, i.e., a collection that satisfies (6.24). For that goal we define a *modified hierarchical basis* $\{\hat\phi_\nu\colon\nu\in\mathfrak{N}_0\}$ by $\hat\phi_\nu := \phi_\nu$ when $\operatorname{gen}(\nu)=0$, and

$$\hat\phi_\nu := \phi_\nu - \frac{\sum_{\{\tilde\nu\in\mathfrak{N}_0\colon\tilde\nu\lhd_\mathfrak{N}\nu\}}\frac{\int_\Omega\phi_\nu\,dx}{\int_\Omega\phi_{\tilde\nu}\,dx}\phi_{\tilde\nu}}{\#\{\tilde\nu\in\mathfrak{N}\colon\tilde\nu\lhd_\mathfrak{N}\nu\}}$$

otherwise. Notice that for those ν with $\operatorname{gen}(\nu)>0$ that have all their parents not on Γ_D it holds that $\int_\Omega\hat\phi_\nu\,dx=0$, i.e., $\hat\phi_\nu$ has a *vanishing moment*, and furthermore that for any $\mathcal{T}\in\mathbb{T}$, $W_\mathcal{T}=\operatorname{span}\{\hat\phi_\nu\colon\nu\in N_{\mathcal{T},0}\}$.

For any $\mathcal{T}\in\mathbb{T}$, it holds that

$$W_\mathcal{T} = \operatorname{span}\{\phi_\nu\colon\nu\in N_{\mathcal{T},0}\} = \operatorname{span}\{\hat\phi_\nu\colon\nu\in N_{\mathcal{T},0}\},$$

and thus for any lower set $I_\delta \subset \vee_\Sigma \times \mathfrak{N}$,

$$X^\delta = \operatorname{span}\{\sigma_\lambda \otimes \hat{\phi}_\nu \colon (\lambda,\nu) \in I_{\delta,0}\} = \operatorname{span}\{\sigma_\lambda \otimes \phi_\nu \colon (\lambda,\nu) \in I_{\delta,0}\}.$$

Moreover, for any $\mathcal{T} \in \mathbb{T}$, the basis transformation from the modified to unmodified hierarchical basis for $W_\mathcal{T}$ can be applied in linear complexity traversing from the leaves to the roots.

Given δ, the collection Θ_δ will be the set of properly normalized functions $\sigma_\lambda \otimes \hat{\phi}_\nu$ for $(\lambda,\nu) \in I_{\underline{\delta},0} \setminus I_{\delta,0}$. In order to demonstrate (6.24), we have to impose some gradedness assumption on the lower sets I_δ.

Definition 6.6.4. The *gradedness constant* of a lower set $I_\delta \subset \vee_\Sigma \times \mathfrak{N}$ is the smallest $L_\delta \in \mathbb{N}$ such that for all $(\lambda,\nu) \in I_\delta$ for which ν has an ancestor $\tilde{\nu} \in \mathfrak{N}$ with $\operatorname{gen}(\nu) - \operatorname{gen}(\tilde{\nu}) = L_\delta$, it holds that $(\breve{\lambda},\tilde{\nu}) \in I_\delta$ for any child $\breve{\lambda} \in \vee_\Sigma$ of λ.

Remark 6.6.5 (Uniform boundedness of the gradedness constants). Under the (unproven) assumption that our adaptive method creates a sequence of spaces X^δ which are quasi-optimal for the approximation of the solution of the the parabolic PDE, one may hope that these spaces have a *uniformly bounded gradedness constant*, unless (locally) the solution u is extremely more smooth as function of t than as function of the spatial variables.

To see this, consider the non-adaptive sparse grid index sets of the form $\{(\lambda,\nu) \in \vee_\Sigma \times \mathfrak{N} \colon \tilde{L}|\lambda| + \operatorname{gen}(\nu) \le N\}$ for some constant $\tilde{L}$ and $N \in \mathbb{N}$, which are appropriate when the behaviour of u as function of t on the one hand and that of the spatial variables on the other is globally similar. Then for $\tilde{L} \le L$, the gradedness constant of this index set is $\le L$, where the smallest spatial resolution in the 'sparse-grid mesh' equals the smallest temporal resolution in this mesh to the power $\tilde{L}/d$. So only when a polynomial decay of the spatial resolution as function of the temporal resolution does not suffice for a proper approximation of u, one cannot expect to have a gradedness constant that is uniformly bounded.

Proposition 6.6.6. *For $(\lambda,\nu) \in \vee_\Sigma \times \mathfrak{N}_0$, let $e_{\lambda\mu} := 1/\sqrt{2^{(\frac{2}{d}-1)\operatorname{gen}(\nu)} + 4^{|\lambda|}2^{(-\frac{2}{d}-1)\operatorname{gen}(\nu)}}$ and $\theta_{\lambda\nu} := e_{\lambda\mu}\sigma_\lambda \otimes \hat{\phi}_\nu$. For any $\delta \in \Delta$, let $\Theta_\delta := \{\theta_{\lambda\nu} \colon (\lambda,\nu) \in J_\delta := I_{\underline{\delta},0} \setminus I_{\delta,0}\}$. Then $X^\delta \oplus \operatorname{span}\Theta_\delta = X^{\underline{\delta}}$, and there exist constants $0 < m_\delta \le M_\delta$, only dependent on the gradedness constant L_δ, such that for any $z \in X^\delta$ and $\mathbf{c} = (c_{\lambda\nu})_{(\lambda,\nu)\in I_{\underline{\delta},0}\setminus I_{\delta,0}} \subset \mathbb{R}$,*

$$m_\delta(\|z\|_X^2 + \|\mathbf{c}\|^2) \le \|z + \mathbf{c}^\top\Theta_\delta\|_X^2 \le M_\delta(\|z\|_X^2 + \|\mathbf{c}\|^2)$$

So under the mild assumption that the gradedness constants of the sets X^δ that we encounter are uniformly bounded, we have shown that the condition (6.24) is satisfied.

Proof. Setting $c_{\lambda\nu} := 0$ when $(\lambda,\nu) \notin I_{\delta,0} \setminus I_{\delta,0}$, and writing $z = \sum_{\lambda\in\vee_\Sigma} \sigma_\lambda \otimes w_\lambda$ where $w_\lambda \in \operatorname{span}\{\hat\phi_\nu \colon (\lambda,\nu)\in I_{\delta,0}\}$, from Σ and $\{2^{-|\lambda|}\sigma_\lambda \colon \lambda\in\vee_\Sigma\}$ being Riesz bases for $L_2(I)$ and $H^1(I)$, and $\mathbf{c}^\top\Theta_\delta = \sum_\lambda \sigma_\lambda \otimes \sum_\nu e_{\lambda\mu}c_{\lambda\nu}\hat\phi_\nu$, an application of Lemma 6.6.7 given below shows that

$$\begin{aligned}
\|z+\mathbf{c}^\top\Theta_\delta\|_X^2 &\eqsim \sum_\lambda \big\{\|w_\lambda + \sum_\nu e_{\lambda\mu}c_{\lambda\nu}\hat\phi_\nu\|_{H^1(\Omega)}^2 + 4^{|\lambda|}\|w_\lambda+\sum_\nu e_{\lambda\mu}c_{\lambda\nu}\hat\phi_\nu\|_{H^1_{0,\Gamma_D}(\Omega)'}^2\big\}\\
&\eqsim \big\{\sum_\lambda \|w_\lambda\|_{H^1(\Omega)}^2 + 4^{|\lambda|}\|w_\lambda\|_{H^1_{0,\Gamma_D}(\Omega)'}^2 + \sum_\nu \big(2^{(\frac{2}{d}-1)\operatorname{gen}(\nu)} + 4^{|\lambda|}2^{(-\frac{2}{d}-1)\operatorname{gen}(\nu)}\big)|e_{\lambda\mu}c_{\lambda\nu}|^2\big\}\\
&= \sum_\lambda \|w_\lambda\|_{H^1(\Omega)}^2 + 4^{|\lambda|}\|w_\lambda\|_{H^1_{0,\Gamma_D}(\Omega)'}^2 + \sum_\nu |c_{\lambda\nu}|^2 \eqsim \|z\|_X^2 + \|\mathbf{c}\|^2,
\end{aligned}$$

with the $\eqsim$-symbol in the second line dependent on the gradedness constant. □

Lemma 6.6.7. *For $\tilde{\mathcal{T}} \in \mathbb{T}$, and either $\mathbb{T} \ni \mathcal{T} \preceq \tilde{\mathcal{T}}$ and $v \in W_{\mathcal{T}}$, or $\mathcal{T} = \emptyset$, $N_{\mathcal{T},0} := \emptyset$, and $v = 0$, and scalars $(d_\nu)_{\nu\in N_{\tilde{\mathcal{T}},0}\setminus N_{\mathcal{T},0}}$, it holds that*

$$\|v + \sum_\nu d_\nu\hat\phi_\nu\|_{H^1(\Omega)}^2 \eqsim \|v\|_{H^1(\Omega)}^2 + \sum_\nu 2^{(\frac{2}{d}-1)\operatorname{gen}(\nu)}|d_\nu|^2 \tag{6.63}$$

$$\|v + \sum_\nu d_\nu\hat\phi_\nu\|_{H^1_{0,\Gamma_D}(\Omega)'}^2 \eqsim \|v\|_{H^{-1}(\Omega)}^2 + \sum_\nu 2^{(-\frac{2}{d}-1)\operatorname{gen}(\nu)}|d_\nu|^2 \tag{6.64}$$

with the constants hidden in the $\eqsim$-symbols only dependent on $M_{\tilde{\mathcal{T}}\mathcal{T}} := \max\{\operatorname{gen}(\tilde T) - \operatorname{gen}(T)\colon \tilde{\mathcal{T}} \ni \tilde T \subset T \in \mathcal{T}\}$ or $M_{\tilde{\mathcal{T}}\mathcal{T}} := \max\{\operatorname{gen}(\tilde T)\colon \tilde T \in \tilde{\mathcal{T}}\}$ for $\mathcal{T} = \emptyset$.

Proof. Once the equivalences are shown uniformly in any $\mathcal{T} \preceq \tilde{\mathcal{T}}$ for which $M_{\tilde{\mathcal{T}}\mathcal{T}} = 1$, a repeated application of these equivalences shows them for the general case, with constants that are only dependent on $M_{\tilde{\mathcal{T}}\mathcal{T}}$. So in the following, it suffices to consider the case that $M_{\tilde{\mathcal{T}}\mathcal{T}} = 1$. The case $\mathcal{T} = \emptyset$ is easy, so we will consider the case that $\mathcal{T} \in \mathbb{T}$.

Let $\Phi_{\tilde{\mathcal{T}}} = \{\phi_{\tilde{\mathcal{T}},\nu} \colon \nu \in N_{\tilde{\mathcal{T}},0}\}$ denote the standard nodal basis for $W_{\tilde{\mathcal{T}}}$. For any weight function $0 < w_{\tilde{\mathcal{T}}} \in \prod_{T\in\tilde{\mathcal{T}}} P_0(T)$, with $\|\cdot\|_{L_{2,w_{\tilde{\mathcal{T}}}}(\Omega)} := \|w_{\tilde{\mathcal{T}}}^{\frac12}\cdot\|_{L_2(\Omega)}$ it holds that $\|\sum_\nu c_\nu\phi_{\tilde{\mathcal{T}},\nu}\|_{L_{2,w_{\tilde{\mathcal{T}}}}(\Omega)}^2 \eqsim \sum_\nu |c_\nu|^2\|\phi_{\tilde{\mathcal{T}},\nu}\|_{L_{2,w_{\tilde{\mathcal{T}}}}(\Omega)}^2$, only dependent on the spectrum of the element mass matrix on a reference element, i.e., on the space dimension d, so independent of the weight function $w_{\tilde{\mathcal{T}}}$. We refer to this equivalence by saying that $\Phi_{\tilde{\mathcal{T}}}$ is (uniformly) stable w.r.t. $\|\cdot\|_{L_{2,w_{\tilde{\mathcal{T}}}}(\Omega)}$.

Notice that for $\nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}$, it holds that $\phi_{\tilde{\mathcal{T}},\nu} = \phi_\nu$. W.r.t. the splitting $N_{\tilde{\mathcal{T}},0} = N_{\mathcal{T},0} + N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}$, the basis transformation from $\Phi_{\mathcal{T}} \cup \{\hat\phi_\nu \colon \nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}\}$ to $\Phi_{\mathcal{T}} \cup \{\phi_\nu \colon \nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}\}$ is of the form $\begin{bmatrix}\mathrm{Id} & *\\ 0 & \mathrm{Id}\end{bmatrix}$, and the

basis transformation from the latter basis to $\Phi_{\tilde{\mathcal{T}}}$ is of the form $\begin{bmatrix} \mathrm{Id} & 0 \\ * & \mathrm{Id} \end{bmatrix}$. The entries in both non-zero off-diagonal blocks are uniformly bounded, where non-zeros can only occur for index pairs $(\nu, \tilde{\nu})$ that are vertices of the same $\tilde{T} \in \tilde{\mathcal{T}}$. Consequently, for a family of weight functions $(w_{\tilde{\mathcal{T}}})_{\tilde{\mathcal{T}} \in \mathbb{T}}$ that have *uniformly bounded jumps* in the sense that

$$\sup_{\tilde{\mathcal{T}} \in \mathbb{T}} \sup_{\{T, T' \in \tilde{\mathcal{T}} \colon T \cap T' \neq \emptyset\}} \frac{w_{\tilde{\mathcal{T}}}|_T}{w_{\tilde{\mathcal{T}}}|_{T'}} < \infty, \tag{6.65}$$

all basis transformations between the $L_{2,w_{\tilde{\mathcal{T}}}}(\Omega)$-normalized bases $\Phi_{\mathcal{T}} \cup \{\hat{\phi}_\nu \colon \nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}\}$, $\Phi_{\mathcal{T}} \cup \{\phi_\nu \colon \nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}\}$ and $\Phi_{\tilde{\mathcal{T}}}$ are uniformly bounded.

Since, as we have seen, $\Phi_{\tilde{\mathcal{T}}}$ is (uniformly) stable w.r.t. $\|\cdot\|_{L_{2,w_{\tilde{\mathcal{T}}}}(\Omega)}$, we conclude that also $\Phi_{\mathcal{T}} \cup \{\hat{\phi}_\nu \colon \nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}\}$ and $\Phi_{\mathcal{T}} \cup \{\phi_\nu \colon \nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}\}$ are (uniformly) stable w.r.t. $\|\cdot\|_{L_{2,w_{\tilde{\mathcal{T}}}}(\Omega)}$. Because of the *uniform K-mesh property* of $\mathcal{T} \in \mathbb{T}$, examples of families of weights that satisfy (6.65) are given by $(h_{\mathcal{T}}^s)_{\tilde{\mathcal{T}} \in \mathbb{T}}$ for any $s \in \mathbb{R}$, where $h_{\tilde{\mathcal{T}}}|_T := 2^{-\operatorname{gen}(T)/d} (\eqsim |T|^{1/d})$ $(T \in \tilde{\mathcal{T}})$.

For showing (6.63), let $P_{\mathcal{T}} \colon W_{\tilde{\mathcal{T}}} \to W_{\mathcal{T}}$ be the projector with $\operatorname{ran} P_{\mathcal{T}} = W_{\mathcal{T}}$ and $\operatorname{ran}(\mathrm{Id} - P_{\mathcal{T}}) = \operatorname{span}\{\hat{\phi}_\nu \colon \nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}\}$. Using the form of the basis transformation from $\Phi_{\mathcal{T}} \cup \{\phi_\nu \colon \nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}\}$ to $\Phi_{\mathcal{T}} \cup \{\hat{\phi}_\nu \colon \nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}\}$, one infers that

$$\mathrm{Id} - P_{\mathcal{T}} = J_{\mathcal{T}} \circ (\mathrm{Id} - I_{\mathcal{T}}),$$

where $I_{\mathcal{T}}$ is the nodal interpolator onto $W_{\mathcal{T}}$, and $J_{\mathcal{T}}$ is defined by $J_{\mathcal{T}} \phi_\nu = \hat{\phi}_\nu$. Since both $\{\hat{\phi}_\nu \colon \nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}\}$ and $\{\phi_\nu \colon \nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}\}$ are uniformly stable w.r.t. $\|h_{\tilde{\mathcal{T}}}^{-1} \cdot\|_{L_2(\Omega)}$, and $\|h_{\tilde{\mathcal{T}}}^{-1} \hat{\phi}_\nu\|_{L_2(\Omega)} \eqsim \|h_{\tilde{\mathcal{T}}}^{-1} \phi_\nu\|_{L_2(\Omega)}$, it follows that $J_{\mathcal{T}}$ is uniformly bounded w.r.t. $\|h_{\tilde{\mathcal{T}}}^{-1} \cdot\|_{L_2(\Omega)}$, i.e. $\|h_{\tilde{\mathcal{T}}}^{-1} J_{\mathcal{T}} h_{\tilde{\mathcal{T}}}\|_{\mathcal{L}(L_2(\Omega), L_2(\Omega))} \lesssim 1$, and so

$$\|h_{\tilde{\mathcal{T}}}^{-1}(\mathrm{Id} - P_{\mathcal{T}})v\|_{L_2(\Omega)} \lesssim \|h_{\tilde{\mathcal{T}}}^{-1}(\mathrm{Id} - I_{\mathcal{T}})v\|_{L_2(\Omega)} \lesssim |v|_{H^1(\Omega)} \quad (v \in W_{\tilde{\mathcal{T}}}).$$

Using the common inverse inequality $\|\cdot\|_{H^1(\Omega)} \lesssim \|h_{\tilde{\mathcal{T}}}^{-1} \cdot\|_{L_2(\Omega)}$ on $W_{\tilde{\mathcal{T}}}$, we infer that $(\mathrm{Id} - P_{\mathcal{T}})$ is uniformly bounded in the $H^1(\Omega)$-norm, and that $\|\cdot\|_{H^1(\Omega)} \eqsim \|h_{\tilde{\mathcal{T}}}^{-1} \cdot\|_{L_2(\Omega)}$ on $\operatorname{ran}(\mathrm{Id} - P_{\mathcal{T}})$. The proof of (6.63) is completed by the uniform stability of $\{\hat{\phi}_\nu \colon \nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}\}$ w.r.t. $\|h_{\tilde{\mathcal{T}}}^{-1} \cdot\|_{L_2(\Omega)}$, and the fact that $\|h_{\tilde{\mathcal{T}}}^{-1} \hat{\phi}_\nu\|_{L_2(\Omega)}^2 \eqsim 2^{(\frac{2}{d}-1)\operatorname{gen}(\nu)}$.

Moving to (6.64), either by $\int_\Omega \hat{\phi}_\nu \, dx = 0$, or otherwise using the proximity of the Dirichlet boundary Γ_D by an application of Poincaré's inequality, it holds that

$$|\langle \hat{\phi}_\nu, v \rangle_{L_2(\Omega)}| \lesssim 2^{-\operatorname{gen}(\nu)/d} \|\hat{\phi}_\nu\|_{L_2(\Omega)} |v|_{H^1(\operatorname{supp} \hat{\phi}_\nu)} \quad (\nu \in \mathfrak{N}_0 \setminus N_{\mathcal{T}_\perp,0}).$$

By using that for $T \in \tilde{\mathcal{T}}$ the number of $\nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}$ for which $\operatorname{supp} \hat{\phi}_\nu$ has non-empty intersection with T is uniformly bounded, and furthermore that $\Phi_{\mathcal{T}} \cup \{\hat{\phi}_\nu \colon \nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}\}$ is uniformly stable w.r.t. $\|h_{\mathcal{T}} \cdot\|_{L_2(\Omega)}$, we infer that for any $z = \sum_{\nu \in N_{\mathcal{T},0}} z_\nu \phi_{\mathcal{T},\nu} \in W_{\mathcal{T}}$ it holds that

$$
\begin{aligned}
&\| \sum_{\nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}} d_\nu \hat{\phi}_\nu \|_{H^1_{0,\Gamma_D}(\Omega)'} = \sup_{0 \neq v \in H^1_{0,\Gamma_D}(\Omega)} \frac{\langle \sum_{\nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}} d_\nu \hat{\phi}_\nu, v \rangle_{L_2(\Omega)}}{\|v\|_{H^1(\Omega)}} \\
&\lesssim \sqrt{\sum_{\nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}} |d_\nu|^2 \|h_{\tilde{\mathcal{T}}} \hat{\phi}_\nu\|^2_{L_2(\Omega)}} \\
&\leq \sqrt{\sum_{\nu \in N_{\mathcal{T},0}} |z_\nu|^2 \|h_{\tilde{\mathcal{T}}} \phi_{\mathcal{T},\nu}\|^2_{L_2(\Omega)} + \sum_{\nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}} |d_\nu|^2 \|h_{\tilde{\mathcal{T}}} \hat{\phi}_\nu\|^2_{L_2(\Omega)}} \\
&\eqsim \|h_{\tilde{\mathcal{T}}}(z + \sum_{\nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}} d_\nu \hat{\phi}_\nu)\|_{L_2(\Omega)} \lesssim \|z + \sum_{\nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}} d_\nu \hat{\phi}_\nu\|_{H^1_{0,\Gamma_D}(\Omega)'}
\end{aligned}
\tag{6.66}
$$

the last inequality by application of a less common inverse inequality which proof can be found in Lemma 2.3.4 for general dimensions d. From (6.66) it follows that $\mathrm{Id} - P_{\mathcal{T}}$ is uniformly bounded in the $H^1_{0,\Gamma_D}(\Omega)'$-norm, and also that $\| \sum_{\nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}} d_\nu \hat{\phi}_\nu \|^2_{H^1_{0,\Gamma_D}(\Omega)'} \eqsim \sum_{\nu \in N_{\tilde{\mathcal{T}},0} \setminus N_{\mathcal{T},0}} |d_\nu|^2 2^{(-\frac{2}{d}-1)\operatorname{gen}(\nu)}$, where we used that $\|h_{\tilde{\mathcal{T}}} \hat{\phi}_\nu\|^2_{L_2(\Omega)} \eqsim 2^{(-\frac{2}{d}-1)\operatorname{gen}(\nu)}$. The proof of (6.64) is completed. □

6.6.3 The wavelet collections Σ and Ψ

As wavelet basis $\Sigma = \{\sigma_\lambda \colon \lambda \in \vee_\Sigma\}$ we select the three-point hierarchical basis illustrated in Figure 6.2. This basis is known to be a Riesz basis for $L_2(I)$, and, after re-normalization, for $H^1(I)$ (see [Ste96]). It also satisfies the other assumptions made in Sect. 6.5.2. The wavelets up to level ℓ span the space of continuous piecewise linear functions on I w.r.t. the uniform partition into 2^ℓ subintervals.

As wavelet basis $\Psi = \{\psi_\mu \colon \mu \in \vee_\Psi\}$ we take the orthonormal (discontinuous) piecewise linear wavelets, see Figure 6.3. The wavelets up to level ℓ span the space of (discontinuous) piecewise linear functions on I w.r.t. the uniform partition into 2^ℓ subintervals.

6.6.4 The family $({}^\delta G, {}^\delta U_0)_{\delta \in \Delta}$

The index set $\vee_\Sigma$ is naturally identified with the set of 'nodal dyadic' points, see Figure 6.4, which is the natural index set for the one-dimensional hierarchical basis that we denote by $\{\phi_\lambda \colon \lambda \in \vee_\Sigma\}$. Recalling that for $\delta \in \Delta$, $X^\delta = \operatorname{span}\{\sigma_\lambda \otimes \phi_\nu \colon (\lambda,\nu) \in I_{\delta,0} = I_\delta \setminus (\vee_\Sigma \times \Gamma_D)\}$ for some lower set $I_\delta \subset \vee_\Sigma \times \mathfrak{N}$,

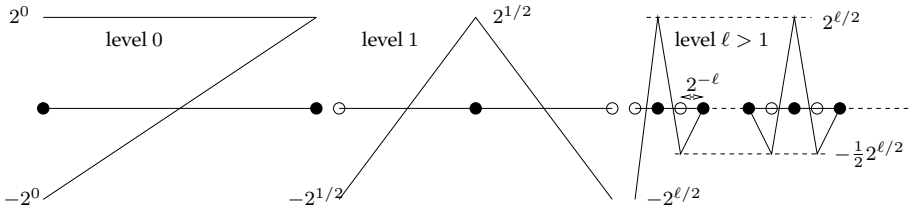


FIGURE 6.2. Three-point hierarchical basis Σ. On level 0 there are two wavelets, and on level 1 there is one wavelet, whose parents are both wavelets on level 0. On each level $\ell > 1$ there are $2^{\ell-1}$ wavelets, among them near each boundary one boundary-adapted wavelet, where each wavelet has one parent being the wavelet on level $\ell-1$ whose support includes the support of its child (so $\mathcal{S}_\Sigma(\lambda)$ can be taken equal to $\operatorname{supp}\sigma_\lambda$). All but one wavelets have one (bdr. wav) or two vanishing moments.

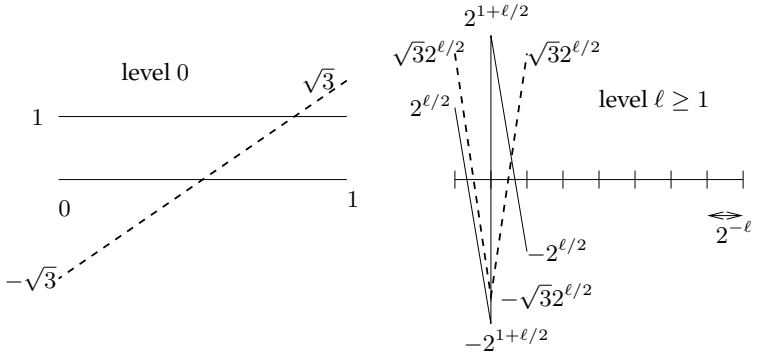


FIGURE 6.3. $L_2(I)$-orthonormal (discontinuous) piecewise linear wavelet basis Ψ. On level $\ell = 0$ there are 2 wavelets. On each level $\ell \geq 1$ there are 2^ℓ wavelets of two types, each of them having 2 parents being the wavelets on level $\ell - 1$ whose supports include the supports of their children (so $\mathcal{S}_\Psi(\mu)$ can be taken equal to $\operatorname{supp}\psi_\mu$). The wavelets on level 0 have either 0 or 1 vanishing moment, all other wavelets have two vanishing moments.

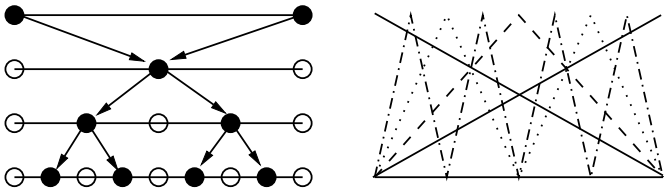


FIGURE 6.4. Index set $\vee_\Sigma$ with parent-child relations, and the one-dimensional hierarchical basis.

we define

$$ {}^\delta G := \operatorname{span}\{\phi_\lambda \otimes \phi_\nu \colon (\lambda,\nu) \in I_\delta\},\ {}^\delta U_0 := \operatorname{span}\{\phi_\nu \colon (\lambda,\nu) \in I_\delta,\ \phi_\lambda(0) \neq 0\}. $$

Since the level of resolution of these spaces is comparable to that of X^δ, based on our experiences with wavelet and finite element methods we *expect* that with this choice of $({}^\delta G, {}^\delta U_0)$ and the definition of $X^{\underline{\delta}\delta}$, that *saturation* holds, i.e., that Assumption 6.3.1 assumption is valid.

Given $g \in Y'$ and $u_0 \in L_2(\Omega)$, it remains to define their approximations $({}^\delta g, {}^\delta u_0) \in ({}^\delta G, {}^\delta U_0)$. In general, the construction of these approximations depends on the data at hand. Below we give a construction that applies to general continuous g and u_0, and that avoids quadrature issues.

For $\nu \in \mathfrak{N}$ with $\operatorname{gen}(\nu) = 0$, let $\tilde\phi_\nu := \delta_\nu$. Each $\nu \in \mathfrak{N}$ with $\operatorname{gen}(\nu) > 0$ is the midpoint of an edge of a $T \in \mathcal{T}$ with $\operatorname{gen}(T) = \operatorname{gen}(\nu) - 1$. Denoting the endpoints of this edge as $\nu_1, \nu_2 \in \mathfrak{N}$, let $\tilde\phi_\nu := \delta_\nu - \frac{1}{2}(\delta_{\nu_1} + \delta_{\nu_2})$. Then $\{\tilde\phi_\nu \colon \nu \in \mathfrak{N}\} \subset C(\overline{\Omega})'$ is biorthogonal to $\{\phi_\nu \colon \nu \in \mathfrak{N}\}$. With $\{\tilde\phi_\lambda \colon \lambda \in \vee_\Sigma\} \subset C(\overline{I})'$ defined analogously for the one-dimensional case, for $g \in C(\overline{I \times \Omega})$ and $u_0 \in C(\overline{\Omega})$ we define the interpolants

$$ {}^\delta g := \sum_{(\lambda,\nu)\in I_\delta} (\tilde\phi_\lambda \otimes \tilde\phi_\nu)(g)\phi_\lambda \otimes \phi_\nu,\ {}^\delta u_0 := \sum_{\{\nu \colon (\lambda,\nu)\in I_\delta,\, \phi_\lambda(0)\neq 0\}} \tilde\phi_\nu(u_0)\phi_\nu. $$

Since we expect that for sufficiently smooth g and u_0, the errors $\|g - {}^\delta g\|_{Y'}$ and $\|u_0 - {}^\delta u_0\|_{L_2(\Omega)}$ are of higher order than the approximation error $\inf_{w\in X^\delta} \|u - w\|_X$, for our convenience in the adaptive Algorithm 6.4.8 we ignore errors caused by data-oscillation by setting $\eta(\cdot) \equiv 0$.

Notice that setting up the matrix vector formulation of the system (6.31) that defines our approximation u^δ requires computing the vectors

$$ \big[\langle {}^\delta g, \psi_\mu \otimes \phi_\nu\rangle_{L_2(I\otimes\Omega)}\big]_{(\mu,\nu)\in I^Y_{\underline{\delta},0}},\quad \big[\langle {}^\delta u_0, \phi_\nu\rangle_{L_2(\Omega)}\big]_{\{\nu \colon (\lambda,\nu)\in I_{\delta,0},\, \sigma_\lambda(0)\neq 0\}} $$

which can be performed in $\mathcal{O}(\dim X^\delta)$ operations because I_δ and $I^Y_{\underline{\delta},0}$ are lower sets (and $\#I^Y_{\underline{\delta},0} \lesssim \#I_\delta$).

6.7 Numerical experiments

We test our algorithm on the heat equation, i.e., the parabolic problem with $a(t;\eta,\zeta) = \int_\Omega \nabla\eta\cdot\nabla\zeta d\mathbf{x}$, posed on a two-dimensional polygonal spatial domain Ω, and Dirichlet boundary $\Gamma_D = \partial\Omega$. Recall from §6.6.3 the three-point continuous piecewise linear temporal wavelet basis Σ, the orthonormal discontinuous piecewise linear temporal wavelet basis Ψ, and the hierarchical continuous piecewise linear spatial basis $\Xi := \{\phi_\nu : \nu \in \mathfrak{N}_0\}$.

We consider 'trial' spaces X^δ which are spanned by finite subsets of $\Sigma\otimes\Xi$ whose index sets are lower sets (more precisely, satisfy (6.59)-(6.60)), and corresponding 'test' spaces Y^δ spanned by finite subsets of $\Psi\otimes\Xi$ as defined in (6.61)-(6.62). We construct the enlarged trial space $X^{\underline{\delta}}$ as defined in Def. 6.6.3, with corresponding test space $Y^{\underline{\delta}}$.

For a given level $N\in\mathbb{N}$, $\operatorname{span}\{\sigma_\lambda : |\lambda|\le N\}$ coincides with the span of the continuous piecewise linears on an N-times recursive dyadic refinement of I, and $\operatorname{span}\{\phi_\nu\in\Xi : \operatorname{gen}(\nu)\le 2N\}$ coincides with that of the continuous piecewise linears, zero at $\partial\Omega$, on a $2N$-times recursive bisection refinement of an initial partition $\mathcal{T}_\perp$. Therefore, the span of the *'full'* tensor product $\{\sigma_\lambda\colon |\lambda|\le N\}\otimes\{\phi_\nu\colon \operatorname{gen}(\nu)\le 2N\}$ equals a space of lowest order continuous finite elements w.r.t. a quasi-uniform shape regular product mesh into prismatic elements.

Taking only those index pairs (λ,ν) for which $2|\lambda|+\operatorname{gen}(\nu)\le 2N$ produces a *'sparse'* tensor product on level N. Sparse tensor products allow to overcome the *curse of dimensionality* in the sense that for smooth solutions they achieve a rate in X-norm that is equal to the best rate in the $H^1(\Omega)$-norm that can be expected for the corresponding stationary problem on the spatial domain, here the Poisson equation; see also Sect. 6.5.5.

We run our adaptive Algorithm 6.4.8 with $\theta = 0.5$ and $\xi = \frac{1}{2}$, computing ${}^\delta g$ and ${}^\delta u_0$ as in Sect. 6.6.4. Since we envisage that in our experiments data-oscillation errors are not dominant, for our convenience we took $\omega = \infty$. We solve the arising linear system of (6.31) using Preconditioned CG, using the previous solution as initial guess. We then perform Dörfler marking on the residual, yielding a minimal set J, and finally choose $I_{\tilde\delta}$ as the smallest lower set containing $J\cup I_\delta$. Due to this constraint generally we add index pairs outside of the marked set, i.e. $I_{\tilde\delta}\setminus I_\delta \supsetneq J$. Still, in our experiments, we *observe* $\#I_{\tilde\delta} - \#I_\delta \lesssim \#J$ with a moderate constant.

Remark 6.7.1. Rather we would have applied an algorithm that produces a $I_{\tilde\delta}$ such that $I_{\tilde\delta}\setminus I_\delta$ is *guaranteed* to have an, up to a multiplicative factor, smallest cardinality among all lower sets $I_{\tilde\delta}\supset I_\delta$ that realize the bulk criterion. Such an algorithm was introduced in [BD04, BFV19] for 'single-tree' approximation, but seems not to be available for the 'double-tree' (i.e. lower set) constraint that we need here.

We compare adaptive refinement with non-adaptive full- and sparse tensor

products, and monitor the error estimator $\mathcal{E}^\delta(\tilde{u}^\delta)$ from Proposition 6.4.5, the residual error estimator from Proposition 6.4.3, and the $L_2(\Omega)$ trace error at $t = 0$.

6.7.1 Condition numbers of preconditioner

For the calibration of our preconditioners, we consider $\Omega := [0,1]^2$, and compare *uniformly refined* space-time meshes with *locally refined* meshes with refinements towards $\{0\} \times \partial\Omega$.

The replacement of the nonlocal operator $(E_Y^{\delta\,\prime} A E_Y^\delta)^{-1}$ in the forward application of $S^{\delta\delta}$ by the block-diagonal preconditioner K_Y^δ from Sect. 6.5.6 is only guaranteed to result in a convergent algorithm when the eigenvalues of $K_Y^\delta E_Y^{\delta\,\prime} A E_Y^\delta$ are sufficiently close to one.

In Table 6.1, we investigate the values

$$\kappa_\delta := \max\{\lambda_{\max}(\mathbf{K}_Y^\delta \mathbf{A}_Y^\delta), 1/\lambda_{\min}(\mathbf{K}_Y^\delta \mathbf{A}_Y^\delta)\}$$

with $\boldsymbol{A}_Y^\delta$ the matrix representation of $E_Y^{\delta\,\prime} A E_Y^\delta$, and $\mathbf{K}_Y^\delta$ built from spatial multigrid preconditioners $\mathbf{K}_\mu^\delta$ corresponding to n V-cycles. In each V-cycle we applied one pre- and one post Gauss-Seidel smoother. In case of a locally refined spatial mesh, on each level these Gauss-Seidel updates were restricted to the vertices whose generation is equal to that level as well as both endpoints of the edge on which these vertices were inserted ([WZ17]). We see that for both uniform and locally refined space-time meshes, κ_δ converges to 1 rapidly in n, and is essentially independent of $\dim X^\delta$. In our examples, κ_δ is sufficiently close to one already for $n = 1$.

Fixing $n = 1$ for the forward application of $S^{\delta\delta}$, we want to precondition $S^{\delta\delta}$ itself as well. Following Sect. 6.5.6, we build a block-diagonal preconditioner taking $\mathbf{K}_\lambda^\delta$ to correspond to m V-cycles of the aforementioned multigrid method now applied to $\mathbf{A}_\lambda^\delta + 2^{|\lambda|}\mathbf{M}_\lambda^\delta$ with $\mathbf{A}_\lambda^\delta$ and $\mathbf{M}_\lambda^\delta$ being stiffness- or mass-matrices.

	$\dim X^\delta$	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$
uniform	729	1.343	1.070	1.017	1.004	1.001	1.000
	35937	1.360	1.075	1.019	1.004	1.001	1.000
	2146689	1.365	1.077	1.019	1.004	1.001	1.000
local	766	1.306	1.058	1.013	1.003	1.001	1.000
	30151	1.307	1.058	1.013	1.003	1.001	1.000
	1964797	1.307	1.058	1.013	1.003	1.001	1.000

TABLE 6.1. Values $\kappa_\delta := \max\{\lambda_{\max}(\mathbf{K}_Y^\delta \mathbf{A}_Y^\delta), 1/\lambda_{\min}(\mathbf{K}_Y^\delta \mathbf{A}_Y^\delta)\}$ using spatial multigrid with n V-cycles.

Table 6.2 shows the condition numbers of the preconditioned matrix. We again see fast stabilization in m as well as in $\dim X^\delta$. We fix $m = 3$ in the sequel. Most interestingly, in every of our example problems, the adaptive algorithm only needs one or two PCG iterations to reach the error tolerance t_δ.

6.7.2 Smooth problem

We consider the square domain $\Omega := [0,1]^2$ and prescribe

$$u(t,x,y) := (1+t^2)x(1-x)y(1-y)$$

with derived data u_0 and g. For this smooth solution, full and sparse tensor products are expected to yield the best possible error decays proportional to $(\dim X^\delta)^{-1/3}$ and $(\dim X^\delta)^{-1/2}$, respectively.

The left side of Figure 6.5 shows the error progressions for the smooth problem. We plot the error estimator $\mathcal{E}^\delta(\tilde{u}^\delta) := \mathcal{E}^\delta(\tilde{u}^\delta; {}^\delta g, {}^\delta u_0) \eqsim \|{}^\delta u - \tilde{u}^\delta\|_X$ from Proposition 6.4.5, the residual error estimator $\|\mathbf{r}^\delta\|$, and $\|\gamma_0({}^\delta u - \tilde{u}^\delta)\|_{L_2(\Omega)}$. We see that the error progressions are as expected. For this solution, adaptive refinement yields no advantage over sparse grid refinement. We observe a higher order of convergence for the trace at $t = 0$ measured in $L_2(\Omega)$.

6.7.3 Moving peak problem

We consider a square domain $\Omega := [0,1]^2$ and select

$$u(t,x,y) := x(1-x)y(1-y)\exp(-100[(x-t)^2 + (y-t)^2]).$$

We took this example from [LS20]. The solution is smooth, and almost zero everywhere except on a small strip near the diagonal from $(0,0,0)$ to $(1,1,1)$ of the space-time cylinder. As u is smooth, we expect sparse grid refinements

	$\dim X^\delta$	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$
uniform	4913	9.196	6.119	6.048	6.042	6.041	6.041
	35937	9.718	6.315	6.263	6.260	6.260	6.260
	274625	9.991	6.750	6.749	6.751	6.752	6.752
	2146689	10.115	7.080	7.087	7.088	7.088	7.088
local	3520	5.707	5.132	5.110	5.111	5.111	5.111
	30151	6.355	5.734	5.706	5.704	5.704	5.704
	244870	7.619	6.879	6.843	6.841	6.841	6.841
	1964797	9.353	8.734	8.703	8.701	8.701	8.701

Table 6.2. Spectral condition numbers of $K_X^\delta S^{\delta\delta}$, using spatial multigrid with m V-cycles.

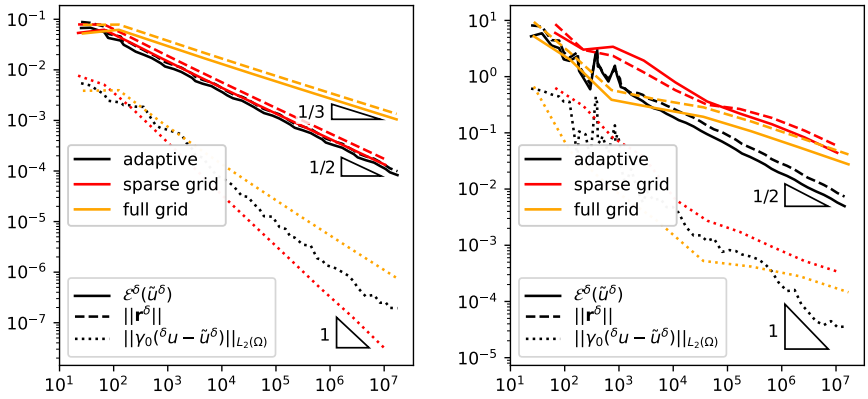


FIGURE 6.5. Error progressions for (left) the *smooth problem* and (right) the *moving peak* problem. Shown: estimated X-norm error (solid line), residual norm (dashed), and $t = 0$ trace error (dotted) as a function of $\dim X^\delta$ for adaptive (black), sparse grid (red), and full grid refinement (orange).

to asymptotically yield the optimal error decay proportional to $(\dim X^\delta)^{-1/2}$, albeit with a terrible constant. Adaptive refinement should be able to achieve the same rate at quantitatively smaller doubletrees.

From the right of Figure 6.5, we see that the sparse grid rate is not (yet) optimal, while our adaptive routine is able to find the optimal rate from $\dim X^\delta \approx 10^3$ onwards. Figure 6.6 shows the number of basis functions $\sigma_\lambda \otimes \phi_\nu$ whose supports intersect given points in the time-space cylinder. We see the adaptation to the moving peak.

6.7.4 Cylinder problem

Selecting the L-shaped domain $\Omega := [-1,1]^2 \setminus [-1,0]^2$ with data $u_0 \equiv 0$ and $g(t,x,y) := t \cdot \mathbb{1}_{\{x^2+y^2<1/4\}}$, the true solution is known to be singular at the re-entrant corner and at the wall of the cylinder $\{(t,x,y)\colon x^2 + y^2 = 1/4\}$. We took this example from [FK21]. The left side of Figure 6.7 shows the error progression for this cylinder problem. We see that the full grid error decay proportional to $(\dim X^\delta)^{-1/4}$ is improved to an error decay proportional to $(\dim X^\delta)^{-1/3}$ by considering sparse grids. Adaptive refinement, however, achieves the best possible error decay proportional to $(\dim X^\delta)^{-1/2}$, recovering the rate for a smooth solution.

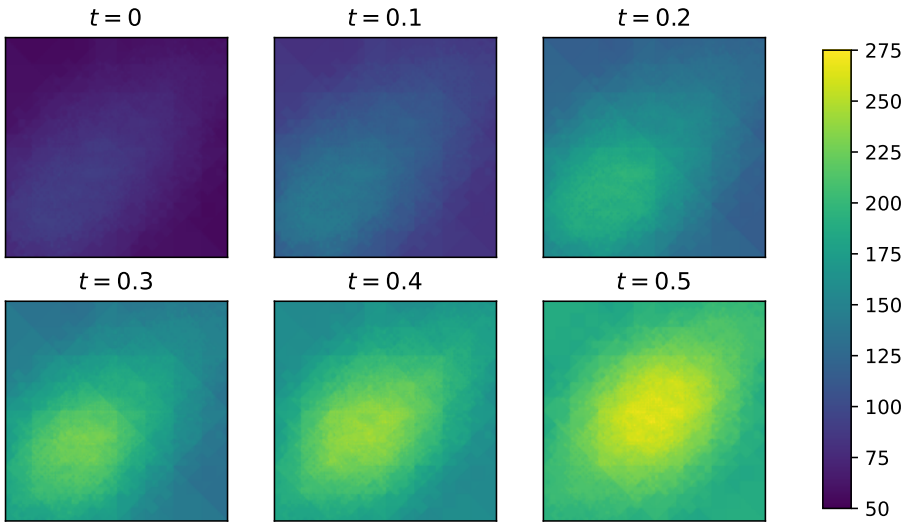


Figure 6.6. *Moving peak* problem, adaptive lower set with $\dim X^\delta = 89\,401$. Shown: $\#\{(\lambda,\nu) \in I_\delta \colon (t,x,y) \in \operatorname{supp} \sigma_\lambda \otimes \phi_\nu\}$ for a selection of times t.

6.7.5 Singular problem

We again select the L-shaped domain $\Omega := [-1,1]^2 \setminus [-1,0]^2$ with data $u_0 \equiv \mathbb{1}$ and $g \equiv 0$. The solution has a strong singularity along $\{0\} \times \partial\Omega$ due to the incompatibility of initial- and boundary conditions, in addition to the singularity at the re-entrant corner $(0,0)$. At the right of Figure 6.7, for uniform refinement, we see the extremely slow error decay proportional to $(\dim X^\delta)^{-1/11}$, already found in [FK21]. Interestingly, sparse grid refinement offers no rate improvement over full grid refinement. The adaptive algorithm yields a much better error decay proportional to $(\dim X^\delta)^{-2/5}$. We observed that increasing the Dörfler marking parameter to $\theta = 0.7$ decreases the convergence rate to $-1/3$, whereas a θ smaller than 0.5 did not improve the rate beyond $-2/5$. Looking at Figure 6.8, we see strong adaptivity towards $\{0\} \times \partial\Omega$ and $I \times \{(0,0)\}$, and observe basis functions $\sigma_\lambda \otimes \phi_\nu$ that span X^δ whose barycenter is at $t = 2^{-14} \approx 10^{-4}$.

6.7.6 Gradedness and error reduction

In Sect. 6.4 we used (6.24) to demonstrate proportionality of $\|\mathbf{r}^\delta\|$ and $\|u-u^\delta\|_X$, as well as a constant error reduction in each iteration of the adaptive algorithm. In Proposition 6.6.6, we showed that (6.24) holds when the gradedness L_δ of Definition 6.6.4 is uniformly bounded.

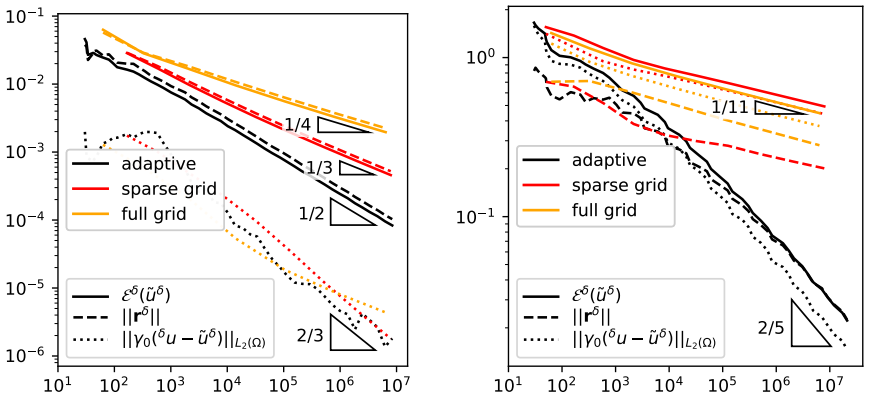


FIGURE 6.7. Error progressions for (left) the *cylinder* problem and (right) the *singular* problem. Shown: estimated X-norm error (solid line), residual norm (dashed), and $t = 0$ trace error (dotted) as a function of $\dim X^\delta$ for adaptive (black), sparse grid (red), and full grid refinement (orange).

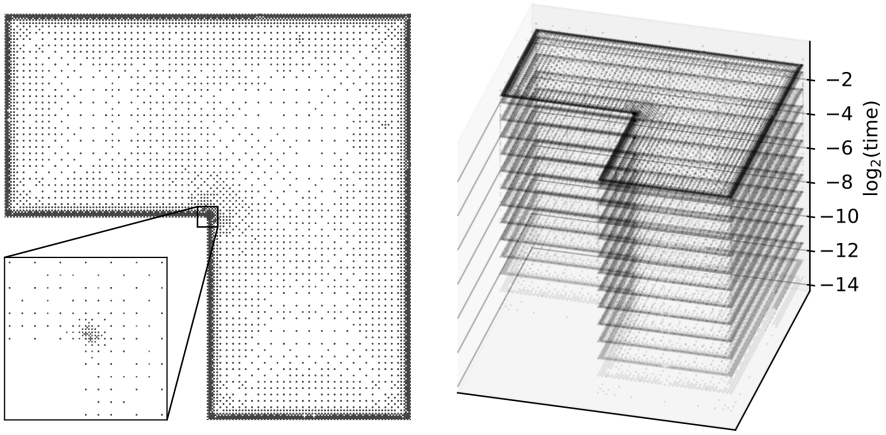


FIGURE 6.8. Barycenters of supports of basis functions $\sigma_\lambda \otimes \phi_\nu$ spanning X^δ generated by Algorithm 6.4.8 of dimension 81 074 for the *singular* problem. Left: a top-down view, with a $10\times$ zoom to the origin; right: centers in spacetime, logarithmic in time.

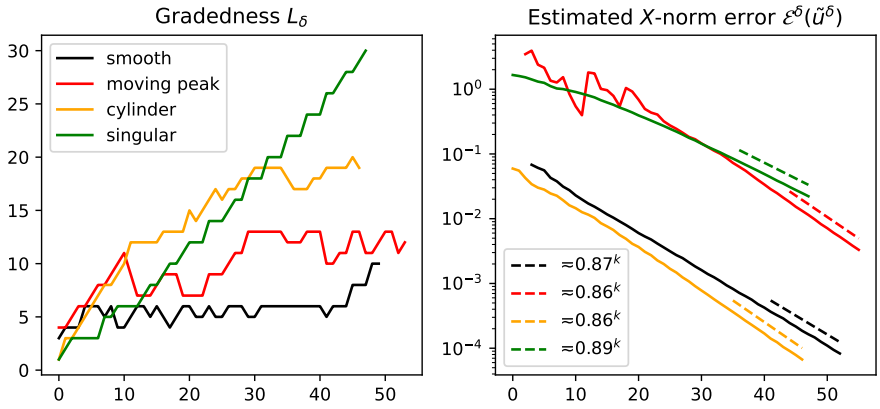


Figure 6.9. Gradedness and estimated X-norm error at every iteration of the adaptive loop, for the four different model problems under consideration.

In the left picture of Figure 6.9, we see however a more than expected increase in gradedness, where in particular for the singular problem we observe a logarithmic increase in terms of $\dim X^\delta$. However, this turns out not to be a problem in practice: Figures 6.5 and 6.7 demonstrate that the residual error $\|\mathbf{r}^\delta\|$ and the estimated X-norm error $\mathcal{E}^\delta(\tilde{u}^\delta)$ are very close, and even converge for the singular problem. Moreover, in the right picture of Figure 6.9, we see a constant error reduction of $\breve{\rho} \approx 0.89$ at every step of the adaptive algorithm, and hence, that the conclusion of Theorem 6.4.9 holds in practice.

6.7.7 Total runtime and memory consumption

Figure 6.10 shows the total runtime and peak memory consumption after every iteration of the adaptive algorithm. The top row shows absolute values, and the bottom row values relative to $\dim X^\delta$.

The left of the figure shows that the adaptive algorithm runs in optimal linear time in the dimension of the current trial space.

The right of the figure shows that the peak memory is linear as well, stabilizing to around 15kB per degree of freedom. This is relatively high, mainly because our implementation uses trees rather than hash maps to represent vectors to ensure a linear-time implementation of the matrix-vector products (cf. Rem. 6.6.2).

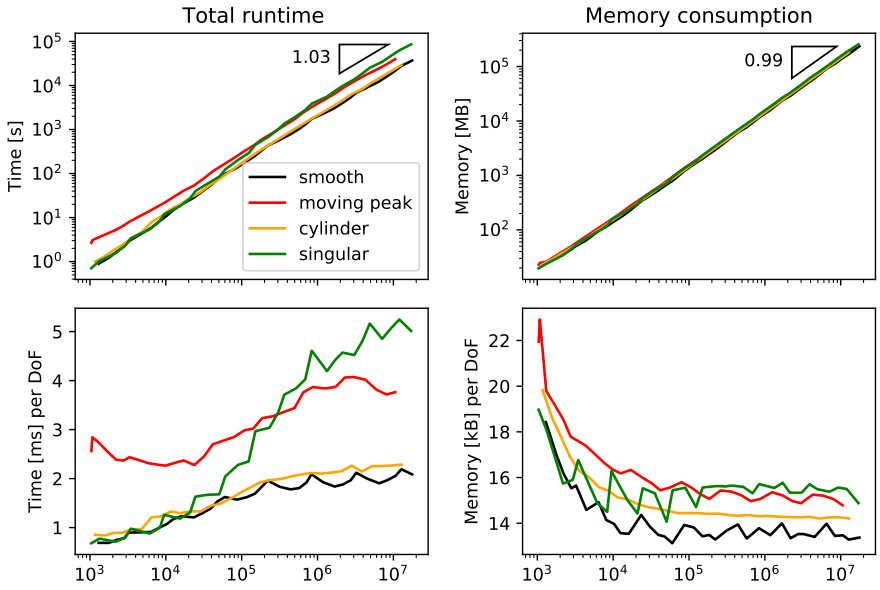


FIGURE 6.10. Total runtime and peak memory consumption as function of $\dim X^\delta$, measured after every iteration of the adaptive loop, for the four different model problems.

6.8 Conclusion

We have constructed an adaptive solver for a space-time variational formulation of parabolic evolution problems. The collection of trial spaces are given by the spans of sets of tensor products of wavelets-in-time and hierarchical basis functions-in-space. Compared to our previous works [CS11, RS19] where we employed 'true' wavelets also in space, the theoretical results are weaker. We have demonstrated r-linear convergence of the adaptive routine, but have not shown optimal rates at linear complexity. On the other hand, the runtimes that we obtained with the current approach are much better.

7 Adaptivity: an efficient implementation

7.1 Introduction

This chapter is about an *efficient* adaptive method for parabolic evolution equations using a simultaneous space-time variational formulation. Compared to the more classical time-stepping schemes, these space-time methods are very flexible. Among other things, they are especially well-suited for massively parallel computation ([NS19, vVW21a]), and some can guarantee quasi-best approximations from the trial space ([And13, FK21, SZ20]).

We are interested in those space-time methods that permit adaptive refinement locally in space *and* time. Within this class, wavelet-based methods (see [SS09, GK11, KSU16]) are attractive, as they can be shown to be *quasi-optimal*: they produce a sequence of solutions that converges at the best possible rate, at optimal linear computational cost. Moreover, they can overcome the *curse of dimensionality* using a form of *sparse tensor-product approximation*, solving the whole time evolution at a runtime proportional to that of solving the corresponding *stationary* problem.

In Chapter 6, we constructed an r-linearly converging space-time adaptive solver for parabolic evolution equations that exploits the product structure of the space-time cylinder to construct a family of trial spaces given as the spans of wavelets-in-time tensorized with (locally refined) finite element spaces-in-space.

The principal difference between this and other wavelet-based methods is that we use wavelets in time *only*, and standard finite elements in space. This eases implementation, and alleviates the need for a suitable spatial wavelet basis, which is generally difficult for general domains ([RS18]). Unfortunately, there is no free lunch: a proof of optimal convergence is, for our method, not yet available.

In this chapter we discuss an implementation of the adaptive algorithm from Chapter 6, in which the different steps (each iteration of the linear algebraic solver, the error estimation, Dörfler marking, and refinement of trial- and test spaces) are of linear complexity.

Special care has to be taken for matrix-vector products. For a bilinear form

that is 'local' and equals (a sum of) tensor-product(s) of bilinear forms in time and space, and 'trial' and 'test' spaces spanned by tensor-product multi-level bases with *double-tree* index sets, the resulting system matrix w.r.t. both bases can be applied in linear complexity, even though this matrix is not sparse. The algorithm that realizes this complexity makes a clever use of multi- to single-scale transformations alternately in time and space. This *unidirectional principle* was introduced in [BZ96] for 'uniform' sparse grids, so without 'local refinements', and it was later extended to general *downward closed* or *lower sets*, also called *adaptive sparse grids*, in [KS14]. The definition of a lower set in [KS14], there called multi-tree, is more restrictive than our current definition that allows more localized refinements.

To the best of our knowledge, other implementations for the efficient evaluation of tensor-product bilinear forms (see [Pfl10, KS14, Pab15, Rek18]) are based on the concept of hash maps. There, a hash function is used to map basis functions to array indices. In an adaptive loop, the final set of basis functions is unknown in advance so it is impossible to construct a hash function that guarantees an upper bound on the number of hash collisions. Aiming at true linear complexity, we implement these operations by traversing *trees* and *double-trees*, so without the use of hash maps.

Organization

In §7.2, we look at the abstract parabolic problem, its stable discretization, and the adaptive routine. In §7.3, we provide an abstract algorithm for the efficient evaluation of tensor-product bilinear forms w.r.t. multilevel bases indexed on *double-trees*. In §7.4, we take the *heat equation* as a model problem, and provide a concrete family of trial- and test spaces with bases indexed by double-trees that permits local space-time adaptivity. In §7.5, we discuss the practical implementation of the adaptive algorithm. Finally, in §7.6, we provide extensive numerical experiments to demonstrate the linear runtime of the algorithm.

Notation

In this work, by $C \lesssim D$ we will mean that C can be bounded by a multiple of D, independently of parameters which C and D may depend on. Obviously, $C \gtrsim D$ is defined as $D \lesssim C$, and $C \eqsim D$ as $C \lesssim D$ and $C \gtrsim D$.

For normed linear spaces E and F, by $\mathcal{L}(E,F)$ we will denote the normed linear space of bounded linear mappings $E \to F$, and by $\mathcal{L}\mathrm{is}(E,F)$ its subset of boundedly invertible linear mappings $E \to F$. We write $E \hookrightarrow F$ to denote that E is continuously embedded into F. For simplicity only, we exclusively consider linear spaces over the scalar field $\mathbb{R}$.

7.2 Space-time adaptivity for a parabolic model problem

In this section, we summarize the relevant parts of Chapter 6.

Let V, H be separable Hilbert spaces of functions on some "spatial domain" such that $V \hookrightarrow H$ with dense and compact embedding. Identifying H with its dual, we obtain the Gelfand triple $V \hookrightarrow H \simeq H' \hookrightarrow V'$.

For a.e.

$$t \in I := (0, T),$$

let $a(t; \cdot, \cdot)$ denote a bilinear form on $V \times V$ so that for any $\eta, \zeta \in V$, $t \mapsto a(t; \eta, \zeta)$ is measurable on I, and such that for a.e. $t \in I$,

$$\begin{aligned} |a(t;\eta,\zeta)| &\lesssim \|\eta\|_V \|\zeta\|_V \quad (\eta, \zeta \in V) \quad (\textit{boundedness}),\\ a(t;\eta,\eta) &\gtrsim \|\eta\|_V^2 \qquad\quad (\eta \in V) \qquad (\textit{coercivity}). \end{aligned}$$

With $(A(t)\cdot)(\cdot) := a(t; \cdot, \cdot) \in \mathcal{L}\mathrm{is}(V, V')$, given a forcing function g and initial value u_0, we want to solve the *parabolic initial value problem* of

$$\text{(7.1)}\quad \text{finding } u : I \to V \text{ such that } \quad \begin{cases} \frac{\mathrm{d}u}{\mathrm{d}t}(t) + A(t)u(t) &= g(t) \quad (t \in I),\\ u(0) &= u_0. \end{cases}$$

Example 7.2.1. For the model problem of the *heat equation* on some spatial domain $\Omega \subset \mathbb{R}^d$ we select $V := H_0^1(\Omega)$, $H := L_2(\Omega)$, and $a(t; \eta, \zeta) := \int_\Omega \nabla_{\mathbf{x}} \eta \cdot \nabla_{\mathbf{x}} \zeta \, \mathrm{d}\mathbf{x}$.

In our simultaneous space-time variational formulation, the parabolic problem is to find u s.t.

$$(Bu)(v) := \int_I \langle \tfrac{\mathrm{d}u}{\mathrm{d}t}(t), v(t)\rangle_H + a(t; u(t), v(t))\mathrm{d}t \; = \int_I \langle g(t), v(t)\rangle_H =: g(v)$$

for all v from some suitable space of functions of time and space. One possibility to enforce the initial condition is by testing against additional test functions.

Theorem 7.2.2 ([SS09]). *With $X := L_2(I; V) \cap H^1(I; V')$, $Y := L_2(I; V)$, we have*

$$\begin{bmatrix} B \\ \gamma_0 \end{bmatrix} \in \mathcal{L}\mathrm{is}(X, Y' \times H),$$

where for $t \in \bar{I}$, $\gamma_t \colon u \mapsto u(t, \cdot)$ denotes the trace map. In other words,

$$\text{(7.2)}\qquad \textit{finding } u \in X \textit{ s.t. } \quad (Bu, \gamma_0 u) = (g, u_0) \quad \textit{given} \quad (g, u_0) \in Y' \times H$$

is a well-posed simultaneous space-time variational formulation of (7.1).

We define $A \in \mathcal{L}\mathrm{is}(Y, Y')$ and $\partial_t \in \mathcal{L}\mathrm{is}(X, Y')$ as

$$(Au)(v) := \int_I a(t; u(t), v(t))\mathrm{d}t\,, \quad \text{and} \quad \partial_t := B - A.$$

Following [SW21b], we assume that A is *self-adjoint*. Morever, in view of an efficient implementation, we assume that A is a finite sum of tensor-product operators. If A does not have this structure, one may alternatively consider (low-rank) tensor-product approximations of A, see e.g. [Hac12] for an overview.

We equip Y and X with *'energy'-norms*

$$\|\cdot\|_Y^2 := (A\cdot)(\cdot), \quad \|\cdot\|_X^2 := \|\partial_t \cdot\|_{Y'}^2 + \|\cdot\|_Y^2 + \|\gamma_T \cdot\|_H^2,$$

which are equivalent to the canonical norms on Y and X.

The solution u of (7.2) equals the solution of the following minimization problem

$$u = \operatorname*{argmin}_{w\in X} \|Bw - g\|_{Y'}^2 + \|\gamma_0 w - u_0\|_H^2, \tag{7.3}$$

which in turn is the second component of the solution of

$$\text{finding } (\mu, u) \in Y \times X \text{ s.t.} \quad \begin{bmatrix} A & B \\ B' & -\gamma_0'\gamma_0 \end{bmatrix} \begin{bmatrix} \mu \\ u \end{bmatrix} = \begin{bmatrix} g \\ -u_0 \end{bmatrix}. \tag{7.4}$$

Indeed, taking the Schur complement of (7.4) w.r.t. the Y-block results in the Euler-Lagrange equations of (7.3).

7.2.1 Discretizations

Take a family $(X^\delta)_{\delta\in\Delta}$ of closed subspaces of X, and define

$$u_\delta = \operatorname*{argmin}_{w\in X^\delta} \|Bw - g\|_{Y'}^2 + \|\gamma_0 w - u_0\|_H^2, \tag{7.5}$$

being the best approximation to u from X^δ w.r.t. $\|\cdot\|_X$. Solving this problem, however, is not feasible because of the presence of the dual norm. Therefore, take $(Y^\delta)_{\delta\in\Delta}$ to be a family of closed subspaces of Y such that

$$X^\delta \subseteq Y^\delta \quad (\delta\in\Delta), \quad \text{and} \quad \gamma_\Delta := \inf_{\delta\in\Delta} \inf_{0\neq w\in X^\delta} \sup_{0\neq v\in Y^\delta} \frac{(\partial_t w)(v)}{\|\partial_t w\|_{Y'}\|v\|_Y} > 0. \tag{7.6}$$

For $\underline{\delta}\in\Delta$ with $Y^{\underline{\delta}} \supseteq Y^\delta$, we replace Y' by ${Y^{\underline{\delta}}}'$ in (7.5) yielding the approximation

$$u^{\underline{\delta}\delta} = \operatorname*{argmin}_{w\in X^\delta} \|Bw - g\|_{{Y^{\underline{\delta}}}'}^2 + \|\gamma_0 w - u_0\|_H^2.$$

Notice that $u^{\underline{\delta}\delta}$ approximates u_δ in that $u^{\underline{\delta}\delta} = u_\delta$ when $Y^{\underline{\delta}} = Y$.

With $E_Y^{\underline{\delta}} : Y^{\underline{\delta}} \to Y$ and $E_X^\delta : X^\delta \to X$ denoting the trivial embeddings, $u^{\underline{\delta}\delta}$ is the second component of the solution of

$$\begin{bmatrix} {E_Y^{\underline{\delta}}}' A E_Y^{\underline{\delta}} & {E_Y^{\underline{\delta}}}' B E_X^\delta \\ {E_X^\delta}' B' E_Y^{\underline{\delta}} & -{E_X^\delta}'\gamma_0'\gamma_0 E_X^\delta \end{bmatrix} \begin{bmatrix} \mu^{\underline{\delta}\delta} \\ u^{\underline{\delta}\delta} \end{bmatrix} = \begin{bmatrix} {E_Y^{\underline{\delta}}}' g \\ -{E_X^\delta}'\gamma_0' u_0 \end{bmatrix}.$$

Taking the Schur complement w.r.t. the $Y^{\underline{\delta}}$-block then leads to the equation

$$E_X^{\delta\,\prime}(B' E_Y^{\underline{\delta}} (E_Y^{\underline{\delta}\,\prime} A E_Y^{\underline{\delta}})^{-1} E_Y^{\underline{\delta}\,\prime} B + \gamma_0' \gamma_0) E_X^{\delta} u^{\underline{\delta}\delta} = E_X^{\delta\,\prime}(B' E_Y^{\underline{\delta}} (E_Y^{\underline{\delta}\,\prime} A E_Y^{\underline{\delta}})^{-1} E_Y^{\underline{\delta}\,\prime} g + \gamma_0' u_0), \tag{7.7}$$

which has a unique solution (cf. Lemma 6.3.3) that satisfies $\|u - u^{\underline{\delta}\delta}\|_X \leq \gamma_\Delta^{-1} \|u - u_\delta\|_X$ whenever $Y^{\underline{\delta}} \supseteq Y^\delta$; cf. [SW21b, Thm. 3.7]. For now, we assume the right-hand side of (7.7) to be evaluated exactly. Later, in §7.4.5, we will discuss approximation of the right-hand side.

In view of obtaining an efficient solver, we want to replace the inverses in (7.7) while aiming to preserve quasi-optimality of the solution. To this end, let $K_Y^{\underline{\delta}} = K_Y^{\underline{\delta}\,\prime} \in \mathcal{L}\mathrm{is}(Y^{\underline{\delta}\prime}, Y^{\underline{\delta}})$ be a uniformly optimal preconditioner for $E_Y^{\underline{\delta}\,\prime} A E_Y^{\underline{\delta}}$ that can be applied in linear complexity. Then, for some $\kappa_\Delta \geq 1$ we have

$$\frac{((K_Y^{\underline{\delta}})^{-1} v)(v)}{(Av)(v)} \in [\kappa_\Delta^{-1}, \kappa_\Delta] \quad (\delta \in \Delta, v \in Y^\delta).$$

Replacing $(E_Y^{\underline{\delta}\,\prime} A E_Y^{\underline{\delta}})^{-1}$ by $K_Y^{\underline{\delta}}$, we denote the solution of (7.7) again by $u^{\underline{\delta}\delta}$. It is quasi-optimal with $\|u - u^{\underline{\delta}\delta}\|_X \leq \frac{\kappa_\Delta}{\gamma_\Delta} \|u - u_\delta\|_X$; cf. [SW21b, Rem. 3.8].

7.2.2 Adaptive refinement loop

Our adaptive loop, given in Algorithm 7.1, takes the familiar Solve, Estimate, Mark and refine steps, and is driven by an efficient and reliable 'hierarchical basis' a posteriori error estimator.

The adaptive loop below requires a saturation assumption. Define a *partial order* on Δ by $\tilde{\delta} \succeq \delta$ whenever $X^{\tilde{\delta}} \supseteq X^\delta$. Let $\delta \mapsto \underline{\delta} \succeq \delta$ be a mapping providing *saturation* in that for some $\zeta < 1$,

$$\|u - u_{\underline{\delta}}\|_X \leq \zeta \|u - u_\delta\|_X \quad (\delta \in \Delta). \tag{7.8}$$

With this choice of $\underline{\delta}$, we are interested in finding $u^\delta := u^{\underline{\delta}\delta} \in X^\delta$ that solves

$$\underbrace{E_X^{\delta\,\prime}(B' E_Y^{\underline{\delta}} K_Y^{\underline{\delta}} E_Y^{\underline{\delta}\,\prime} B + \gamma_0' \gamma_0) E_X^{\underline{\delta}}}_{S^{\underline{\delta}\delta} :=} u^\delta = \underbrace{E_X^{\delta\,\prime}(B' E_Y^{\underline{\delta}} K_Y^{\underline{\delta}} E_Y^{\underline{\delta}\,\prime} g + \gamma_0' u_0)}_{f^\delta :=}. \tag{7.9}$$

Notice that (7.9) is uniquely solvable even with $X^{\underline{\delta}}$ as 'trial space', and we use this 'room' between X^δ and $X^{\underline{\delta}}$ to our advantage. Expanding X^δ to some intermediate space $X^\delta \subset X^{\tilde{\delta}} \subset X^{\underline{\delta}}$ yields a $u^{\tilde{\delta}}$ that is a better approximation to u than u^δ; cf. Proposition 6.4.2. This function will be the successor of u^δ in our loop, and we will show that the resulting sequence of functions converges r-linearly to u; see Algorithm 7.1 and Theorem 7.2.4.

Solving

Instead of solving the symmetric positive definite system (7.9) exactly, we construct an approximate solution $\hat{u}^\delta$ using Preconditioned Conjugate Gradients (PCG). To this end, let $K_X^\delta = {K_X^\delta}' \in \mathcal{L}\mathrm{is}(X^{\delta\prime}, X^\delta)$ be a uniformly optimal preconditioner for $S^{\underline{\delta}\delta}$. Then $((K_X^\delta)^{-1}w)(w) \eqsim \|w\|_X^2 \eqsim \|K_X^\delta S^{\underline{\delta}\delta} w\|_X^2$ for $w \in X^\delta$. Writing $w = K_X^\delta S^{\underline{\delta}\delta}(u^\delta - v^\delta)$ reveals that this induces an algebraic error estimator

(7.10)
$$\beta^\delta(v^\delta) := \sqrt{(f^\delta - S^{\underline{\delta}\delta}v^\delta)(K_X^\delta(f^\delta - S^{\underline{\delta}\delta}v^\delta))} \eqsim \|u^\delta - v^\delta\|_X \quad (v^\delta \in X^\delta, \delta \in \Delta).$$

With $\hat{u}_k^\delta$ denoting the approximant at iteration k of the PCG loop, $\beta^\delta(\hat{u}_k^\delta)$ is already available as $\sqrt{\beta_k}$, for β_k the variable used in computing the next search direction.

Error estimation

Let $\Theta_\delta := \{\theta_\lambda : \lambda \in J_\delta\}$ be some uniformly X-stable basis satisfying $X^\delta \oplus \operatorname{span}\Theta_\delta = X^{\underline{\delta}}$, in that

$$\|z + \mathbf{c}^\top \Theta_\delta\|_X^2 \eqsim \|z\|_X^2 + \|\mathbf{c}\|^2 \quad (\mathbf{c} \in \ell_2(J_\delta), z \in X^\delta, \delta \in \Delta). \tag{7.11}$$

Define the trivial embedding $P^\delta : X^\delta \to X^{\underline{\delta}}$. Akin to (7.9), we define $S^{\underline{\delta}\underline{\delta}}$ and $f^{\underline{\delta}\underline{\delta}}$, and with it, the residual-based a posteriori error estimator $\mathbf{r}^\delta : X^\delta \to \ell_2(J_\delta)$, as

(7.12)
$$S^{\underline{\delta}\underline{\delta}} := {E_X^{\underline{\delta}}}'(B' E_Y^{\underline{\delta}} K_Y^{\underline{\delta}} {E_Y^{\underline{\delta}}}' B + \gamma_0'\gamma_0) E_X^{\underline{\delta}}, \quad f^{\underline{\delta}\underline{\delta}} := {E_X^{\underline{\delta}}}'(B' E_Y^{\underline{\delta}} K_Y^{\underline{\delta}} {E_Y^{\underline{\delta}}}' g + \gamma_0' u_0),$$
$$\mathbf{r}^\delta(\hat{u}^\delta) := (f^{\underline{\delta}\underline{\delta}} - S^{\underline{\delta}\underline{\delta}} P^\delta \hat{u}^\delta)(\Theta_\delta).$$

For $\hat{u}^\delta$ close to u^δ, the error estimator $\|\mathbf{r}^\delta(\hat{u}^\delta)\|$ is reliable and efficient:

Lemma 7.2.3. *Assume* (7.8) *and* (7.11), $\frac{\kappa_\Delta}{\gamma_\Delta} < \frac{1}{\zeta}$, *and fix some* $\xi > 0$ *small enough. For* $\hat{u}^\delta \in X^\delta$ *satisfying* $\beta(\hat{u}^\delta) \le \frac{\xi}{1-\xi}\|\mathbf{r}^\delta(\hat{u}^\delta)\|$, *we have*

$$\|\mathbf{r}^\delta(\hat{u}^\delta)\| \eqsim \|u - \hat{u}^\delta\|_X \quad \textit{and} \quad \|u - \hat{u}^\delta\|_X \lesssim \|u - u^\delta\|_X \quad (\delta \in \Delta).$$

Proof. For convenience, we write $\hat{\mathbf{r}}^\delta := \mathbf{r}^\delta(\hat{u}^\delta)$ and $\mathbf{r}^\delta := \mathbf{r}^\delta(u^\delta)$.

By (7.8), (7.11) and $\frac{\kappa_\Delta}{\gamma_\Delta} < \frac{1}{\zeta}$, Proposition 6.4.4 shows that

$$\|\mathbf{r}^\delta\| \eqsim \|u - u^\delta\|_X \quad (\delta \in \Delta). \tag{7.13}$$

From (7.11) one deduces that $\|\mathbf{r}^\delta - \hat{\mathbf{r}}^\delta\| \lesssim \|u^\delta - \hat{u}^\delta\|_X$; cf. (6.25). By assumption, for $\xi < 1$, we find $\beta^\delta(\hat{u}^\delta) \lesssim \xi\|\hat{\mathbf{r}}^\delta\|$. Combined this reveals

$$\|\mathbf{r}^\delta - \hat{\mathbf{r}}^\delta\| \overset{(7.11)}{\lesssim} \|u^\delta - \hat{u}^\delta\|_X \overset{(7.10)}{\eqsim} \beta^\delta(\hat{u}^\delta) \lesssim \xi\|\hat{\mathbf{r}}^\delta\|. \tag{7.14}$$

Using this, we can show reliability of the estimator by

$$\|u-\hat{u}^\delta\|_X \le \|u-u^\delta\|_X + \|u^\delta-\hat{u}^\delta\|_X \overset{(7.13),(7.10)}{\eqsim} \|\mathbf{r}^\delta\| + \beta^\delta(\hat{u}^\delta) \le \|\hat{\mathbf{r}}^\delta\| + \|\mathbf{r}^\delta-\hat{\mathbf{r}}^\delta\| + \beta^\delta(\hat{u}^\delta) \overset{(7.14)}{\lesssim} \|\hat{\mathbf{r}}^\delta\|.$$

For efficiency of the estimator, we deduce

$$\|\hat{\mathbf{r}}^\delta\| \overset{(7.13)}{\lesssim} \|u-u^\delta\|_X + \|\mathbf{r}^\delta-\hat{\mathbf{r}}^\delta\| \le \|u-\hat{u}^\delta\|_X + \|u^\delta-\hat{u}^\delta\|_X + \|\mathbf{r}^\delta-\hat{\mathbf{r}}^\delta\| \overset{(7.14)}{\lesssim} \|u-\hat{u}^\delta\|_X + \xi\|\hat{\mathbf{r}}^\delta\|,$$

so taking ξ sufficiently small and kicking back $\|\hat{\mathbf{r}}^\delta\|$ yields

$$\|\hat{\mathbf{r}}^\delta\| \lesssim \|u-\hat{u}^\delta\|_X. \tag{7.15}$$

Similarly, from (7.13) and (7.14) it follows that

$$\|\hat{\mathbf{r}}^\delta\| \lesssim \|u-u^\delta\|_X. \tag{7.16}$$

We infer quasi-optimality of $\hat{u}^\delta$ from

$$\|u-\hat{u}^\delta\|_X \overset{(7.14)}{\lesssim} \|u-u^\delta\|_X + \xi\|\hat{\mathbf{r}}^\delta\| \overset{(7.16)}{\lesssim} \|u-u^\delta\|_X. \qquad \square$$

In the solve step, we need to iterate PCG until $\beta^\delta(\hat{u}_k^\delta)/\|\mathbf{r}^\delta(\hat{u}_k^\delta)\|$ is small enough. In the algorithm below, this is ensured by the do-while loop which also avoids the (expensive) recomputation of the residual at every PCG iteration.

Marking and refinement

Denoting the output of the solve step by $\hat{u}^\delta$, we drive the adaptive loop by performing Dörfler marking on the residual $\hat{\mathbf{r}}^\delta := \mathbf{r}^\delta(\hat{u}^\delta)$, i.e., for some $\theta \in (0,1]$, we mark the smallest set $J \subset J_\delta$ for which $\|\hat{\mathbf{r}}^\delta|_J\| \ge \theta\|\hat{\mathbf{r}}^\delta\|$. We then construct the smallest $\tilde{\delta} \succeq \delta$ such that $X^{\tilde{\delta}}$ contains $\operatorname{span}\Theta_\delta|_J$.

Theorem 7.2.4 (Theorem 6.4.9 with $\eta = 0$)**.** *Assume* (7.8) *and* (7.11)*. For ξ and $\frac{\kappa_\Delta}{\gamma_\Delta} - 1$ sufficiently small with $\frac{\kappa_\Delta}{\gamma_\Delta} - 1 \downarrow 0$ when $\theta \downarrow 0$, the sequence of approximations produced by Algorithm 7.1 converges r-linearly to u, in that after every iteration, $\|u-\hat{u}^\delta\|_X$ decreases with a factor at least $\rho < 1$.*

Remark 7.2.5. In a practical implementation, to ensure termination, Algorithm 7.1 has to be complemented by an appropriate stopping criterium; cf. Algorithm 6.30.

Algorithm 7.1: Space-time adaptive refinement loop.

Data: $\theta \in (0,1]$, $\xi \in (0,1)$, $\delta := \delta_{\text{init}} \in \Delta$;
$t_\delta := \mathcal{E}^\delta(0) = \sqrt{(E_Y^\delta{}' g)(K_Y^\delta E_Y^\delta{}' g) + \|u_0\|_H^2}$;
repeat
  Solve:
  do
    Compute $\hat{u}_*^\delta \in X^\delta$ with $\beta^\delta(\hat{u}_*^\delta) \leq t_\delta/2$;
    $t_\delta := \beta^\delta(\hat{u}_*^\delta)$;
    $e_\delta := \|\mathbf{r}^\delta(\hat{u}_*^\delta)\| + t_\delta$;
  while $t_\delta > \xi e_\delta$;
  $\hat{u}^\delta := \hat{u}_*^\delta$;
  Estimate: Set $\hat{\mathbf{r}}^\delta := \mathbf{r}^\delta(\hat{u}^\delta)$;
  Mark: Mark a smallest $J \subset J_\delta$ for which $\|\hat{\mathbf{r}}^\delta|_J\| \geq \theta\|\hat{\mathbf{r}}^\delta\|$;
  Refine: Determine the smallest $\tilde{\delta} \in \Delta$ such that
  $X^{\tilde{\delta}} \supset X^\delta \oplus \operatorname{span} \Theta_\delta|_J$;
  $t_{\tilde{\delta}} := e_\delta$, $\delta := \tilde{\delta}$;

Proof. For convenience, we denote $\mathbf{r}^\delta := \mathbf{r}^\delta(u^\delta)$ and $\hat{\mathbf{r}}^\delta := \mathbf{r}^\delta(\hat{u}^\delta)$. The stopping criterium of the solve step ensures that $\beta^\delta(\hat{u}^\delta) \leq \xi\big(\|\hat{\mathbf{r}}^\delta\| + \beta^\delta(\hat{u}^\delta)\big)$, so for $\xi < 1$ we are in the situation of Lemma 7.2.3.

We have

$$\|\hat{\mathbf{r}}^\delta - \mathbf{r}^\delta\| \overset{(7.14)}{\lesssim} \xi\|\hat{\mathbf{r}}^\delta\| \leq \xi\big(\|\mathbf{r}^\delta\| + \|\hat{\mathbf{r}}^\delta - \mathbf{r}^\delta\|\big),$$

so taking ξ sufficiently small and kicking back $\|\hat{\mathbf{r}}^\delta - \mathbf{r}^\delta\|$ yields

$$\|\hat{\mathbf{r}}^\delta - \mathbf{r}^\delta\| \lesssim \xi\|\mathbf{r}^\delta\|. \tag{7.17}$$

After marking, we have $\|\hat{\mathbf{r}}^\delta\| \leq \theta^{-1}\|\hat{\mathbf{r}}^\delta|_J\|$, which shows that

$$\|\mathbf{r}^\delta\| \overset{(7.14)}{\lesssim} \|\hat{\mathbf{r}}^\delta\| \lesssim \|\hat{\mathbf{r}}^\delta|_J\| \leq \|\mathbf{r}^\delta|_J\| + \|\mathbf{r}^\delta - \hat{\mathbf{r}}^\delta\| \overset{(7.17)}{\lesssim} \|\mathbf{r}^\delta|_J\| + \xi\|\mathbf{r}^\delta\|,$$

so for ξ small enough, kicking back $\|\mathbf{r}^\delta\|$ reveals that for a $\hat{\theta} > 0$ dependent on θ,

$$\|\mathbf{r}^\delta|_J\| \geq \hat{\theta}\|\mathbf{r}^\delta\|.$$

From Proposition 6.4.3 we now find that, for $\frac{\kappa_\Delta}{\gamma_\Delta} - 1 \downarrow 0$ when $\theta \downarrow 0$, there is a $\bar{\rho} < 1$ for which

$$\|u - u^{\tilde{\delta}}\|_X \leq \bar{\rho}\|u - u^\delta\|_X. \tag{7.18}$$

Combining the results shows that

$$\begin{aligned}\|u-\hat{u}^{\bar{\delta}}\|_X &\leq \|u-u^{\bar{\delta}}\|_X + \|u^{\bar{\delta}}-\hat{u}^{\bar{\delta}}\|_X \\ &\overset{(7.14),(7.16)}{\leq} (1+\mathcal{O}(\xi))\|u-u^{\bar{\delta}}\|_X \\ &\overset{(7.18)}{\leq} (1+\mathcal{O}(\xi))\bar{\rho}\|u-u^{\delta}\|_X \\ &\leq (1+\mathcal{O}(\xi))\bar{\rho}(\|u-\hat{u}^{\delta}\|_X + \|u^{\delta}-\hat{u}^{\delta}\|_X) \\ &\overset{(7.14),(7.15)}{\leq} \underbrace{(1+\mathcal{O}(\xi))\bar{\rho}}_{=:\rho}\|u-\hat{u}^{\delta}\|_X,\end{aligned}$$

so for ξ small enough, $\rho < 1$, completing the proof of r-linear convergence. □

7.2.3 Adaptive trial- and test spaces

The convergence rate of our adaptive loop is determined by the approximation properties of the family $(X^\delta)_{\delta\in\Delta}$. We want to construct a family that allows for *local* refinements. Here, the crucial problem is guaranteeing the inf-sup stability condition (7.6). It is known that inf-sup stability is satisfied for *full* tensor-products of (non-uniform) finite element spaces, and in [And13, Prop. 4.2], this result was generalized to families of *sparse* tensor-products. Unfortunately, neither family allows for adaptive refinements both locally in time and space.

In §7.4 we will solve this by first equipping X with a tensor-product of (infinite) bases: a wavelet basis Σ in time, and a hierarchical basis in space. We then construct X^δ as the span of a (finite) subset of this tensor-product basis, which we grow by adding particular functions.

By imposing a *double-tree* constraint on the index set of the basis of X^δ, we can apply tensor-product operators in linear complexity; see §7.3. Moreover, this constraint implies that for our model problem the inf-sup condition (7.6) is satisfied and we can construct optimal preconditioners K_Y^δ and K_X^δ.

7.3 The application of linear operators in linear complexity

An efficient implementation of our adaptive loop requires the efficient application of the operators ${E_Y^\delta}' B E_X^\delta$ and ${E_X^\delta}' \gamma_0' \gamma_0 E_X^\delta$ appearing in (7.9). Both terms are finite sums of tensor-products of operators in time and space. When we equip our trial and test spaces with tensor-products of multilevel bases, it turns out that we can evaluate these operators in linear complexity.

More precisely, this section will show the abstract result that given

- tensor-products $\Psi := \Psi^0 \times \Psi^1$, $\breve{\Psi} := \breve{\Psi}^0 \times \breve{\Psi}^1$ of multilevel bases Ψ^0, Ψ^1, $\breve{\Psi}^0$, $\breve{\Psi}^1$ indexed by $\vee^0$, $\vee^1$, $\breve{\vee}^0$, $\breve{\vee}^1$, and

- (finite) subsets $\boldsymbol{\Lambda} \subset \vee^0 \times \vee^1, \breve{\boldsymbol{\Lambda}} \subset \breve{\vee}^0 \times \breve{\vee}^1$ that are *double-trees*, and
- linear operators $A_i : \operatorname{span} \Psi^i \to (\operatorname{span} \breve{\Psi}^i)'$ that are *local* ($i \in \{0,1\}$),

we can apply the matrix $((A_0 \otimes A_1)\Psi|_{\boldsymbol{\Lambda}})(\breve{\Psi}|_{\breve{\boldsymbol{\Lambda}}})$ in $\mathcal{O}(\#\boldsymbol{\Lambda} + \#\breve{\boldsymbol{\Lambda}})$ operations even though this matrix is not uniformly sparse.

Example 7.3.1. For our model problem, Ψ^0 and $\breve{\Psi}^0$ will be wavelets for $H^1(I)$ or $L_2(I)$ in time, and $\Psi^1 = \breve{\Psi}^1$ will be a hierarchical finite element basis for $H^1_0(\Omega)$ in space. We will apply the result of this section to the operators $\gamma_0'\gamma_0$ and $B = \partial_t + A$.

We will achieve this complexity using a variant of the *unidirectional principle*. Denote with $I_{\boldsymbol{\Lambda}}$ the extension with zeros of a vector supported on $\boldsymbol{\Lambda}$ to one on $\vee^0 \times \vee^1$, and with $R_{\boldsymbol{\Lambda}}$ its adjoint; define $I_{\breve{\boldsymbol{\Lambda}}}$ and $R_{\breve{\boldsymbol{\Lambda}}}$ analogously. Define $\mathbf{A}_i := (A_i\Psi^i)(\breve{\Psi}^i)$. We will split $\mathbf{A}_0$ in its upper and strictly lower triangular parts $\mathbf{U}_0$ and $\mathbf{L}_0$, so that

$$R_{\breve{\boldsymbol{\Lambda}}}(\mathbf{A}_0 \otimes \mathbf{A}_1) I_{\boldsymbol{\Lambda}} = R_{\breve{\boldsymbol{\Lambda}}}(\mathbf{L}_0 \otimes \mathrm{Id})(\mathrm{Id} \otimes \mathbf{A}_1) I_{\boldsymbol{\Lambda}} + R_{\breve{\boldsymbol{\Lambda}}}(\mathbf{U}_0 \otimes \mathrm{Id})(\mathrm{Id} \otimes \mathbf{A}_1) I_{\boldsymbol{\Lambda}}.$$

This in itself is not useful, as $(\mathrm{Id} \otimes \mathbf{A}_1) I_{\boldsymbol{\Lambda}}$ maps into a vector space which dimension we cannot control. However, the restriction $R_{\breve{\boldsymbol{\Lambda}}}$ gives us elbow room: in Theorem 7.3.13 we construct double-trees $\boldsymbol{\Sigma}, \boldsymbol{\Theta}$ with $\#\boldsymbol{\Sigma} + \#\boldsymbol{\Theta} \lesssim \#\breve{\boldsymbol{\Lambda}} + \#\boldsymbol{\Lambda}$ s.t.

$$\begin{cases} R_{\breve{\boldsymbol{\Lambda}}}(\mathbf{L}_0 \otimes \mathrm{Id})(\mathrm{Id} \otimes \mathbf{A}_1) I_{\boldsymbol{\Lambda}} = R_{\breve{\boldsymbol{\Lambda}}}(\mathbf{L}_0 \otimes \mathrm{Id}) R_{\boldsymbol{\Sigma}} I_{\boldsymbol{\Sigma}} (\mathrm{Id} \otimes \mathbf{A}_1) I_{\boldsymbol{\Lambda}}, \\ R_{\breve{\boldsymbol{\Lambda}}}(\mathbf{U}_0 \otimes \mathrm{Id})(\mathrm{Id} \otimes \mathbf{A}_1) I_{\boldsymbol{\Lambda}} = R_{\breve{\boldsymbol{\Lambda}}}(\mathbf{U}_0 \otimes \mathrm{Id}) R_{\boldsymbol{\Theta}} I_{\boldsymbol{\Theta}} (\mathrm{Id} \otimes \mathbf{A}_1) I_{\boldsymbol{\Lambda}}. \end{cases} \tag{7.19}$$

These right hand sides we *can* apply efficiently, and their application boils down to applications of $\mathbf{L}_0$, $\mathbf{U}_0$, and $\mathbf{A}_1$ in a *single* coordinate direction only. Simple matrix-vector products are inefficient though, as these matrices are again not uniformly sparse. However, by using the properties of a double-tree and the sparsity of the operator in *single scale*, we can evaluate $\mathbf{U}_0, \mathbf{L}_0$ and $\mathbf{A}_1$ in linear time; see §7.3.1.

We follow the structure of [KS14, §3], which applies the aforementioned idea to *multi-trees* though with a slightly more restrictive definition of a *tree*. For readability, we defer the proofs of Theorems 7.3.7, 7.3.9, 7.3.11, and 7.3.13 to Appendix 7.A.

7.3.1 Evaluation of linear operators w.r.t. trees

Let Ψ be a (multilevel) collection of functions on some domain Q.

Example 7.3.2. In our application, Q will be either the time interval I with Ψ being a collection of wavelets, or the spatial domain Ω, in which case Ψ is a collection of hierarchical basis functions.

Writing $\Psi = \{\psi_\lambda : \lambda \in \vee\}$, we assume that the ψ_λ are *locally supported* in the sense that with $|\lambda| \in \mathbb{N}_0$ denoting the *level* of λ,

$$\sup_{\lambda\in\vee} 2^{|\lambda|} \operatorname{diam}\operatorname{supp}\psi_\lambda < \infty, \tag{7.20}$$

$$\sup_{\ell\in\mathbb{N}_0}\sup_{x\in Q} \#\{\lambda \in \vee : |\lambda| = \ell \wedge \operatorname{supp}\psi_\lambda \cap B(x; 2^{-\ell}) \neq \emptyset\} < \infty. \tag{7.21}$$

We will refer to the functions ψ_λ as being *wavelets*, although not necessarily they have vanishing moments or other specific wavelet properties.

For $\ell \in \mathbb{N}_0$, and any $\Lambda \subset \vee$, we set $\Lambda_\ell := \{\lambda \in \Lambda : |\lambda| = \ell\}$ and $\Lambda_{\ell\uparrow} := \{\lambda \in \Lambda : |\lambda| \geq \ell\}$, and write $\Psi_\ell := \Psi|_{\vee_\ell}$.

For $\ell \in \mathbb{N}_0$, we assume a collection $\Phi_\ell = \{\phi_\lambda : \lambda \in \Delta_\ell\}$, whose members will be referred to as being *scaling functions*, with

$$\operatorname{span}\Phi_{\ell+1} \supseteq \operatorname{span}\Phi_\ell \cup \Psi_{\ell+1}, \quad \Phi_0 = \Psi_0 \quad (\Delta_0 := \vee_0), \tag{7.22}$$

$$\sup_{\ell\in\mathbb{N}_0}\sup_{\lambda\in\Delta_\ell} 2^\ell \operatorname{diam}\operatorname{supp}\phi_\lambda < \infty, \tag{7.23}$$

$$\sup_{\ell\in\mathbb{N}_0}\sup_{x\in Q} \#\{\lambda \in \Delta_\ell : \operatorname{supp}\phi_\lambda \cap B(x; 2^{-\ell}) \neq \emptyset\} < \infty, \tag{7.24}$$

$$\{\phi_\lambda|_\Sigma : \lambda \in \Delta_\ell,\ \phi_\lambda|_\Sigma \not\equiv 0\} \text{ is independent (for all open } \Sigma \subset Q,\ \ell \in \mathbb{N}_0). \tag{7.25}$$

W.l.o.g. we assume that the index sets Δ_ℓ for different ℓ are mutually disjoint, and set $\Phi := \cup_{\ell\in\mathbb{N}_0}\Phi_\ell$ with index set $\Delta := \cup_{\ell\in\mathbb{N}_0}\Delta_\ell$. For $\lambda \in \Delta$, we set $|\lambda| := \ell$ when $\lambda \in \Delta_\ell$.

Viewing Ψ_ℓ, Φ_ℓ as column vectors, the assumptions we made so far guarantee the existence of matrices $\mathfrak{p}_\ell$, $\mathfrak{q}_\ell$ such that

$$\left[(\Phi_{\ell-1})^\top \quad (\Psi_\ell)^\top\right] = (\Phi_\ell)^\top \left[\mathfrak{p}_\ell \quad \mathfrak{q}_\ell\right],$$

where the number of non-zeros per row and column of $\mathfrak{p}_\ell$ and $\mathfrak{q}_\ell$ is finite, uniformly in the rows and columns and in $\ell \in \mathbb{N}$ (here also (7.25) has been used). We refer to $\mathfrak{p}_\ell$ as the *prolongation matrix*. Columns of $\mathfrak{p}_\ell$ contain the *mask* of the scaling functions, whereas columns of $\mathfrak{q}_\ell$ contain the mask of the wavelets.

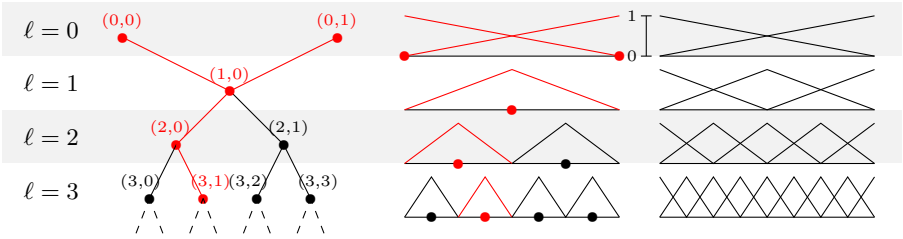
To each $\lambda \in \vee$ with $|\lambda| > 0$, we associate one or more $\mu \in \vee$ with $|\mu| = |\lambda| - 1$ and $|\operatorname{supp}\psi_\lambda \cap \operatorname{supp}\psi_\mu| > 0$. We call μ a *parent* of λ, and so λ a *child* of μ.

To each $\lambda \in \vee$, we associate some neighbourhood $S(\lambda)$ of $\operatorname{supp}\psi_\lambda$, with diameter $\lesssim 2^{-|\lambda|}$, such that for $|\lambda| > 0$, $S(\lambda) \subset \cup_{\mu\in\mathtt{parent}(\lambda)} S(\mu)$.

Remark 7.3.3. Such a neighborhood always exists even when a child has only one parent. Indeed with $C := \sup_{\lambda\in\vee} 2^{|\lambda|}\operatorname{diam}\operatorname{supp}\psi_\lambda$ and $S(\lambda) := \{x \in Q : \operatorname{dist}(x, \operatorname{supp}\psi_\lambda) < C2^{-|\lambda|}\}$, for μ being a parent of λ and $x \in S(\lambda)$, $\operatorname{dist}(x, \operatorname{supp}\psi_\mu) \leq \operatorname{dist}(x, \operatorname{supp}\psi_\lambda) + \operatorname{diam}\operatorname{supp}\psi_\lambda < 2C2^{-|\lambda|} = C2^{-|\mu|}$, i.e., $x \in S(\mu)$.

Definition 7.3.4 (Tree)**.** A finite $\Lambda \subset \vee_{\ell\uparrow}$ is called an *ℓ-tree*, or simply a *tree* when $\ell = 0$, when for any $\lambda \in \Lambda$ its parents in $\vee_{\ell\uparrow}$ are in Λ. This is not a tree in the graph-theoretical sense, but rather one in the sense of a family history tree.

Example 7.3.5 (Hierarchical basis in one dimension). Figure 7.1 shows an example multilevel collection Ψ of functions defined on the interval $[0,1]$. Its index set $\vee_{\mathfrak{I}}$ with parent-child relations is shown left, with a tree $\Lambda \subset \vee_{\mathfrak{I}}$ visualised in red. This collection is called the *hierarchical basis*. With $S(\lambda) := \operatorname{supp} \psi_\lambda$ for $\lambda \in \vee_{\mathfrak{I}}$, the hierarchical basis satisfies conditions mentioned above.



Index set $\vee_{\mathfrak{I}}$ and tree $\Lambda \subset \vee_{\mathfrak{I}}$ Multilevel functions Ψ Scaling functions Φ

Figure 7.1. Hierarchical basis for the interval $[0,1]$.

A routine eval

Let (Ψ, Φ) and $(\breve{\Psi}, \breve{\Phi})$ satisfy the conditions of the previous subsection, and let $A\colon \operatorname{span}\Phi \to (\operatorname{span}\breve{\Phi})'$ be *local* in that $(Au)(v) = (Au|_{\operatorname{supp} v})(v)$. Typically, A is a (partial) differential operator in variational form; e.g. $A \in \mathcal{L}(H^1(I), L_2(I)')$ with $(Au)(v) = \int_I \frac{\mathrm{d}u}{\mathrm{d}t} v \,\mathrm{d}t$. For trees $\Lambda \subset \vee$ and $\breve{\Lambda} \in \breve{\vee}$, we are interested in the efficient application of the matrix $(A\Psi|_\Lambda)(\breve{\Psi}|_{\breve{\Lambda}})$.

Just for brevity of the following argument, assume $\Psi = \breve{\Psi}$ and $\Phi = \breve{\Phi}$. The matrix $(A\Psi|_\Lambda)(\Psi|_\Lambda)$ is not uniformly sparse, so a straight-forward matrix-vector product is not of linear complexity. However, for Λ a uniform tree up to level ℓ, i.e. $\Lambda = \{\lambda \in \vee : |\lambda| \le \ell\}$, a solution is provided by the multi- to single-scale transform T characterized by $\Psi|_\Lambda = T^\top \Phi_\ell$ through the equality $(A\Psi|_\Lambda)(\Psi|_\Lambda) = T^\top (A\Phi_\ell)(\Phi_\ell) T$, as the transforms can be applied in linear complexity and the single-scale matrix is uniformly sparse.

For general trees however, we don't have $\dim \Phi_\ell \lesssim \dim \Psi|_\Lambda$ so the previous approach is not of linear complexity. Clever level-by-level multi-to-singlescale transformations and the prolongation of *only* relevant functions *does* allow applying $(A\Psi|_\Lambda)(\breve{\Psi}|_{\breve{\Lambda}})$ in linear complexity; see Algorithm 7.2 below.

On several places the restriction of a vector (of scalars or of functions) to its indices in some subset of the index set should be read as the vector of full length where the entries with indices outside this subset are replaced by zeros. For index sets Δ and $\breve{\Delta}$, matrix $\mathtt{m} \in \mathbb{R}^{\#\breve{\Delta}\times\#\Delta}$, and subset $\Pi \subset \Delta$, we write $\operatorname{supp}(\mathtt{m}, \Pi) \subset \breve{\Delta}$ for the index set corresponding to the image of $\mathtt{m}$ under $\{\mathbf{x}|_\Pi : \mathbf{x} \in \mathbb{R}^{\#\Delta}\}$.

Algorithm 7.2: Function $\mathtt{eval}(A)$.

Data: $\ell \in \mathbb{N}$, $\breve{\Pi} \subset \breve{\Delta}_{\ell-1}$, $\Pi \subset \Delta_{\ell-1}$, ℓ-trees $\breve{\Lambda} \subset \breve{\vee}_{\ell\uparrow}$ and $\Lambda \subset \vee_{\ell\uparrow}$, $\mathbf{d} \in \mathbb{R}^{\#\Pi}$, $\mathbf{c} \in \mathbb{R}^{\#\Lambda}$.

Result: $[\mathbf{e}, \mathbf{f}]$ where $\mathbf{e} = (Au)(\breve{\Phi}|_{\breve{\Pi}})$, $\mathbf{f} = (Au)(\breve{\Psi}|_{\breve{\Lambda}})$, with $u := \mathbf{d}^\top \Phi|_\Pi + \mathbf{c}^\top \Psi|_\Lambda$.

if $\breve{\Pi} \cup \breve{\Lambda} \neq \emptyset$ **then**

$\breve{\Pi}_B := \{\lambda \in \breve{\Pi} : \big|\operatorname{supp}\breve{\phi}_\lambda \cap \cup_{\mu\in\Lambda_\ell} S(\mu)\big| > 0\}$, $\breve{\Pi}_A := \breve{\Pi} \setminus \breve{\Pi}_B$

$\Pi_B := \{\lambda \in \Pi : \big|\operatorname{supp}\phi_\lambda \cap \big(\cup_{\mu\in\breve{\Lambda}_\ell} \breve{S}(\mu) \cup_{\gamma\in\breve{\Pi}_B} \operatorname{supp}\breve{\phi}_\gamma\big)\big| > 0\}$, $\Pi_A := \Pi \setminus \Pi_B$

$\underline{\breve{\Pi}} := \operatorname{supp}(\breve{\mathfrak{p}}_\ell, \breve{\Pi}_B) \cup \operatorname{supp}(\breve{\mathfrak{q}}_\ell, \breve{\Lambda}_\ell)$

$\underline{\Pi} := \operatorname{supp}(\mathfrak{p}_\ell, \Pi_B) \cup \operatorname{supp}(\mathfrak{q}_\ell, \Lambda_\ell)$

$\underline{\mathbf{d}} := \mathfrak{p}_\ell \mathbf{d}|_{\Pi_B} + \mathfrak{q}_\ell \mathbf{c}|_{\Lambda_\ell}$

$[\underline{\mathbf{e}}, \underline{\mathbf{f}}] := \mathtt{eval}(A)(\ell+1, \underline{\breve{\Pi}}, \breve{\Lambda}_{\ell+1\uparrow}, \underline{\Pi}, \Lambda_{\ell+1\uparrow}, \underline{\mathbf{d}}, \mathbf{c}|_{\Lambda_{\ell+1\uparrow}})$

$$\mathbf{e} = \begin{bmatrix} \mathbf{e}|_{\breve{\Pi}_A} \\ \mathbf{e}|_{\breve{\Pi}_B} \end{bmatrix} := \begin{bmatrix} (A\Phi|_\Pi)(\breve{\Phi}|_{\breve{\Pi}_A})\mathbf{d} \\ (\breve{\mathfrak{p}}_\ell^\top \underline{\mathbf{e}})|_{\breve{\Pi}_B} \end{bmatrix}$$

$$\mathbf{f} = \begin{bmatrix} \mathbf{f}|_{\breve{\Lambda}_\ell} \\ \mathbf{f}|_{\breve{\Lambda}_{\ell+1\uparrow}} \end{bmatrix} := \begin{bmatrix} (\breve{\mathfrak{q}}_\ell^\top \underline{\mathbf{e}})|_{\breve{\Lambda}_\ell} \\ \underline{\mathbf{f}} \end{bmatrix}$$

Remark 7.3.6. Let $\breve{\Lambda} \subset \breve{\vee}$, $\Lambda \subset \vee$ be trees, and $\mathbf{c} \in \ell_2(\Lambda)$, then

$$[\mathbf{e}, \mathbf{f}] := \mathtt{eval}(A)(1, \breve{\Lambda}_0, \breve{\Lambda}_{1\uparrow}, \Lambda_0, \Lambda_{1\uparrow}, \mathbf{c}|_{\Lambda_0}, \mathbf{c}|_{\Lambda_{1\uparrow}}),$$

satisfies

$$(A\Psi|_\Lambda)(\breve{\Psi}|_{\breve{\Lambda}})\mathbf{c} = \begin{bmatrix} \mathbf{e} \\ \mathbf{f} \end{bmatrix}.$$

Theorem 7.3.7. *A call of* `eval` *yields the output as specified, at the cost of* $\mathcal{O}(\#\breve{\Pi} + \#\breve{\Lambda} + \#\Pi + \#\Lambda)$ *operations.*

Proof. See Appendix 7.A. □

Routines `evalupp` and `evallow`

Let $A\colon \operatorname{span}\Phi \to (\operatorname{span}\breve{\Phi})'$ be local and *linear*. Set

$$\mathbf{A} := (A\Psi)(\breve{\Psi}) = [(A\psi_\mu)(\breve{\psi}_\lambda)]_{(\lambda,\mu)\in\breve{\vee}\times\vee}$$

as well as $\mathbf{U} := [(A\psi_\mu)(\breve{\psi}_\lambda)]_{|\lambda|\le|\mu|}$ and $\mathbf{L} := [(A\psi_\mu)(\breve{\psi}_\lambda)]_{|\lambda|>|\mu|}$ so that $\mathbf{A} = \mathbf{L}+\mathbf{U}$. As sketched in the introduction of this section, this splitting is going to be necessary for the application of system matrices in the tensor-product setting; see also (7.19). Algorithms 7.3 and 7.4 below can be used to evaluate $\mathbf{U}$ and $\mathbf{L}$ in linear complexity.

Algorithm 7.3: Function `evalupp`(A).

Data: $\ell\in\mathbb{N}$, $\breve{\Pi}\subset\breve{\Delta}_{\ell-1}$, $\Pi\subset\Delta_{\ell-1}$, ℓ-trees $\breve{\Lambda}\subset\breve{\vee}_{\ell\uparrow}$ and $\Lambda\subset\vee_{\ell\uparrow}$, $\mathbf{d}\in\mathbb{R}^{\#\Pi}$, $\mathbf{c}\in\mathbb{R}^{\#\Lambda}$.

Result: $[\mathbf{e},\mathbf{f}]$ where $\mathbf{e}=(Au)(\breve{\Phi}|_{\breve{\Pi}})$, $\mathbf{f}=\mathbf{U}|_{\breve{\Lambda}\times\Lambda}\mathbf{c}$, with $u:=\mathbf{d}^\top\Phi|_\Pi+\mathbf{c}^\top\Psi|_\Lambda$.

if $\breve{\Pi}\cup\breve{\Lambda}\neq\emptyset$ **then**

$\breve{\Pi}_B := \{\lambda\in\breve{\Pi}:\big|\operatorname{supp}\breve{\phi}_\lambda\cap\cup_{\mu\in\Lambda_\ell}S(\mu)\big|>0\}$, $\breve{\Pi}_A := \breve{\Pi}\setminus\breve{\Pi}_B$

$\underline{\breve{\Pi}} := \operatorname{supp}(\breve{\mathfrak{p}}_\ell,\breve{\Pi}_B)\cup\operatorname{supp}(\breve{\mathfrak{q}}_\ell,\breve{\Lambda}_\ell)$

$\underline{\Pi} := \operatorname{supp}(\mathfrak{q}_\ell,\Lambda_\ell)$

$\underline{\mathbf{d}} := \mathfrak{q}_\ell\mathbf{c}|_{\Lambda_\ell}$

$[\underline{\mathbf{e}},\underline{\mathbf{f}}] := \texttt{evalupp}(A)(\ell+1,\underline{\breve{\Pi}},\breve{\Lambda}_{\ell+1\uparrow},\underline{\Pi},\Lambda_{\ell+1\uparrow},\underline{\mathbf{d}},\mathbf{c}|_{\Lambda_{\ell+1\uparrow}})$

$$\mathbf{e}=\begin{bmatrix}\mathbf{e}|_{\breve{\Pi}_A}\\ \mathbf{e}|_{\breve{\Pi}_B}\end{bmatrix} := \begin{bmatrix}(A\Phi|_\Pi)(\breve{\Phi}|_{\breve{\Pi}_A})\mathbf{d}\\ (A\Phi|_\Pi)(\breve{\Phi}|_{\breve{\Pi}_B})\mathbf{d}+(\breve{\mathfrak{p}}_\ell^\top\underline{\mathbf{e}})|_{\breve{\Pi}_B}\end{bmatrix}$$

$$\mathbf{f}=\begin{bmatrix}\mathbf{f}|_{\breve{\Lambda}_\ell}\\ \mathbf{f}|_{\breve{\Lambda}_{\ell+1\uparrow}}\end{bmatrix} := \begin{bmatrix}(\breve{\mathfrak{q}}_\ell^\top\underline{\mathbf{e}})|_{\breve{\Lambda}_\ell}\\ \underline{\mathbf{f}}\end{bmatrix}$$

Remark 7.3.8. Let $\breve{\Lambda}\subset\breve{\vee}$, $\Lambda\subset\vee$ be trees, and $\mathbf{c}\in\ell_2(\Lambda)$, then

$$[\mathbf{e},\,\mathbf{f}] := \texttt{evalupp}(A)(1,\breve{\Lambda}_0,\breve{\Lambda}_{1\uparrow},\Lambda_0,\Lambda_{1\uparrow},\mathbf{c}|_{\Lambda_0},\mathbf{c}|_{\Lambda_{1\uparrow}}),$$

satisfies

$$\mathbf{U}|_{\breve{\Lambda}\times\Lambda}\mathbf{c}=\begin{bmatrix}\mathbf{e}\\ \mathbf{f}\end{bmatrix}.$$

Theorem 7.3.9. *A call of* `evalupp` *yields the output as specified, at the cost of* $\mathcal{O}(\#\breve{\Pi}+\#\breve{\Lambda}+\#\Pi+\#\Lambda)$ *operations.*

Proof. See Appendix 7.A. □

Algorithm 7.4: Function $\texttt{evallow}(A)$.

Data: $\ell \in \mathbb{N}$, $\Pi \subset \Delta_{\ell-1}$, ℓ-trees $\breve{\Lambda} \subset \breve{\vee}_{\ell\uparrow}$ and $\Lambda \subset \vee_{\ell\uparrow}$, $\mathbf{d} \in \mathbb{R}^{\#\Pi}$, $\mathbf{c} \in \mathbb{R}^{\#\Lambda}$.
Result: $\mathbf{f} = (A\Phi|_{\Pi})(\breve{\Psi}|_{\breve{\Lambda}})\mathbf{d} + \mathbf{L}|_{\breve{\Lambda}\times\Lambda}\mathbf{c}$.
if $\breve{\Pi} \cup \breve{\Lambda} \neq \emptyset$ **then**
$\Pi_B := \{\lambda \in \Pi : \left|\operatorname{supp}\phi_\lambda \cap \cup_{\mu\in\breve{\Lambda}_\ell} \breve{S}(\mu)\right| > 0\}$,
$\underline{\Pi} := \operatorname{supp}(\mathfrak{p}_\ell, \Pi_B) \cup \operatorname{supp}(\mathfrak{q}_\ell, \Lambda_\ell)$
$\underline{\Pi}_B := \operatorname{supp}(\mathfrak{p}_\ell, \Pi_B)$
$\breve{\underline{\Pi}} := \operatorname{supp}(\breve{\mathfrak{q}}_\ell, \breve{\Lambda}_\ell)$
$\underline{\mathbf{d}} := \mathfrak{p}_\ell\mathbf{d}|_{\Pi_B} + \mathfrak{q}_\ell\mathbf{c}|_{\Lambda_\ell}$
$\underline{\mathbf{e}} := (A\Phi|_{\underline{\Pi}_B})(\breve{\Phi}|_{\breve{\underline{\Pi}}})\mathfrak{p}_\ell\mathbf{d}|_{\Pi_B}$
$\mathbf{f} = \begin{bmatrix}\mathbf{f}|_{\breve{\Lambda}_\ell} \\ \mathbf{f}|_{\breve{\Lambda}_{\ell+1\uparrow}}\end{bmatrix} := \begin{bmatrix}(\breve{\mathfrak{q}}_\ell^\top \underline{\mathbf{e}})|_{\breve{\Lambda}_\ell} \\ \texttt{evallow}(A)(\ell+1, \breve{\Lambda}_{\ell+1\uparrow}, \underline{\Pi}, \Lambda_{\ell+1\uparrow}, \underline{\mathbf{d}}, \mathbf{c}|_{\Lambda_{\ell+1\uparrow}})\end{bmatrix}$

Remark 7.3.10. Let $\breve{\Lambda} \subset \breve{\vee}$, $\Lambda \subset \vee$ be trees, and $\mathbf{c} \in \ell_2(\Lambda)$, then

$$\mathbf{L}|_{\breve{\Lambda}\times\Lambda}\mathbf{c} = \texttt{evallow}(A)(1, \breve{\Lambda}_{1\uparrow}, \Lambda_0, \Lambda_{1\uparrow}, \mathbf{c}|_{\Lambda_0}, \mathbf{c}|_{\Lambda_{1\uparrow}}).$$

Theorem 7.3.11. *A call of* `evallow` *yields the output as specified, at the cost of* $\mathcal{O}(\#\breve{\Lambda} + \#\Pi + \#\Lambda)$ *operations.*

Proof. See Appendix 7.A. □

7.3.2 Application of tensor-product operators w.r.t. double-trees

For $i \in \{0, 1\}$, let $A_i \colon \operatorname{span}\Phi_i \to \operatorname{span}\breve{\Phi}_i'$ be local and linear and let

$$\mathbf{A}_i = (A\Psi_i)(\breve{\Psi}_i) = [(A\psi_\mu^i)(\breve{\psi}_\lambda^i)]_{\lambda\in\breve{\vee}^i, \mu\in\vee^i} = \mathbf{L}_i + \mathbf{U}_i.$$

where $\mathbf{U}_i := [(\mathbf{A}_i)_{\lambda,\mu}]_{|\lambda|\leq|\mu|}$ and $\mathbf{L}_i := [(\mathbf{A}_i)_{\lambda,\mu}]_{|\lambda|>|\mu|}$. For $i \in \{0, 1\}$, let $\neg i := 1 - i$.

Definition 7.3.12 (Double-tree)**.** Define the coordinate projector $P_i(b_0, b_1) := b_i$. We call $\mathbf{\Lambda} \subset \{\breve{\vee}^0 \times \breve{\vee}^1, \vee^0 \times \breve{\vee}^1, \breve{\vee}^0 \times \vee^1, \vee^0 \times \vee^1\}$, a *double-tree* when for $i \in \{0, 1\}$ and any $\mu \in P_{\neg i}\mathbf{\Lambda}$, the *fiber*

$$\mathbf{\Lambda}_{i,\mu} := P_i(P_{\neg i}|_{\mathbf{\Lambda}})^{-1}\{\mu\}$$

is a tree (in $\breve{\vee}^i$ or $\vee^i$), i.e., $\mathbf{\Lambda}$ is a double-tree when 'frozen' in each of its coordinates, at any value of that coordinate, it is a tree in the remaining coordinate.

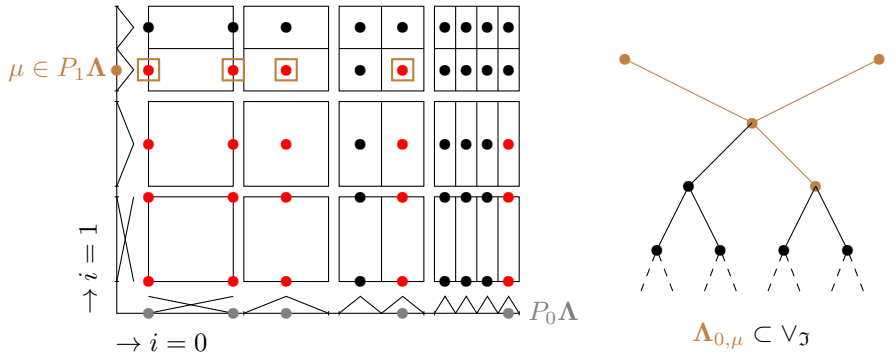


Figure 7.2. With $\vee_{\mathfrak{I}}$ from Figure 7.1: $\vee_{\mathfrak{I}} \times \vee_{\mathfrak{I}}$ in black; a double-tree $\mathbf{\Lambda} \subset \vee_{\mathfrak{I}} \times \vee_{\mathfrak{I}}$ in red; the projection $P_0\mathbf{\Lambda}$ in gray, and a fiber $\mathbf{\Lambda}_{0,\mu}$ for $\mu \in P_1\mathbf{\Lambda}$ in brown.

From $\mathbf{\Lambda} = \cup_{\mu \in P_{\neg i}\mathbf{\Lambda}} (P_{\neg i}|_{\mathbf{\Lambda}})^{-1}\{\mu\}$, we have $P_i\mathbf{\Lambda} = \cup_{\mu \in P_{\neg i}\mathbf{\Lambda}} \mathbf{\Lambda}_{i,\mu}$, which, being a union of trees, is a tree itself. See also Figure 7.2.

For a subset $\lhd$ of a (double) index set $\Diamond$, let $I_{\lhd}^{\Diamond}$ denote the extension operator with zeros of a vector supported on $\lhd$ to one on $\Diamond$, and let $R_{\lhd}^{\Diamond}$ denotes its (formal) adjoint, being the restriction operator of a vector supported on $\Diamond$ to one on $\lhd$. Since the set $\Diamond$ will always be clear from the context, we will denote these operators simply by $I_{\lhd}$ and $R_{\lhd}$.

As sketched in the introduction of this section, the pieces are now in place to apply $R_{\breve{\mathbf{\Lambda}}}(\mathbf{A}_0 \otimes \mathbf{A}_1) I_{\mathbf{\Lambda}}$ in linear complexity.

Theorem 7.3.13. *Let $\breve{\mathbf{\Lambda}} \subset \breve{\vee}^0 \times \breve{\vee}^1$, $\mathbf{\Lambda} \subset \vee^0 \times \vee^1$ be finite double-trees. Then*

$$\mathbf{\Sigma} := \bigcup_{\lambda \in P_0\mathbf{\Lambda}} \Big(\{\lambda\} \times \bigcup_{\{\mu \in P_0\breve{\mathbf{\Lambda}} : |\mu| = |\lambda|+1,\ |\breve{S}^0(\mu) \cap S^0(\lambda)| > 0\}} \breve{\mathbf{\Lambda}}_{1,\mu} \Big),$$

$$\mathbf{\Theta} := \bigcup_{\lambda \in P_1\mathbf{\Lambda}} \Big(\{\mu \in P_0\breve{\mathbf{\Lambda}} : \exists \gamma \in \mathbf{\Lambda}_{0,\lambda} \text{ s.t. } |\gamma| = |\mu|,\ |\breve{S}^0(\mu) \cap S^0(\gamma)| > 0\} \times \{\lambda\} \Big),$$

are double-trees with $\#\mathbf{\Sigma} \lesssim \#\breve{\mathbf{\Lambda}}$ and $\#\mathbf{\Theta} \lesssim \#\mathbf{\Lambda}$, and

$$\begin{aligned} R_{\breve{\mathbf{\Lambda}}}(\mathbf{A}_0 \otimes \mathbf{A}_1) I_{\mathbf{\Lambda}} = {} & R_{\breve{\mathbf{\Lambda}}}(\mathbf{L}_0 \otimes \mathrm{Id}) I_{\mathbf{\Sigma}} R_{\mathbf{\Sigma}} (\mathrm{Id} \otimes \mathbf{A}_1) I_{\mathbf{\Lambda}} + \\ & R_{\breve{\mathbf{\Lambda}}}(\mathrm{Id} \otimes \mathbf{A}_1) I_{\mathbf{\Theta}} R_{\mathbf{\Theta}} (\mathbf{U}_0 \otimes \mathrm{Id}) I_{\mathbf{\Lambda}}. \end{aligned}$$

Proof. See Appendix 7.A. □

The application of $R_{\breve{\mathbf{\Lambda}}}(\mathbf{L}_0 \otimes \mathrm{Id}) I_{\mathbf{\Sigma}}$ boils down to the application of $R_{\breve{\mathbf{\Lambda}}_{0,\mu}} \mathbf{L}_0 I_{\mathbf{\Sigma}_{0,\mu}}$ for every $\mu \in P_1\mathbf{\Sigma} \cap P_1\breve{\mathbf{\Lambda}}$. Such an application can be performed in $\mathcal{O}(\#\breve{\mathbf{\Lambda}}_{0,\mu} + \#\mathbf{\Sigma}_{0,\mu})$ operations by means of a call of `evallow`(A_0); see also Algorithm 7.9.

Since $\sum_{\mu\in\breve{\vee}_1} \#\breve{\mathbf{\Lambda}}_{0,\mu} + \#\mathbf{\Sigma}_{0,\mu} = \#\breve{\mathbf{\Lambda}} + \#\mathbf{\Sigma}$, we conclude that the application of $R_{\breve{\mathbf{\Lambda}}}(\mathbf{L}_0 \otimes \mathrm{Id})I_{\mathbf{\Sigma}}$ can be performed in $\mathcal{O}(\#\breve{\mathbf{\Lambda}} + \#\mathbf{\Sigma})$ operations.

Similarly, applications of $R_{\mathbf{\Sigma}}(\mathrm{Id} \otimes \mathbf{A}_1)I_{\mathbf{\Lambda}}$, $R_{\breve{\mathbf{\Lambda}}}(\mathrm{Id} \otimes \mathbf{A}_1)I_{\mathbf{\Theta}}$, and $R_{\mathbf{\Theta}}(\mathbf{U}_0 \otimes \mathrm{Id})I_{\mathbf{\Lambda}}$ using calls of $\texttt{eval}(A_1)$, $\texttt{eval}(A_1)$, and $\texttt{evalupp}(A_0)$ respectively, can be done in $\mathcal{O}(\#\mathbf{\Sigma} + \#\mathbf{\Lambda})$, $\mathcal{O}(\#\breve{\mathbf{\Lambda}} + \#\mathbf{\Theta})$, and $\mathcal{O}(\#\mathbf{\Theta} + \#\mathbf{\Lambda})$ operations. From $\#\mathbf{\Sigma} \lesssim \#\breve{\mathbf{\Lambda}}$ and $\#\mathbf{\Theta} \lesssim \#\mathbf{\Lambda}$ we conclude the following.

Corollary 7.3.14. *Let $\breve{\mathbf{\Lambda}} \subset \breve{\vee}^0 \times \breve{\vee}^1$, $\mathbf{\Lambda} \subset \vee^0 \times \vee^1$ be finite double-trees, then $R_{\breve{\mathbf{\Lambda}}}(\mathbf{A}_0 \otimes \mathbf{A}_1)I_{\mathbf{\Lambda}}$ can be applied in $\mathcal{O}(\#\breve{\mathbf{\Lambda}} + \#\mathbf{\Lambda})$ operations.*

7.4 The heat equation and practical realization

In this section, we consider the numerical approximation of the *heat equation*

$$\left\{\begin{array}{rl} \frac{\mathrm{d}u}{\mathrm{d}t}(t) - (\Delta_{\mathbf{x}}u)(t) &= g(t) \quad (t \in I), \\ u(0) &= u_0. \end{array}\right. \tag{7.26}$$

For some bounded domain $\Omega \subset \mathbb{R}^2$, we take $H := L_2(\Omega)$ and $V := H^1_0(\Omega)$, so that $X = L_2(I; H^1_0(\Omega)) \cap H^1(I; H^{-1}(\Omega))$ and $Y = L_2(I; H^1_0(\Omega))$. We define

$$a(t; \eta, \zeta) := \int_\Omega \nabla\eta \cdot \nabla\zeta \,\mathrm{d}\mathbf{x},$$

and aim to solve the parabolic initial value problem (7.1) numerically. The bilinear forms present in our variational formulation (7.4) satisfy

$$A = M_t \otimes A_{\mathbf{x}}, \quad B = D_t \otimes M_{\mathbf{x}} + A, \quad \text{and} \quad \gamma_0'\gamma_0 = G_t \otimes M_{\mathbf{x}}$$

where

$$\begin{aligned} (M_t v)(w) &:= \int_I vw \,\mathrm{d}t, \;\; (D_t v)(w) := \int_I v'w \,\mathrm{d}t, \;\; (G_t v)(w) := v(0)w(0), \\ (A_{\mathbf{x}}\eta)(\zeta) &:= \int_\Omega \nabla\eta \cdot \nabla\zeta \,\mathrm{d}\mathbf{x}, \quad (M_{\mathbf{x}}\eta)(\zeta) := \int_\Omega \eta\zeta \,\mathrm{d}\mathbf{x}. \end{aligned} \tag{7.27}$$

In this section, we first construct suitable tensor-product bases for X and Y which functions are wavelets in time and hierarchical finite element functions in space. We then build our discrete 'trial' and 'test' spaces $(X^\delta, Y^\delta)_{\delta\in\Delta}$ as the span of subsets of these tensor-product bases. We finish with concrete uniformly optimal preconditioners K_X^δ and K_Y^δ, the basis necessary for error estimation in the adaptive loop, and evaluation of the right-hand side of (7.9) using interpolants.

7.4.1 Wavelets in time

We construct piecewise linear wavelet bases Σ for $H^1(I)$ and Ξ for $L_2(I)$.

Basis on the trial side

For Σ, we choose the three-point wavelet basis from [Ste98]; for completeness, we include its construction. For $\ell \geq 0$, define the scaling functions as the nodal continuous piecewise linears w.r.t. a uniform partition into 2^ℓ subintervals, that is $\Phi_\ell^\Sigma := \{\phi_{(\ell,n)} : 0 \leq n \leq 2^\ell\}$ with $\phi_{(\ell,n)}(k2^{-\ell}) = \delta_{kn}$ for $0 \leq k \leq 2^\ell$. Define $\Sigma_0 := \Phi_0^\Sigma$, and for $\ell \geq 1$, define $\Sigma_\ell := \{\sigma_\lambda : \lambda := (\ell, n) \text{ with } 0 \leq n < 2^{\ell-1}\}$ with $\sigma_\lambda = \sigma_{(\ell,n)}$ as in the right of Figure 7.3. Note that each σ_λ is a linear combination of three nodal functions from Φ_ℓ^Σ, hence the name *three-point wavelet*.

By imposing the parent-child structure

$$\tilde{\lambda} \lhd_\Sigma \lambda \iff |\tilde{\lambda}| + 1 = |\lambda| \text{ and } |\operatorname{supp}\sigma_\lambda \cap \operatorname{supp}\sigma_{\tilde{\lambda}}| > 0, \tag{7.28}$$

on any two indices $\tilde{\lambda}, \lambda$, we get the tree shown left in Figure 7.3.

Define $\Sigma := \cup_{\ell \geq 0}\Sigma_\ell$, $\vee_\Sigma := \{\lambda : \sigma_\lambda \in \Sigma\}$, and $S(\sigma_\lambda) := \operatorname{supp}\sigma_\lambda$. We see that Σ satisfies (7.20)–(7.21) and that the Φ_ℓ^Σ satisfy (7.22)–(7.25). Moreover, one can show that Σ is a Riesz basis for $L_2(I)$ (cf. [Ste98, Thm. 4.2]), and that $\{2^{-|\lambda|}\sigma_\lambda\}$ is a Riesz basis for $H^1(I)$ (cf. [Ste98, Thm. 4.3]).

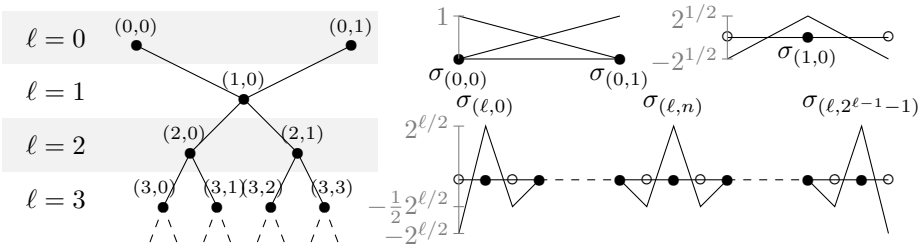


Figure 7.3. Left: three-point wavelet index set $\vee_\Sigma$ with parent-child relations; right: three-point wavelets.

Basis on the test side

We construct an $L_2(I)$-orthonormal basis Ξ.

For $\ell \geq 0$, define the (discontinuous) piecewise linear scaling functions w.r.t. a uniform partition into 2^ℓ subintervals by $\Phi_\ell^\Xi := \{\phi_{(\ell,n)} : 0 \leq n < 2^{\ell+1}\}$ where $\phi_{(0,0)}(t) := \mathbb{1}_{[0,1]}(t)$ and $\phi_{(0,1)}(t) := \sqrt{3}(2t-1)\mathbb{1}_{[0,1]}$, and for $\ell \geq 1$, $\phi_{(\ell,2k)}(t) := \phi_{(0,0)}(2^\ell t - k)$ and $\phi_{(\ell,2k+1)}(t) := \phi_{(0,1)}(2^\ell t - k)$. Let $\Xi_0 := \Phi_0^\Xi$, and define $\Xi_1 := \{\xi_{(1,0)}, \xi_{(1,1)}\}$ as in the right of Figure 7.4. For $\ell \geq 2$, we take $\Xi_\ell := \{\xi_{(\ell,n)} : 0 \leq n < 2^\ell\}$ with

$$\xi_{(\ell,2k)}(t) := 2^{(\ell-1)/2}\xi_{(1,0)}(2^{\ell-1}t - k), \quad \xi_{(\ell,2k+1)}(t) := 2^{(\ell-1)/2}\xi_{(1,1)}(2^{\ell-1}t - k).$$

The resulting $\Xi := \cup_{\ell\geq 0}\Xi_\ell$ is an *orthonormal* basis for $L_2(I)$, and together with its scaling functions $\cup_\ell \Phi_\ell^\Xi$, the conditions from §7.3.1 are satisfied with $S(\xi_\mu) := \operatorname{supp}\xi_\mu$. We impose a parent-child relation analogously to (7.28); see the left of Figure 7.4.

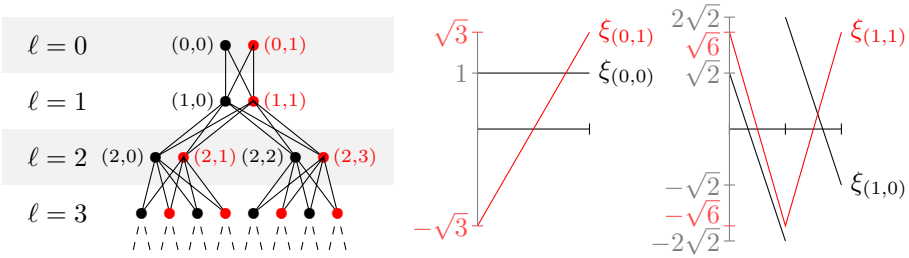


FIGURE 7.4. Left: orthonormal wavelet index set $\vee_\Xi$ with parent-child relations; right: the wavelets at levels 0 and 1.

7.4.2 Finite elements in space

Let $\mathbb{T}$ be the family of all *conforming* partitions of Ω into triangles that can be created by Newest Vertex Bisection from some given conforming initial triangulation $\mathcal{T}_\perp$ with an assignment of newest vertices satisfying the matching condition; cf. [Ste08b].

Define $\mathfrak{T} := \cup_{\mathcal{T}\in\mathbb{T}}\{T : T \in \mathcal{T}\}$. For $T \in \mathfrak{T}$, set $\operatorname{gen}(T)$ as the number of bisections needed to create T from its 'ancestor' $T' \in \mathcal{T}_\perp$. With $\mathfrak{N}$ the set of all vertices of all $T \in \mathfrak{T}$, for $\nu \in \mathfrak{N}$ we set $\operatorname{gen}(\nu) := \min\{\operatorname{gen}(T) : \nu \text{ is a vertex of } T \in \mathfrak{T}\}$.

Any $\nu \in \mathfrak{N}$ with $\operatorname{gen}(\nu) > 0$ is the midpoint of an edge e_ν of one or two $T \in \mathfrak{T}$ with $\operatorname{gen}(T) = \operatorname{gen}(\nu) - 1$. The set of newest vertices $\tilde{\nu}$ of these T, so those vertices of T with $|\tilde{\nu}| = \operatorname{gen}(\nu) - 1$, are defined as the parents of ν, denoted $\tilde{\nu} \lhd_{\mathfrak{N}} \nu$. The set of *godparents* of ν, denoted $\operatorname{gp}(\nu)$, are defined as the two endpoints of e_ν. Vertices with $\operatorname{gen}(\nu) = 0$ have no parents or godparents.

Example 7.4.1. In Figure 7.5, the parents of ν_4 are ν_1 and ν_3 and its godparents are ν_0, ν_2; the sole parent of ν_5 is ν_4, and its godparents are ν_0 and ν_3.

Proposition 7.4.2 ([DKS16]). *An (essentially) non-overlapping partition $\mathcal{T}$ of $\overline{\Omega}$ into triangles is in $\mathbb{T}$ if and only if the set $N_\mathcal{T}$ of vertices of all $T \in \mathcal{T}$ forms a* tree *in the sense of §7.3.1, meaning that it contains every vertex of generation zero as well as all parents of any $\nu \in N_\mathcal{T}$; see also Figure 7.5.*

Let $\mathcal{O}$ be the collection of spaces $W_\mathcal{T}$ of continuous piecewise linears w.r.t. $\mathcal{T} \in \mathbb{T}$ vanishing on $\partial\Omega$. For $\nu \in \mathfrak{N}$, we set ψ_ν as that continuous piecewise linear function on the *uniform partition* $\mathcal{T}_\nu := \{T \in \mathfrak{T} : \operatorname{gen}(T) = \operatorname{gen}(\nu)\} \in \mathbb{T}$

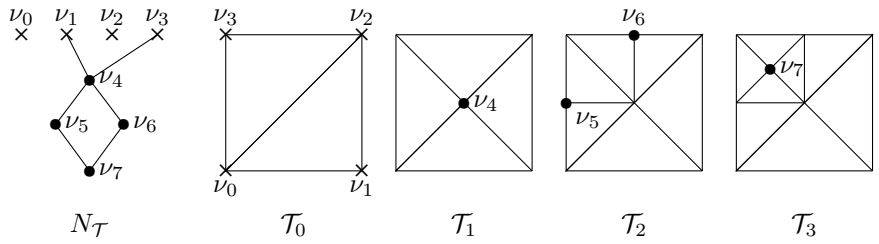


Figure 7.5. Vertex tree $N_{\mathcal{T}}$ and its triangulation $\mathcal{T}$ shown level-by-level.

for which $\psi_\nu(\tilde{\nu}) = \delta_{\nu\tilde{\nu}}$ for $\tilde{\nu} \in \mathcal{T}_\nu$. Setting $\mathfrak{N}_0 := \mathfrak{N} \setminus \partial\Omega$, the collection $\{\psi_\nu : \nu \in \mathfrak{N}_0\}$ is known as the *hierarchical basis*. For $\mathcal{T} \in \mathbb{T}$, write $N_{\mathcal{T},0} := N_{\mathcal{T}} \setminus \partial\Omega$ and $\Psi_{\mathcal{T}} := \{\psi_\nu : \nu \in N_{\mathcal{T},0}\}$; it holds that $W_{\mathcal{T}} = \operatorname{span}\Psi_{\mathcal{T}}$.

Applying stiffness matrices

The hierarchical basis satisfies conditions (7.20) and (7.21), and so, the application of stiffness matrices $(A\Psi_{\mathcal{T}})(\Psi_{\mathcal{T}})$ for $A \in \{A_{\mathbf{x}}, M_{\mathbf{x}}\}$ can be done through $\texttt{eval}(A)$.[1] However, the computation in Theorem 7.3.13 does not involve the lower and upper parts of A. This crucial insight allows for a faster and easier approach using standard finite element techniques: $\operatorname{span}\Psi_{\mathcal{T}}$ is a continuous piecewise linear finite element space, so it has a *canonical* single-scale basis $\Phi_{\mathcal{T}} := \operatorname{span}\{\phi_{\mathcal{T},\nu}\}$ characterized by $\phi_{\mathcal{T},\nu}(\tilde{\nu}) = \delta_{\nu\tilde{\nu}}$ for $\tilde{\nu} \in N_{\mathcal{T},0}$, for which the application of $(A\Phi_{\mathcal{T}})(\Phi_{\mathcal{T}})$ at linear cost using local element matrices is standard. This is different from the general setting in §7.3.1, in that $\dim\Phi_{\mathcal{T}} = \dim\Psi_{\mathcal{T}}$ also for locally refined triangulations. Let T be the transformation characterized by $\Psi_{\mathcal{T}} = T^\top\Phi_{\mathcal{T}}$, we find

$$(A\Psi_{\mathcal{T}})(\Psi_{\mathcal{T}}) = T^\top(A\Phi_{\mathcal{T}})(\Phi_{\mathcal{T}})T. \tag{7.29}$$

We can apply T in linear complexity by iterating over the vertices bottom-up while applying elementary local transformations in which not parent-child, but *godparent*-child relations play a role.

7.4.3 Inf-sup stable family of trial- and test spaces

With Σ and Ξ from §7.4.1 and $\Psi_{\mathfrak{N}_0} := \{\psi_\nu : \nu \in \mathfrak{N}_0\}$ from §7.4.2, we find that $X = \operatorname{span}(\Sigma \otimes \Psi_{\mathfrak{N}_0})$ and $Y = \operatorname{span}(\Xi \otimes \Psi_{\mathfrak{N}_0})$. We now turn to the construction of X^δ and Y^δ.

[1] This would require the definition of a suitable single-scale basis.

Definition 7.4.3. For a double-tree $\mathbf{\Lambda}^\delta \subset \vee_\Sigma \times \mathfrak{N}$, define $\mathbf{\Lambda}_0^\delta := \mathbf{\Lambda}^\delta \setminus \vee_\Sigma \times \partial\Omega$. We construct our 'trial' space as

$$X^\delta := \operatorname{span}\{\sigma_\lambda \otimes \psi_\nu : (\lambda, \nu) \in \mathbf{\Lambda}_0^\delta\}.$$

Defining the double-tree $\mathbf{\Lambda}_{Y,0}^\delta \subset \vee_\Xi \times \mathfrak{N}_0$ as

$$\mathbf{\Lambda}_{Y,0}^\delta := \{(\mu, \nu) : \exists (\lambda, \nu) \in \mathbf{\Lambda}_0^\delta,\ \mu \in \vee_\Xi,\ |\mu| = |\lambda|,\ |\operatorname{supp} \xi_\mu \cap \operatorname{supp} \sigma_\lambda| > 0\},$$

we construct our 'test' space as $Y^\delta = Y^\delta(X^\delta) := \operatorname{span}\{\xi_\mu \otimes \psi_\nu : (\mu, \nu) \in \mathbf{\Lambda}_{Y,0}^\delta\}$.

Theorem 7.4.4. *Define $\Delta := \{\delta : \mathbf{\Lambda}^\delta \subset \vee_\Sigma \times \mathfrak{N}$ is a double-tree$\}$ equipped with the partial ordering $\delta \preceq \tilde{\delta} \iff \mathbf{\Lambda}^\delta \subseteq \mathbf{\Lambda}^{\tilde{\delta}}$. With X^δ and Y^δ as above, uniform inf-sup stability holds; cf.* (7.6).

Proof. See Proposition 6.5.2. □

Definition 7.4.5. Given a double-tree $\mathbf{\Lambda}^\delta \subset \vee_\Sigma \times \mathfrak{N}$, we define $\mathbf{\Lambda}^{\underline{\delta}} \supset \mathbf{\Lambda}^\delta$ by adding, for $(\lambda, \nu) \in \mathbf{\Lambda}^\delta$ and any child $\tilde{\lambda}$ of λ and descendant $\tilde{\nu}$ of ν up to generation 2, all pairs $(\tilde{\lambda}, \nu)$ and $(\lambda, \tilde{\nu})$. We expect this choice of $X^{\underline{\delta}}$ to provide saturation; cf. (7.8).

7.4.4 Preconditioners

We follow §6.5.6 for the construction of optimal preconditioners K_Y^δ for ${E_Y^\delta}' A E_Y^\delta$ and K_X^δ for $S^{\underline{\delta}\delta}$ necessary for solving (7.9). With notation from Definition 7.3.12, we equip X^δ and Y^δ with bases

$$\begin{cases} \bigcup\limits_{\lambda \in P_0 \mathbf{\Lambda}_0^\delta} \sigma_\lambda \otimes \Psi_\lambda^\delta \quad \text{with} \quad \Psi_\lambda^\delta := \{\psi_\nu : \nu \in (\mathbf{\Lambda}_0^\delta)_{1,\lambda}\}, \\ \bigcup\limits_{\mu \in P_0 \mathbf{\Lambda}_{Y,0}^\delta} \xi_\mu \otimes \Psi_\mu^\delta \quad \text{with} \quad \Psi_\mu^\delta := \{\psi_\nu : \nu \in (\mathbf{\Lambda}_{Y,0}^\delta)_{1,\mu}\}. \end{cases}$$

Matrix representations of preconditioners from §6.5.6 are then given by

$$\begin{cases} \mathbf{K}_Y^\delta := \operatorname{blockdiag}[\mathbf{K}_\mu^\delta]_{\mu \in P_0 \mathbf{\Lambda}_{Y,0}^\delta} \quad \text{where} \quad \mathbf{K}_\mu^\delta \eqsim (\mathbf{A}_\mu^\delta)^{-1}, \\ \mathbf{K}_X^\delta := \operatorname{blockdiag}[\mathbf{K}_\lambda^\delta \mathbf{A}_\lambda^\delta \mathbf{K}_\lambda^\delta]_{\lambda \in P_0 \mathbf{\Lambda}_0^\delta} \quad \text{where} \quad \mathbf{K}_\lambda^\delta \eqsim (\mathbf{A}_\lambda^\delta + 2^{|\lambda|} \mathbf{M}_\lambda^\delta)^{-1} \end{cases}$$

with $\mathbf{A}_\mu^\delta := (A_{\mathbf{x}} \Psi_\mu^\delta)(\Psi_\mu^\delta)$, $\mathbf{A}_\lambda^\delta := (A_{\mathbf{x}} \Psi_\lambda^\delta)(\Psi_\lambda^\delta)$, and $\mathbf{M}_\lambda^\delta := (M_{\mathbf{x}} \Psi_\lambda^\delta)(\Psi_\lambda^\delta)$. Suitable spatial preconditioners $\mathbf{K}_\mu^\delta$ are provided by multigrid methods. In [OR00] it was shown that for quasi-uniform triangulations, satisfying a 'full-regularity' assumption, a multiplicative multigrid method yields suitable $\mathbf{K}_\lambda^\delta$, and we assume these results to hold for our locally refined triangulations $\mathcal{T} \in \mathbb{T}$ as well. In §7.5.1 below, we detail our linear-complexity multigrid implementation following [WZ17].

7.4.5 Right-hand side

We follow §6.6.4. For $g \in C(\overline{I \times \Omega})$, $u_0 \in C(\overline{\Omega})$, we can approximate the right-hand side of (7.9) by interpolants, avoiding quadrature issues.

The procedure of §7.4.2 for constructing the hierarchical basis $\Psi_{\mathfrak{N}} := \{\psi_\nu : \nu \in \mathfrak{N}\}$ can be applied in time as well, yielding the basis $\{\psi_\lambda : \lambda \in \vee_{\mathfrak{I}}\}$ from Figure 7.1 which index set $\vee_{\mathfrak{I}}$ coincides with $\vee_\Sigma$. We construct $\{\tilde{\psi}_\nu : \nu \in \mathfrak{N}\} \subset C(\overline{\Omega})'$ biorthogonal to $\Psi_{\mathfrak{N}}$, with $\tilde{\psi}_\nu := \delta_\nu - \sum_{\tilde{\nu} \in \mathrm{gp}(\nu)} \delta_{\tilde{\nu}}/2$. In time, define $\{\tilde{\psi}_\lambda : \lambda \in \vee_{\mathfrak{I}}\} \subset C(\overline{I})'$ analogously. Define the vectors $\mathbf{g} := [(\tilde{\psi}_\lambda \otimes \tilde{\psi}_\nu)(g)]_{(\lambda,\nu)\in\mathbf{\Lambda}^\delta}$ and $\mathbf{u}_0 := [\tilde{\psi}_\nu(u_0)]_{\nu \in P_1\mathbf{\Lambda}^\delta}$. Upon replacing (g, u_0) in (7.9) by the interpolants

$$ {}^\delta g := \sum_{(\lambda,\nu)\in\mathbf{\Lambda}^\delta} \mathbf{g}_{(\lambda,\nu)} \psi_\lambda \otimes \psi_\nu, \quad {}^\delta u_0 := \sum_{\nu \in P_1\mathbf{\Lambda}^\delta} \mathbf{u}_{0,\nu} \psi_\nu, $$

we can evaluate its right-hand side in linear complexity by computing the quantities

$$ \begin{aligned} [\langle \xi_\mu \otimes \psi_\nu, {}^\delta g\rangle_{L_2(I\times\Omega)}]_{(\mu,\nu)\in\mathbf{\Lambda}^\delta_{Y,0}} &= R_{\mathbf{\Lambda}^\delta_{Y,0}} (M_t \otimes M_{\mathbf{x}}) I_{\mathbf{\Lambda}^\delta} \mathbf{g}, \\ [\sigma_\lambda(0)\langle \psi_\nu, {}^\delta u_0\rangle_{L_2(\Omega)}]_{(\lambda,\nu)\in\mathbf{\Lambda}^\delta_0} &= [\sigma_\lambda(0)\mathbf{w}_\nu]_{(\lambda,\nu)\in\mathbf{\Lambda}^\delta_0} \\ \text{where} \quad \mathbf{w} &= (M_{\mathbf{x}} \Psi_{\mathfrak{N}}|_{P_1\mathbf{\Lambda}^\delta})(\Psi_{\mathfrak{N}}|_{P_1\mathbf{\Lambda}^\delta})\mathbf{u}_0. \end{aligned} $$

7.4.6 Two-level basis

We now discuss the construction of a uniformly X-stable basis Θ_δ, needed in the local error estimator $\mathbf{r}^\delta$ of (7.12). Following §6.6.3, define a *modified hierarchical basis* $\{\hat{\psi}_\nu : \nu \in \mathfrak{N}_0\}$ by

$$ \hat{\psi}_\nu = \psi_\nu \ \text{ when } \ \mathrm{gen}(\nu) = 0, \quad \text{else} \quad \hat{\psi}_\nu := \psi_\nu - \frac{\sum_{\{\tilde{\nu}\in\mathfrak{N}:\tilde{\nu}\lhd_{\mathfrak{N}}\nu\}} \frac{\int_\Omega \psi_\nu \,\mathrm{d}\mathbf{x}}{\int_\Omega \psi_{\tilde{\nu}} \,\mathrm{d}\mathbf{x}} \psi_{\tilde{\nu}}}{\#\{\tilde{\nu} \in \mathfrak{N} : \tilde{\nu} \lhd_{\mathfrak{N}} \nu\}}. $$

For any $\mathcal{T} \in \mathbb{T}$, $W_\mathcal{T} = \mathrm{span}\{\hat{\psi}_\nu : \nu \in N_{\mathcal{T},0}\} = \mathrm{span}\,\Psi_\mathcal{T}$ and the transformation from modified to unmodified hierarchical basis can be performed in linear complexity. For $\underline{\mathcal{T}} \succeq \mathcal{T} \in \mathbb{T}$, $\mathbf{d} \in \ell_2(N_{\underline{\mathcal{T}},0} \setminus N_{\mathcal{T},0})$ and $v \in W_\mathcal{T}$, Lemma 6.6.7 shows that

$$ \text{(7.30)} \qquad \begin{cases} \|v + \sum_\nu \mathbf{d}_\nu \hat{\psi}_\nu\|^2_{H^1(\Omega)} \eqsim \|v\|^2_{H^1(\Omega)} + \|\mathbf{d}\|^2, \\ \|v + \sum_\nu \mathbf{d}_\nu \hat{\psi}_\nu\|^2_{H^{-1}(\Omega)} \eqsim \|v\|^2_{H^{-1}(\Omega)} + \sum_\nu 4^{-\mathrm{gen}(\nu)} |\mathbf{d}_\nu|^2, \end{cases} $$

with the constants in the $\eqsim$-symbols dependent on $\max_{\{\underline{\mathcal{T}}\ni\underline{T}\subset T\in\mathcal{T}\}} \{\mathrm{gen}(\underline{T}) - \mathrm{gen}(T)\}$ only. We then construct a basis for $X^{\underline{\delta}} \ominus X^\delta$ as

$$ \Theta_\delta := \{e_{\lambda\nu}\sigma_\lambda \otimes \hat{\psi}_\nu : (\lambda,\nu) \in \mathbf{\Lambda}^\delta_0 \setminus \mathbf{\Lambda}^\delta_0\} \quad \text{where} \quad \frac{1}{e_{\lambda\nu}} = \sqrt{1 + 4^{|\lambda| - \mathrm{gen}(\nu)}}. $$

Define the *gradedness* of a double-tree $\mathbf{\Lambda}^\delta \subset \vee_\Sigma \times \mathfrak{N}$ as the smallest $L_\delta \in \mathbb{N}$ for which every $(\lambda, \nu) \in \mathbf{\Lambda}^\delta$ with $\tilde{\nu}$ an ancestor of ν with $\operatorname{gen}(\nu) - \operatorname{gen}(\tilde{\nu}) = L_\delta$, it holds that $(\breve{\lambda}, \tilde{\nu}) \in \mathbf{\Lambda}^\delta$ for all $\breve{\lambda} \lhd_\Sigma \lambda$. Thanks to Σ being a (scaled) Riesz basis for $L_2(I)$ and $H^1(I)$, together with the $H^1(\Omega)$- and $H^{-1}(\Omega)$-stable splittings of (7.30), it holds that

$$\|z + \mathbf{c}^\top \Theta_\delta\|_X^2 \eqsim \|z\|_X^2 + \|\mathbf{c}\|^2 \quad (\mathbf{c} \in \ell_2(\mathbf{\Lambda}_0^\delta \setminus \mathbf{\Lambda}_0^\delta), z \in X^\delta),$$

with the constant in the $\eqsim$-symbol dependent on L_δ only, so when L_δ is uniformly bounded, condition (7.11) is satisfied.

7.5 Implementation

A tree-based implementation of the aforementioned adaptive algorithm in C++ can be found at [vVW21d]. In this section, we describe our design choices for a linear complexity implementation.

7.5.1 Trees and linear operators in one axis

In §7.3, we consider an abstract multilevel collection Ψ indexed on $\vee_\Psi$. Endowed with a parent-child relation, $\vee_\Psi$ has a tree-like structure that we call a *mother tree*; see also Figures 7.3 and 7.4.

In our applications, the support of a wavelet ψ_λ is a union of simplices of generation $|\lambda|$. In time, these simplices are subintervals of I found by dyadic refinement. In space, they are elements of $\mathfrak{T}$, the collection of all triangles found by newest vertex bisection. Endowed with the natural parent-child relation, both collections of simplices have a tree structure we call the *domain mother tree*. Every wavelet ψ_λ stores references to the simplices T of generation $|\lambda|$ that make up its support; conversely, every T stores a reference to ψ_λ.

Every mother tree $\vee$ is stored once in memory, and every node $\lambda \in \vee$ stores references to its parents, children, and siblings. We treat the mother tree as infinite by *lazy initialization*, constructing new nodes as they are needed.

Trees

We store a tree $\Lambda \subset \vee$ using the parent-child relation, and additionally, at each $\lambda \in \Lambda$ store a reference to the corresponding node in $\vee$. This allows us to compare different trees subject to the same mother tree. This tree-like representation does not allow direct access of arbitrary nodes: in any operation, we traverse Λ from its roots in breadth-first, or level-wise, order.

Tree operations

One important operation is the `union` of one tree Λ into another $\breve{\Lambda}$. This can be implemented by traversing both trees simultaneously in breadth-first order.

The union allows us to easily perform high-level operations, such as vector addition: given two vectors $\mathbf{c} \in \ell_2(\Lambda)$, $\mathbf{d} \in \ell_2(\breve{\Lambda})$ on the same mother tree $\vee$, we use the union to perform $\mathbf{c} := \mathbf{c} + \mathbf{d}$. See Figure 7.6 for an example.

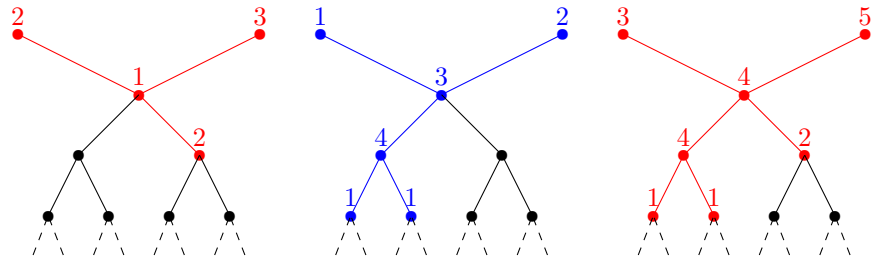


Figure 7.6. Left: $\mathbf{c} \in \ell_2(\Lambda)$ for $\Lambda \subset \vee_{\mathfrak{I}}$; Middle: $\mathbf{d} \in \ell_2(\breve{\Lambda})$ for $\breve{\Lambda} \subset \vee_{\mathfrak{I}}$; Right: in-place sum $\mathbf{c} := \mathbf{c} + \mathbf{d}$.

Tree operations in time

The routines `eval`, `evalupp`, and `evallow` from §7.3.1 involve various level-wise index sets (represented as arrays of references into their mother trees). One example is $\breve{\Pi}_B = \{\lambda \in \breve{\Pi} : \left|\operatorname{supp} \breve{\phi}_\lambda \cap \cup_{\mu\in\Lambda_\ell} S(\mu)\right| > 0\}$, which we constructed efficiently using the domain mother tree; see Algorithm 7.5.

Algorithm 7.5: The construction of $\breve{\Pi}_B$.

Data: $\ell \in \mathbb{N}$, $\breve{\Pi} \subset \breve{\Delta}_{\ell-1}$, $\Lambda_\ell \subset \vee_\ell$.
Result: $[\breve{\Pi}_A, \breve{\Pi}_B]$ where $\breve{\Pi}_A = \breve{\Pi} \setminus \breve{\Pi}_B$,
$\breve{\Pi}_B = \{\lambda \in \breve{\Pi} : \left|\operatorname{supp} \breve{\phi}_\lambda \cap \cup_{\mu\in\Lambda_\ell} S(\mu)\right| > 0\}$.

```
Π̆_A := ∅;
Π̆_B := ∅;
for μ ∈ Λ_ℓ do
    for T ∈ ψ_μ.support do                    // We have S(ψ_μ) = supp ψ_μ.
        T.parent.marked := true;
for λ ∈ Π̆ do
    if ∃T ∈ φ̆_λ.support with T.marked = true then
        Π̆_B.insert(λ);
    else
        Π̆_A.insert(λ);
for μ ∈ Λ_ℓ do
    for T ∈ ψ_μ.support do
        T.parent.marked := false;
```

We can apply the linear operators appearing in the routines of §7.3.1 efficiently by again traversing the domain mother tree; for example, Algorithm 7.6 details a matrix-free application of $(A\Phi|_\Pi)(\breve{\Phi}|_{\breve{\Pi}})$.

Algorithm 7.6: The computation of $\mathbf{e} = (A\Phi|_\Pi)(\breve{\Phi}|_{\breve{\Pi}})\mathbf{d}$.

Data: Index sets $\Pi \subset \Delta_\ell, \breve{\Pi} \subset \breve{\Delta}_\ell$, $\mathbf{d} \in \ell_2(\Pi)$, local and linear $A\colon \operatorname{span}\Phi \to \operatorname{span}\breve{\Phi}'$.
Result: $\mathbf{e} = (A\Phi|_\Pi)(\breve{\Phi}|_{\breve{\Pi}})\mathbf{d}$
for $\lambda \in \Pi$ **do** $\phi_\lambda.\mathtt{data} := \mathbf{d}_\lambda$;
for $\mu \in \breve{\Pi}$ **do**
  $\mathbf{e}_\lambda := 0$;
  for $T \in \breve{\phi}_\mu.\mathtt{support}$ **do**
    for $\phi_\lambda \in T.\mathtt{functions}(\Delta_\ell)$ **do**
      // $\{\phi_\lambda : \lambda \in \Delta_\ell, |\operatorname{supp}\phi_\lambda \cap T| > 0\}$
      $\mathbf{e}_\lambda := \mathbf{e}_\lambda + A(\phi_\lambda)(\breve{\phi}_\mu|_T)\cdot\phi_\lambda.\mathtt{data}$;
for $\lambda \in \Pi$ **do** $\phi_\lambda.\mathtt{data} := 0$;

Operations in space

We can construct a triangulation $\mathcal{T}$ from a vertex tree $N_\mathcal{T}$ in linear complexity. First mark every $\nu \in N_\mathcal{T}$ in its mother tree, then traverse the domain mother tree $\mathfrak{T}$. A triangle T visited in this traversal is in $\mathcal{T}$ exactly when the newest vertex of its children is not marked.

For the preconditioners $\mathbf{K}_\mu^\delta$ and $\mathbf{K}_\lambda^\delta$ from §7.4.4 we use multigrid. We apply multiplicative V-cycle multigrid, in each cycle applying one pre- and one post Gauss-Seidel smoother with reversed ordering of the unknowns.

In view of obtaining a linear complexity algorithm, at level k we restrict smoothing to the vertices of generation k as well as their godparents, cf. [WZ17]. For $\mathcal{T} \in \mathbb{T}$ we consider $W_\mathcal{T}$, the space of continuous piecewise linears w.r.t. $\mathcal{T}$, zero on $\partial\Omega$, now equipped with the single-scale basis $\Phi_\mathcal{T}$. Set $L = L(\mathcal{T}) := \max_{T\in\mathcal{T}} \operatorname{gen}(T)$, and define the sequence

$$\mathcal{T}_\perp = \mathcal{T}_0 \prec \mathcal{T}_1 \prec \cdots \prec \mathcal{T}_L = \mathcal{T} \subset \mathbb{T}$$

where $\mathcal{T}_{k-1}$ is constructed from $\mathcal{T}_k$ by removing all vertices $\nu \in N_{\mathcal{T}_k}$ for which $\operatorname{gen}(\nu) = k$. For $1 \le k \le L$, let M_k be the set of new vertices and their godparents, i.e., $M_k := \bigcup_{\nu\in N_{\mathcal{T}_k}\setminus N_{\mathcal{T}_{k-1}}} \{\nu\} \cup \operatorname{gp}(\nu)$, and let $M_{k,0} := M_k \setminus \partial\Omega$ be the vertices not on the boundary. We consider the multilevel decomposition, cf. [WZ17],

$$W_{\mathcal{T}_L} = W_{\mathcal{T}_0} + \sum_{k=1}^{L} \sum_{\nu\in M_{k,0}} \operatorname{span}\phi_{k,\nu}, \quad \text{where} \quad \phi_{k,\nu} := \phi_{\mathcal{T}_k,\nu}. \tag{7.31}$$

For $1 \leq k \leq L$, let $\mathbf{P}_k$ be the *prolongation matrix*, i.e., the matrix representation of the embedding $W_{\mathcal{T}_{k-1}} \to W_{\mathcal{T}_k}$, and enumerate the vertices $M_{k,0}$ as $(\nu_k^i)_{i=1}^{n_k}$. Algorithm 7.7 details a (non-recursive) implementation of a single multiplicative V-cycle for the multilevel decomposition (7.31) using Gauss-Seidel smoothing. We assume the availability of an efficient coarse-grid solver; in our application, a direct solve suffices. For linear complexity, we use in-place vector updates restricted to non-zeros.

Note that this multigrid method is given in terms of the single-scale basis $\Phi_{\mathcal{T}}$; it can be transformed to the hierarchical basis $\Psi_{\mathcal{T}}$ similarly to (7.29). Multiple V-cycles are done by setting $u_0 := 0$ and iterating $u_k := \mathrm{MG}(A, f - Au_{k-1})$.

Algorithm 7.7: Single multiplicative V-cycle multigrid $\mathrm{MG}(A, f)$.

Data: Some $f \in W'_{\mathcal{T}}$ and a linear operator $A\colon W_{\mathcal{T}} \to W'_{\mathcal{T}}$.
Result: $u = \mathbf{u}^\top \Phi_{\mathcal{T}} \in W_{\mathcal{T}}$, the result of a single V-cycle applied to f.
$\mathbf{r} := f(\Phi_{\mathcal{T}})$;
for $L \geq k \geq 1$ **do**
  for $\nu = \nu_k^1, \dots, \nu_k^{n_k}$ **do**
    $r_{k,\nu} := \mathbf{r}_\nu$;
    $e_{k,\nu} := r_{k,\nu}/(A\phi_{k,\nu})(\phi_{k,\nu})$;
    $\mathbf{r} := \mathbf{r} - e_{k,\nu}(A\phi_{k,\nu})(\Phi_{\mathcal{T}_k})$;
  $\mathbf{r} := \mathbf{P}_k^\top \mathbf{r}$;
Solve $(A\Phi_{\mathcal{T}_0})(\Phi_{\mathcal{T}_0})\mathbf{u} = \mathbf{r}$;

for $1 \leq k \leq L$ **do**
  $\mathbf{u} := \mathbf{P}_k \mathbf{u}$;
  for $\nu = \nu_k^{n_k}, \dots, \nu_k^1$ **do**
    $\mathbf{u}_\nu := \mathbf{u}_\nu + e_{k,\nu}$;
    $\mathbf{u}_\nu := \mathbf{u}_\nu + (r_{k,\nu} - (A\phi_{k,\nu})(\mathbf{u}^\top \Phi_{\mathcal{T}_k}))/(A\phi_{k,\nu})(\phi_{k,\nu})$;

7.5.2 Double-trees and tensor-product operators

For every node in a double-tree $\mathbf{\Lambda} \subset \vee^0 \times \vee^1$, we store a reference to the underlying pair of nodes in their mother trees. This allows growing double-trees intuitively, and allows comparing different double-trees over the same pair of mother trees. C++ templates allow us to re-use much of the tree code without runtime performance loss.

In §7.3.2 we saw how to apply a tensor-product operator. For this, we first construct the double-trees $\mathbf{\Sigma}$ and $\mathbf{\Theta}$; construction of $\mathbf{\Sigma}$ is illustrated in Algorithm 7.8. Evaluation of the operator then reduces to the four simple steps of Algorithm 7.9.

Algorithm 7.8: Function `GenerateSigma`($\breve{\boldsymbol{\Lambda}}, \boldsymbol{\Lambda}$).

Data: $\breve{\boldsymbol{\Lambda}} \subset \breve{\vee}^0 \times \breve{\vee}^1$, $\boldsymbol{\Lambda} \subset \vee^0 \times \vee^1$
Result: $\boldsymbol{\Sigma}$ for application of Theorem 7.3.13 with $\breve{\boldsymbol{\Lambda}}$ and $\boldsymbol{\Lambda}$.
$\boldsymbol{\Sigma} := P_0\boldsymbol{\Lambda} \times \{\nu \in P_1\breve{\boldsymbol{\Lambda}} : |\nu| = 0\}$;
for $\lambda \in \boldsymbol{\Sigma}$.`project`$(0)$ **do**
  for $T \in \phi_\lambda$.`support` **do**
    for $\mu \in T$.`functions`$(\breve{\vee}^0_{|\lambda|})$ **do**
      $\boldsymbol{\Sigma}$.`fiber`$(1, \lambda)$.`union`($\breve{\boldsymbol{\Lambda}}$.`fiber`$(1, \mu)$);

Algorithm 7.9: Algorithm to evaluate $\mathbf{d} = R_{\breve{\boldsymbol{\Lambda}}}(\boldsymbol{A}_0 \otimes \boldsymbol{A}_1) I_{\boldsymbol{\Lambda}} \mathbf{c}$.

Data: $\boldsymbol{\Lambda} \subset \vee^0 \times \vee^1$, $\breve{\boldsymbol{\Lambda}} \subset \breve{\vee}^0 \times \breve{\vee}^1$, $\mathbf{c} \in \ell_2(\boldsymbol{\Lambda})$, $\mathbf{d} \in \ell_2(\breve{\boldsymbol{\Lambda}})$.
$\boldsymbol{\Sigma} :=$ `GenerateSigma`($\breve{\boldsymbol{\Lambda}}, \boldsymbol{\Lambda}$);
$\boldsymbol{\Theta} :=$ `GenerateTheta`($\breve{\boldsymbol{\Lambda}}, \boldsymbol{\Lambda}$);
$\mathbf{s} := \mathbf{0} \in \ell_2(\boldsymbol{\Sigma})$;
$\mathbf{t} := \mathbf{0} \in \ell_2(\boldsymbol{\Theta})$;
$\mathbf{l} := \mathbf{0} \in \ell_2(\breve{\boldsymbol{\Lambda}})$;

for $\lambda \in \mathbf{s}$.`project`(0) **do** `eval`$(A_1)(\mathbf{s}$.`fiber`$(1, \lambda), \mathbf{c}$.`fiber`$(1, \lambda))$;
for $\mu \in \mathbf{l}$.`project`(1) **do** `evallow`$(A_0)(\mathbf{l}$.`fiber`$(0, \mu), \mathbf{s}$.`fiber`$(0, \mu))$;
for $\mu \in \mathbf{t}$.`project`(1) **do** `evalupp`$(A_0)(\mathbf{t}$.`fiber`$(0, \mu), \mathbf{c}$.`fiber`$(0, \mu))$;
for $\lambda \in \mathbf{d}$.`project`(0) **do** `eval`$(A_1)(\mathbf{d}$.`fiber`$(1, \lambda), \mathbf{t}$.`fiber`$(1, \lambda))$;

$\mathbf{d} := \mathbf{d} + \mathbf{l}$;

Memory optimizations

As the memory consumption of a double-tree is significant, at around 280 bytes per node, we want to have as few double-trees in memory as possible. By storing the nodes of $\boldsymbol{\Lambda}$ in a persistent container, every node is uniquely identified with its index in the container. This induces a mapping $\mathbb{R}^{\#\boldsymbol{\Lambda}} \leftrightarrow \ell_2(\boldsymbol{\Lambda})$ and allows us to overlay multiple vectors on the same underlying double-tree in a memory-friendly way.

The $\boldsymbol{\Sigma}$ generated by Algorithm 7.8 for the application of a tensor-product operator can play the role of $\boldsymbol{\Theta}$ necessary for the application of its transpose operator (and vice versa). This allows tensor-product operators and their transposes to share the double-trees $\boldsymbol{\Sigma}$ and $\boldsymbol{\Theta}$.

With these insights, our implementation of the heat equation has at most 5 different double-trees in memory.

7.5.3 The adaptive loop

In the refine step of the adaptive loop, we first mark a set J of nodes in $\boldsymbol{\Lambda}^{\underline{\delta}} \setminus \boldsymbol{\Lambda}^{\delta}$ using Dörfler marking (possible in linear complexity; cf. [PP20]). We then refine $\boldsymbol{\Lambda}^{\delta}$ to the smallest double-tree containing J:

1. mark all nodes in $\boldsymbol{\Lambda}^{\underline{\delta}}$ that are also present in $\boldsymbol{\Lambda}^{\delta}$ ((ii) in Fig. 7.7);
2. traverse $\boldsymbol{\Lambda}^{\underline{\delta}}$ from every node in J, top-down in level-wise order, until hitting a previously marked node. Mark all nodes along the way ((iii–iv) in Fig. 7.7);
3. `union` the marked nodes of $\boldsymbol{\Lambda}^{\underline{\delta}}$ into $\boldsymbol{\Lambda}^{\delta}$ ((v) in Fig. 7.7).

As $\#\boldsymbol{\Lambda}^{\underline{\delta}} \lesssim \#\boldsymbol{\Lambda}^{\delta}$ and we visit every node of $\boldsymbol{\Lambda}^{\underline{\delta}}$ at most twice, the traversal is linear in $\#\boldsymbol{\Lambda}^{\delta}$. See also Figure 7.7.

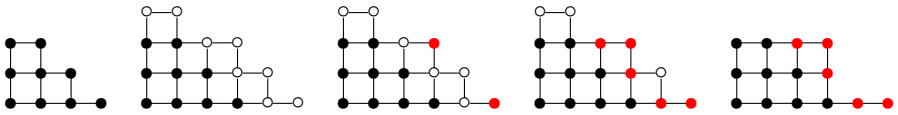


Figure 7.7. Adaptive refinement of a double-tree with underlying *unary* mother trees. Left to right: (i) $\boldsymbol{\Lambda}^{\delta}$; (ii) $\boldsymbol{\Lambda}^{\underline{\delta}}$ with nodes in $\boldsymbol{\Lambda}^{\underline{\delta}} \setminus \boldsymbol{\Lambda}^{\delta}$ in white; (iii) nodes in J marked in red; (iv) nodes marked in the top-down traversal; (v) refined $\boldsymbol{\Lambda}^{\delta}$.

7.6 Numerical experiments

We consider the heat equation (7.26), and assess our implementation of the adaptive Algorithm 7.1 for its numerical solution. Complementing the convergence results gathered in §6.7, here we provide results on the practical performance of the adaptive loop. Results were gathered on a multi-core 2.2 GHz machine, provided by the Dutch national e-infrastructure with the support of SURF Cooperative.

7.6.1 The adaptive loop

We summarize the main results from §6.7. We run Algorithm 7.1 with $\theta = \frac{1}{2}$ and $\xi = \frac{1}{2}$. We consider four problems.

In the *smooth problem*, we select $\Omega := [0,1]^2$ and prescribe the solution

$$u(t,x,y) := (1+t^2)x(1-x)y(1-y).$$

In the *moving peak* problem, we again select $\Omega := [0,1]^2$ with prescribed solution

$$u(t,x,y) := x(1-x)y(1-y)\exp(-100[(x-t)^2+(y-t)^2]);$$

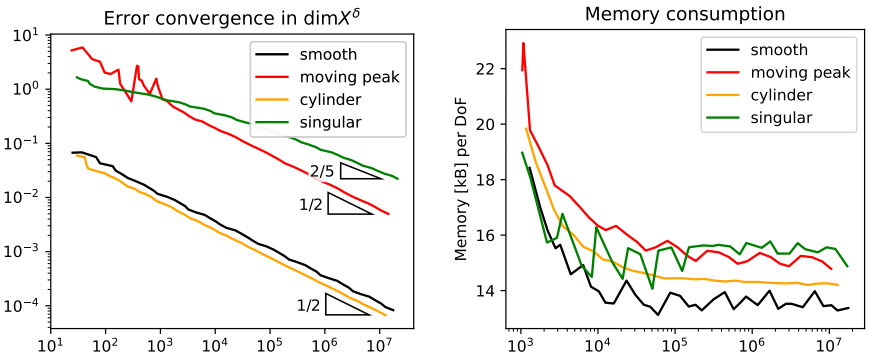


FIGURE 7.8. Error convergence and peak memory usage of the adaptive loop for the four problems of §7.6.1.

here, u is essentially zero outside a small strip along the diagonal $(0,0,0)$ to $(1,1,1)$.

In the *cylinder* problem, we select $\Omega := [-1,1]^2 \setminus [-1,0]^2$ with data

$$u_0 \equiv 0, \quad \text{and} \quad g(t,x,y) := t \cdot \mathbb{1}_{\{x^2+y^2<1/4\}}.$$

The solution has singularities in the re-entrant corner and along the wall of the cylinder $\{(t,x,y) : x^2+y^2 = 1/4\}$.

In the *singular* problem, we select $\Omega := [-1,1]^2 \setminus [-1,0]^2$ with data $u_0 \equiv 1$ and $g \equiv 0$; the solution then has singularities along $\{0\} \times \partial\Omega$ and $I \times \{(0,0)\}$.

Convergence

To estimate the error $\|u - \hat{u}^\delta\|_X$, we measure the residual error estimator $\|\mathbf{r}^\delta(\hat{u}^\delta)\|$ from (7.12); see also Lemma 7.2.3. In the left pane of Figure 7.8, for the first three problems, we observe a convergence rate of $1/2$, which is the best that can be expected from our family of trial spaces $(X^\delta)_{\delta\in\Delta}$. For the singular problem, the reduced rate 0.4 is found; it is unknown if a better rate can be expected.

Memory

The right pane of Figure 7.8 shows the peak memory consumption after every iteration of the adaptive algorithm. We see that the peak memory is linear in $\dim X^\delta$, stabilizing to around 15kB per degree of freedom. This is relatively high due to our implementation based on double-trees. In fact, the double-trees together make up around 85% of the total memory. For the singular problem, the largest double-tree $\mathbf{\Lambda}_Y^\delta$ occupies around 40% of the total memory.

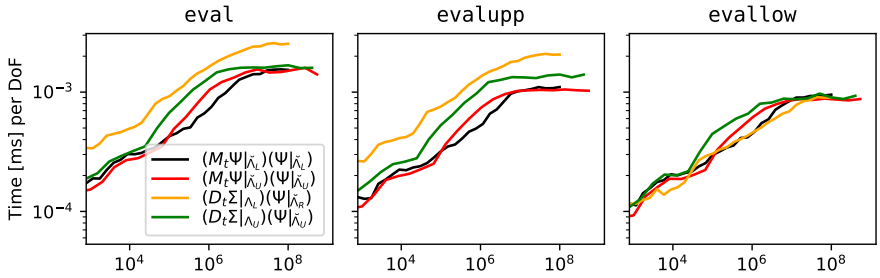


Figure 7.9. Time (in ms) per DoF of bilinear form evaluations in time.

7.6.2 Linearity of operations

The majority of our runtime is spent in the application of bilinear forms. In this section, we measure the application times to assert their linear complexity.

In time

We select three sequences $\{\Lambda_U\}$, $\{\Lambda_L\}$, $\{\Lambda_R\}$ of trees in $\vee_\Sigma$, one uniformly refined and two graded towards the left and right respectively. For each such tree $\Lambda \subset \vee_\Sigma$, we define a corresponding tree $\check{\Lambda} := \{\mu \in \vee_\Xi : \exists \lambda \in \Lambda, |\lambda| = |\mu|, |\operatorname{supp} \xi_\mu \cap \operatorname{supp} \sigma_\lambda| > 0\} \subset \vee_\Xi$.

We select the bilinear forms M_t and D_t from (7.27), and run the algorithms from §7.3.1. We see in Figure 7.9 that the runtime per degree of freedom stabilizes to 10^{-3} ms, essentially independent of the bilinear form and the trees. We suspect the increase until 10^7 degrees of freedom has to do with cache locality.

In space

On the L-shaped domain $\Omega := [-1,1]^2 \setminus [-1,0]^2$, we select two sequences of hierarchical basis trees, one uniformly refined and the other refined by a standard adaptive loop on $-\Delta u = 1$, $u|_{\partial\Omega} = 0$.

For a hierarchical basis tree $\Psi_\mathcal{T} = \{\psi_\nu : \nu \in N_{\mathcal{T},0}\}$, we denote the stiffness matrix $\langle \nabla\Psi_\mathcal{T}, \nabla\Psi_\mathcal{T}\rangle_{L_2(\Omega)}$ as $\mathbf{A}_\mathcal{T}$. We measure the runtime of the conversion from vertex tree $N_\mathcal{T}$ to triangulation $\mathcal{T}$ (cf. §7.5.1), the application time of $\mathbf{A}_\mathcal{T}$ through (7.29), and that of multigrid on $\mathbf{A}_\mathcal{T}$ (with 1 and 3 V-cycles) through Algorithm 7.7. Figure 7.10 confirms that the relative runtime of every operation is essentially independent of the refinement strategy. Interesting is again the increase until 10^5 degrees of freedom.

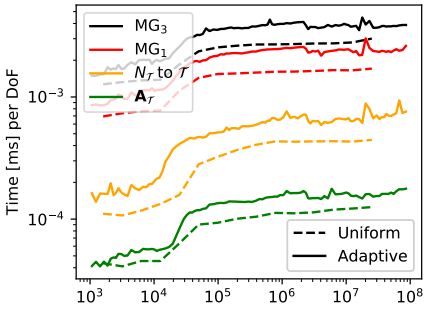


FIGURE 7.10. Time (in ms) per DoF of important operations in space, for uniform and adaptive refinements.

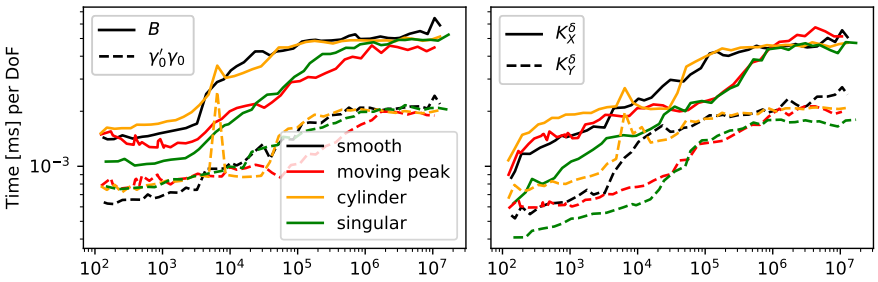


FIGURE 7.11. Time (in ms) per DoF of the four bilinear forms applied in the solve step of the adaptive algorithm.

In space-time

Solving (7.9) using PCG requires the application of the four linear operators $E_Y^\delta B E_X^\delta$, ${E_X^\delta}' \gamma_0' \gamma_0 E_X^\delta$, K_X^δ, and K_Y^δ. For the first two, Corollary 7.3.14 asserts that their application time is of linear complexity, while for the preconditioners K_X^δ and K_Y^δ, this follows from the block-diagonal structure of their matrix representation.

We run the adaptive algorithm on the four problems of §7.6.1. Figure 7.11 shows that the application time of the aforementioned operators is essentially independent of the problem, even though the underlying double-trees are vastly different. We again see an increase in relative runtime until 10^6 degrees of freedom.

Figure 7.12 shows the runtimes of the solve, estimate, mark and refine steps of the adaptive loop. We confirm that each step is of linear complexity, and that the total runtime is governed by the solve and estimate steps.

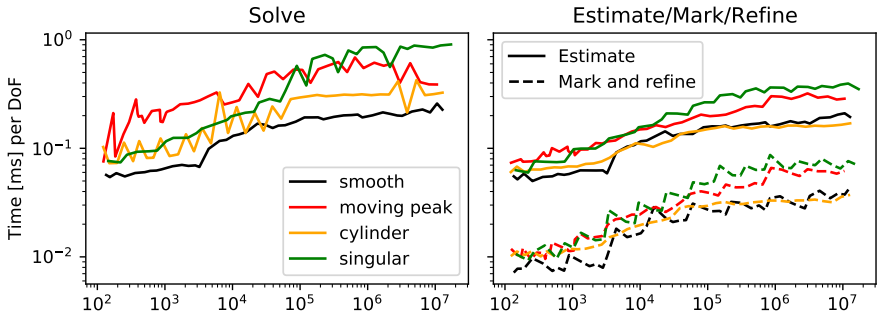


Figure 7.12. Time (in ms) per DoF of the steps in the adaptive loop.

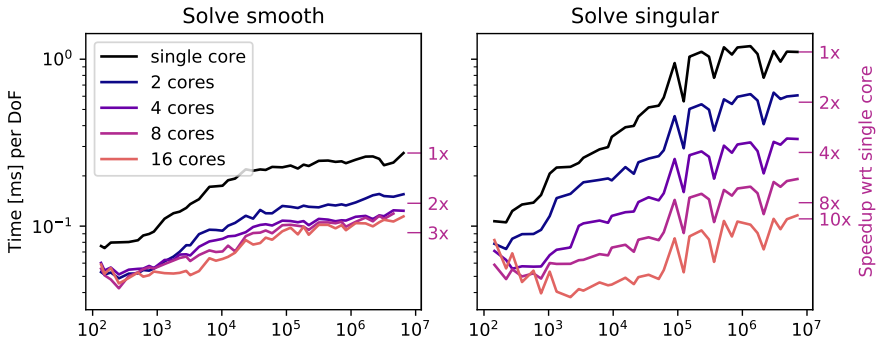


Figure 7.13. Speedup and time (in ms) per DoF of the solve step in the adaptive loop, for different number of parallel processors.

7.6.3 Shared-memory parallelism

Most of our execution time is spent applying the linear operators from Figure 7.11. We can obtain a significant speedup with multithreading. In Algorithm 7.9, all fibers inside each of the four for-loops are disjoint, and we can easily parallelize each loop using OpenMP.

We run the parallel code on the smooth and singular problems. The right pane of Figure 7.13 shows decent parallel performance for the singular problem, with $10\times$ speedup at 16 cores. The left pane however reveals a load balancing issue: as u is smooth, the two fibers $(\mathbf{\Lambda}_0^\delta)_{1,\lambda}$ with $|\lambda| = 0$ contain the majority of the degrees of freedom. This results in poor parallel efficiency for the first and fourth loop in Algorithm 7.9.

7.7 Conclusion

We discussed an implementation of an adaptive solver for a space-time variational formulation of parabolic evolution equations where every step is of linear complexity.

We constructed a family of trial spaces spanned by tensor-products of wavelets in time and hierarchical basis functions in space. The resulting adaptive loop is able to resolve singularities locally in space and time, and we proved its r-linear convergence.

After imposing a *double-tree* constraint on the index set of the trial spaces, we devised an abstract algorithm that is able to apply the system matrices in linear complexity. We achieve this complexity in practice by a *tree-based* implementation. The numerical results show high performance of the adaptive loop as a whole.

7.A Proofs of Theorems in §7.3

Theorem 7.3.7. *A call of* `eval` *yields the output as specified, at the cost of* $\mathcal{O}(\#\breve{\Pi} + \#\breve{\Lambda} + \#\Pi + \#\Lambda)$ *operations.*

Proof. By locality of the collections $\breve{\Phi}$ and $\breve{\Psi}$, and sparsity of the matrices $\breve{\mathfrak{p}}_\ell$ and $\breve{\mathfrak{q}}_\ell$, we see that $\#\underline{\breve{\Pi}} \lesssim \#\breve{\Pi}_B + \#\breve{\Lambda}_\ell \lesssim \#\Lambda_\ell + \#\breve{\Lambda}_\ell$. So after sufficiently many recursive calls, the current set $\breve{\Pi} \cup \breve{\Lambda}$ will be empty. For use later, we note that similarly $\#\underline{\Pi} \lesssim \#\Pi_B + \#\Lambda_\ell \lesssim \#\breve{\Lambda}_\ell + \#\breve{\Pi}_B + \#\Lambda_\ell \lesssim \#\Lambda_\ell + \#\breve{\Lambda}_\ell$.

For $\breve{\Pi} \cup \breve{\Lambda} = \emptyset$, the call produces nothing, which is correct.

Now let $\breve{\Pi} \cup \breve{\Lambda} \neq \emptyset$. From Λ being an ℓ-tree, the definitions of $S(\cdot)$ and $\breve{\Pi}_A$, and the locality of A, one has

$$\mathbf{e}|_{\breve{\Pi}_A} = (Au)(\breve{\Phi}|_{\breve{\Pi}_A}) = (A(\mathbf{d}^\top \Phi|_\Pi))(\breve{\Phi}|_{\breve{\Pi}_A}).$$

By choice of $\underline{\Pi}$ we have

$$\underline{u} := \underline{\mathbf{d}}^\top \Phi|_{\underline{\Pi}} + \mathbf{c}|_{\Lambda_{\ell+1\uparrow}}^\top \Psi|_{\Lambda_{\ell+1\uparrow}} = (\mathbf{d}|_{\Pi_B})^\top \Phi|_{\Pi_B} + \mathbf{c}^\top \Psi|_\Lambda = u - (\mathbf{d}|_{\Pi_A})^\top \Phi|_{\Pi_A}.$$

By induction the recursive call yields $\underline{\mathbf{e}} = (A\underline{u})(\breve{\Phi}|_{\underline{\breve{\Pi}}})$, and $\underline{\mathbf{f}} = (A\underline{u})(\breve{\Psi}|_{\breve{\Lambda}_{\ell+1\uparrow}})$. From $\breve{\Lambda}$ being an ℓ-tree, the definitions of $\breve{S}(\cdot)$ and Π_A, and the locality of A, we have

$$(Au)(\breve{\Psi}|_{\breve{\Lambda}_{\ell\uparrow}}) = (A\underline{u})(\breve{\Psi}|_{\breve{\Lambda}_{\ell\uparrow}}),$$

and so in particular $\mathbf{f}|_{\Lambda_{\ell+1,\uparrow}} = \underline{\mathbf{f}}$.

The definition of $\underline{\breve{\Pi}}$ shows that

$$\breve{\Phi}|_{\breve{\Pi}_B} = (\breve{\mathfrak{p}}_\ell^\top \breve{\Phi}|_{\underline{\breve{\Pi}}})|_{\breve{\Pi}_B}, \quad \breve{\Psi}|_{\breve{\Lambda}_\ell} = (\breve{\mathfrak{q}}_\ell^\top \breve{\Phi}|_{\underline{\breve{\Pi}}})|_{\breve{\Lambda}_\ell}.$$

We conclude that

$$\mathbf{f}|_{\breve{\Lambda}_\ell} = (Au)(\breve{\Psi}|_{\breve{\Lambda}_\ell}) = (A\underline{u})(\breve{\Psi}|_{\breve{\Lambda}_\ell}) = \big(\breve{\mathfrak{q}}_\ell^\top \underline{\mathbf{e}}\big)|_{\breve{\Lambda}_\ell},$$

and from $|\operatorname{supp}\phi_\lambda \cap \operatorname{supp}\breve{\phi}_\mu| = 0$ for $(\lambda,\mu) \in \Pi_A \times \breve{\Pi}_B$, that

$$\mathbf{e}|_{\breve{\Pi}_B} = (Au)(\breve{\Phi}|_{\breve{\Pi}_B}) = (A\underline{u})(\breve{\Phi}|_{\breve{\Pi}_B}) = \big(\breve{\mathfrak{p}}_\ell^\top \underline{\mathbf{e}}\big)|_{\breve{\Pi}_B}.$$

From the assumptions on the collections Φ, $\breve{\Phi}$, $\breve{\Psi}$, and Ψ, and their consequences on the sparsity of the matrices $\mathfrak{p}_\ell$, $\breve{\mathfrak{p}}_\ell$, $\mathfrak{q}_\ell$, and $\breve{\mathfrak{q}}_\ell$, one infers that the total cost of the evaluations of the statements in `eval` is $\mathcal{O}(\#\breve{\Pi} + \#\breve{\Lambda}_\ell + \#\Pi + \#\Lambda_\ell)$ plus the cost of the recursive call. Using $\#\underline{\breve{\Pi}} + \#\underline{\Pi} \lesssim \#\breve{\Lambda}_\ell + \#\Lambda_\ell$ and induction, we conclude the second statement of the theorem. □

Theorem 7.3.9. *A call of* `evalupp` *yields the output as specified, at the cost of* $\mathcal{O}(\#\breve{\Pi} + \#\breve{\Lambda} + \#\Pi + \#\Lambda)$ *operations.*

Proof. By locality of the collections $\breve{\Phi}$ and $\breve{\Psi}$, and sparsity of the matrices $\breve{\mathfrak{p}}_\ell$ and $\breve{\mathfrak{q}}_\ell$, we see that $\#\underline{\breve{\Pi}} \lesssim \#\breve{\Pi}_B + \#\breve{\Lambda}_\ell \lesssim \#\Lambda_\ell + \#\breve{\Lambda}_\ell$. So after sufficiently many recursive calls, the current set $\breve{\Pi} \cup \breve{\Lambda}$ will be empty. Notice that $\#\underline{\Pi} \lesssim \#\Lambda_\ell$.

For $\breve{\Pi} \cup \breve{\Lambda} = \emptyset$, the call produces nothing, which is correct.

Now let $\breve{\Pi} \cup \breve{\Lambda} \neq \emptyset$. From Λ being an ℓ-tree, the definitions of $S(\cdot)$ and $\breve{\Pi}_A$, and the locality of A, one has

$$\mathbf{e}|_{\breve{\Pi}_A} = (Au)(\breve{\Phi}|_{\breve{\Pi}_A}) = (A(\mathbf{d}^\top \Phi|_\Pi))(\breve{\Phi}|_{\breve{\Pi}_A}).$$

By definition of $\underline{\Pi}$ we have

$$\underline{u} := \underline{\mathbf{d}}^\top \Phi|_{\underline{\Pi}} + \mathbf{c}|_{\Lambda_{\ell+1\uparrow}}^\top \Psi|_{\Lambda_{\ell+1\uparrow}} = \mathbf{c}^\top \Psi|_\Lambda = u - \mathbf{d}^\top \Phi|_\Pi.$$

By induction the recursive call yields

$$\underline{\mathbf{e}} = (A\underline{u})(\breve{\Phi}|_{\underline{\breve{\Pi}}}), \quad \underline{\mathbf{f}} = \mathbf{U}_{\breve{\Lambda}_{\ell+1\uparrow} \times \Lambda_{\ell+1\uparrow}} c|_{\Lambda_{\ell+1\uparrow}} = \mathbf{f}|_{\breve{\Lambda}_{\ell+1\uparrow}}.$$

The definition of $\underline{\breve{\Pi}}$ shows that

$$\breve{\Phi}|_{\breve{\Pi}_B} = (\breve{\mathfrak{p}}_\ell^\top \breve{\Phi}|_{\underline{\breve{\Pi}}})|_{\breve{\Pi}_B}, \quad \breve{\Psi}|_{\breve{\Lambda}_\ell} = (\breve{\mathfrak{q}}_\ell^\top \breve{\Phi}|_{\underline{\breve{\Pi}}})|_{\breve{\Lambda}_\ell}.$$

We conclude that

$$\mathbf{f}|_{\breve{\Lambda}_\ell} = (A(\mathbf{c}^\top \Psi|_\Lambda))(\breve{\Psi}|_{\breve{\Lambda}_\ell}) = (A\underline{u})(\breve{\Psi}|_{\breve{\Lambda}_\ell}) = \big(\breve{\mathfrak{q}}_\ell^\top \underline{\mathbf{e}}\big)|_{\breve{\Lambda}_\ell},$$

and

$$\begin{aligned}\mathbf{e}|_{\breve{\Pi}_B} = (Au)(\breve{\Phi}|_{\breve{\Pi}_B}) &= (A\underline{u})(\breve{\Phi}|_{\breve{\Pi}_B}) + (A(\mathbf{d}^\top \Phi|_\Pi))(\breve{\Phi}|_{\breve{\Pi}_B}) \\ &= \big(\mathfrak{p}_\ell^\top \underline{\mathbf{e}}\big)|_{\breve{\Pi}_B} + (A(\mathbf{d}^\top \Phi|_\Pi))(\breve{\Phi}|_{\breve{\Pi}_B}).\end{aligned}$$

From the assumptions on the collections Φ, $\breve{\Phi}$, $\breve{\Psi}$, and Ψ, and their consequences on the sparsity of the matrices $\mathfrak{p}_\ell$, $\breve{\mathfrak{p}}_\ell$, $\mathfrak{q}_\ell$, and $\breve{\mathfrak{q}}_\ell$, one infers that the total cost of the evaluations of the statements in `eval` is $\mathcal{O}(\#\breve{\Pi} + \#\breve{\Lambda}_\ell + \#\Pi + \#\Lambda_\ell)$ plus the cost of the recursive call. Using $\#\underline{\breve{\Pi}} + \#\underline{\Pi} \lesssim \#\breve{\Lambda}_\ell + \#\Lambda_\ell$ and induction, we conclude the second statement of the theorem. □

Theorem 7.3.11. *A call of* `evallow` *yields the output as specified, at the cost of* $\mathcal{O}(\#\breve{\Lambda} + \#\Pi + \#\Lambda)$ *operations.*

Proof. Notice that $\#\underline{\Pi} \lesssim \#\Lambda_\ell + \#\Pi_B \lesssim \#\Lambda_\ell + \#\breve{\Lambda}_\ell$.

For $\breve{\Pi} \cup \breve{\Lambda} = \emptyset$, the call produces nothing, which is correct.

Now let $\breve{\Pi} \cup \breve{\Lambda} \neq \emptyset$. The definitions of $\underline{\breve{\Pi}}$ and $\underline{\Pi}_B$ show that

$$\begin{aligned}\mathbf{f}|_{\breve{\Lambda}_\ell} &= (A\Phi|_\Pi)(\breve{\Psi}|_{\breve{\Lambda}_\ell})\mathbf{d} = (A\Phi|_\Pi)(\breve{\Psi}|_{\breve{\Lambda}_\ell})\mathbf{d}|_{\Pi_B} \\ &= \big(\breve{\mathfrak{q}}_\ell^\top (A\Phi|_{\underline{\Pi}_B})(\breve{\Phi}|_{\underline{\breve{\Pi}}})\mathfrak{p}_\ell \mathbf{d}|_{\Pi_B}\big)|_{\breve{\Lambda}_\ell} = (\breve{\mathfrak{q}}_\ell^\top \underline{\mathbf{e}})|_{\breve{\Lambda}_\ell}\end{aligned}$$

From $\breve{\Lambda}$ being an ℓ-tree, the definitions of $\breve{S}(\cdot)$ and Π_B, and the locality of a, and for the third equality, the definition of $\underline{\Pi}$, one has

$$\begin{aligned}f|_{\breve{\Lambda}_{\ell+1\uparrow}} &= a(\breve{\Psi}|_{\breve{\Lambda}_{\ell+1\uparrow}}, \Phi|_\Pi)\mathbf{d} + \mathbf{L}|_{\breve{\Lambda}_{\ell+1\uparrow}\times\Lambda_\ell}\mathbf{c}|_{\Lambda_\ell} + \mathbf{L}|_{\breve{\Lambda}_{\ell+1\uparrow}\times\Lambda_{\ell+1\uparrow}}\mathbf{c}|_{\Lambda_{\ell+1\uparrow}} \\ &= (A\Phi|_\Pi)(\breve{\Psi}|_{\breve{\Lambda}_{\ell+1\uparrow}})\mathbf{d}|_{\Pi_B} + (A\Psi|_{\Lambda_\ell})(\breve{\Psi}|_{\breve{\Lambda}_{\ell+1\uparrow}})\mathbf{c}|_{\Lambda_\ell} + \mathbf{L}|_{\breve{\Lambda}_{\ell+1\uparrow}\times\Lambda_{\ell+1\uparrow}}\mathbf{c}|_{\Lambda_{\ell+1\uparrow}} \\ &= (A\Phi|_{\underline{\Pi}})(\breve{\Psi}|_{\breve{\Lambda}_{\ell+1\uparrow}})\underline{\mathbf{d}} + \mathbf{L}|_{\breve{\Lambda}_{\ell+1\uparrow}\times\Lambda_{\ell+1\uparrow}}\mathbf{c}|_{\Lambda_{\ell+1\uparrow}} \\ &= \texttt{evallow}(A)(\ell+1, \breve{\Lambda}_{\ell+1\uparrow}, \underline{\Pi}, \Lambda_{\ell+1\uparrow}, \underline{\mathbf{d}}, \mathbf{c}|_{\Lambda_{\ell+1\uparrow}})\end{aligned}$$

by induction.

From the assumptions on the collections Φ, $\breve{\Psi}$, and Ψ, and their consequences on the sparsity of the matrices $\mathfrak{p}_\ell$, $\mathfrak{q}_\ell$, and $\breve{\mathfrak{q}}_\ell$, one easily infers that the total cost of the evaluations of the statements in `evallow` is $\mathcal{O}(\#\breve{\Lambda}_\ell + \#\Pi + \#\Lambda_\ell)$ plus the cost of the recursive call. Using $\#\underline{\Pi} \lesssim \#\breve{\Lambda}_\ell + \#\Lambda_\ell$ and induction, we conclude the second statement of the theorem. □

Theorem 7.3.13. *Let* $\breve{\boldsymbol{\Lambda}} \subset \breve{\vee}^0 \times \breve{\vee}^1$, $\boldsymbol{\Lambda} \subset \vee^0 \times \vee^1$ *be finite double-trees. Then*

$$\boldsymbol{\Sigma} := \bigcup_{\lambda \in P_0\boldsymbol{\Lambda}} \Big(\{\lambda\} \times \bigcup_{\{\mu \in P_0\breve{\boldsymbol{\Lambda}} : |\mu| = |\lambda|+1,\ |\breve{S}^0(\mu) \cap S^0(\lambda)| > 0\}} \breve{\boldsymbol{\Lambda}}_{1,\mu}\Big),$$

$$\boldsymbol{\Theta} := \bigcup_{\lambda \in P_1\boldsymbol{\Lambda}} \Big(\{\mu \in P_0\breve{\boldsymbol{\Lambda}} : \exists \gamma \in \boldsymbol{\Lambda}_{0,\lambda} \text{ s.t. } |\gamma| = |\mu|,\ |\breve{S}^0(\mu) \cap S^0(\gamma)| > 0\} \times \{\lambda\}\Big),$$

are double-trees with $\#\boldsymbol{\Sigma} \lesssim \#\breve{\boldsymbol{\Lambda}}$ *and* $\#\boldsymbol{\Theta} \lesssim \#\boldsymbol{\Lambda}$*, and*

$$\begin{aligned}R_{\breve{\boldsymbol{\Lambda}}}(\mathbf{A}_0 \otimes \mathbf{A}_1)I_{\boldsymbol{\Lambda}} = {} & R_{\breve{\boldsymbol{\Lambda}}}(\mathbf{L}_0 \otimes \mathrm{Id})I_{\boldsymbol{\Sigma}}R_{\boldsymbol{\Sigma}}(\mathrm{Id} \otimes \mathbf{A}_1)I_{\boldsymbol{\Lambda}} + \\ & R_{\breve{\boldsymbol{\Lambda}}}(\mathrm{Id} \otimes \mathbf{A}_1)I_{\boldsymbol{\Theta}}R_{\boldsymbol{\Theta}}(\mathbf{U}_0 \otimes \mathrm{Id})I_{\boldsymbol{\Lambda}}.\end{aligned}$$

Proof. We write

$$R_{\breve{\mathbf{\Lambda}}}(\mathbf{A}_0 \otimes \mathbf{A}_1) I_{\mathbf{\Lambda}} = R_{\breve{\mathbf{\Lambda}}}((\mathbf{L}_0 + \mathbf{U}_0) \otimes \mathbf{A}_1) I_{\mathbf{\Lambda}}$$

$$= R_{\breve{\mathbf{\Lambda}}}(\mathbf{L}_0 \otimes \mathrm{Id})(\mathrm{Id} \otimes \mathbf{A}_1) I_{\mathbf{\Lambda}} + \tag{7.32}$$

$$R_{\breve{\mathbf{\Lambda}}}(\mathrm{Id} \otimes \mathbf{A}_1)(\mathbf{U}_0 \otimes \mathrm{Id}) I_{\mathbf{\Lambda}}. \tag{7.33}$$

Considering (7.32), the range of $(\mathrm{Id} \otimes \mathbf{A}_1) I_{\mathbf{\Lambda}}$ consists of vectors whose entries with first index outside $P_0\mathbf{\Lambda}$ are zero. In view of the subsequent application of $\mathbf{L}_0 \otimes \mathrm{Id}$, furthermore only those indices $(\lambda, \gamma) \in P_0\mathbf{\Lambda} \times \breve{\vee}^1$ of these vectors might be relevant for which $\exists (\mu, \gamma) \in \breve{\mathbf{\Lambda}}$, i.e. $\gamma \in \mathbf{\Lambda}_{1,\mu}$, with $|\mu| > |\lambda|$ and $|\breve{S}^0(\mu) \cap S^0(\lambda)| > 0$. Indeed $|\breve{S}^0(\mu) \cap S^0(\lambda)| = 0$ implies $|\operatorname{supp} \breve{\psi}^0_\mu \cap \operatorname{supp} \psi^0_\lambda| = 0$, and so $A_0(\breve{\psi}^0_\mu, \psi^0_\lambda) = 0$. If for given (λ, γ) such a pair (μ, γ) exists for $|\mu| > |\lambda|$, then such a pair exists for $|\mu| = |\lambda| + 1$ as well, because $\breve{\mathbf{\Lambda}}_{0,\gamma}$ is a tree, and $\breve{S}^0(\mu') \supset \breve{S}^0(\mu)$ for any ancestor μ' of μ. In order words, the condition $|\mu| > |\lambda|$ can be read as $|\mu| = |\lambda| + 1$. The set of (λ, γ) that we just described is given by the set $\mathbf{\Sigma}$, and so we infer that

$$R_{\breve{\mathbf{\Lambda}}}(\mathbf{L}_0 \otimes \mathrm{Id})(\mathrm{Id} \otimes \mathbf{A}_1) I_{\mathbf{\Lambda}} = R_{\breve{\mathbf{\Lambda}}}(\mathbf{L}_0 \otimes \mathrm{Id}) I_{\mathbf{\Sigma}} R_{\mathbf{\Sigma}} (\mathrm{Id} \otimes \mathbf{A}_1) I_{\mathbf{\Lambda}}.$$

Now let $(\lambda, \gamma) \in \mathbf{\Sigma}$. Using that $P_0\mathbf{\Lambda}$ is a tree, and $S^0(\lambda) \subset S^0(\lambda')$ for any ancestor λ' of λ, we infer that $(\lambda', \gamma) \in \mathbf{\Sigma}$. Using that for any $\mu \in P_0\breve{\mathbf{\Lambda}}$, $\breve{\mathbf{\Lambda}}_{1,\mu}$ is a tree, we infer that for any ancestor γ' of γ, $(\lambda, \gamma') \in \mathbf{\Sigma}$, so that $\mathbf{\Sigma}$ is a double-tree.

For any $\mu \in \breve{\vee}^0$, the number of $\lambda \in \vee^0$ with $|\mu| = |\lambda| + 1$ and $|\breve{S}^0(\mu) \cap S^0(\lambda)| > 0$ is uniformly bounded, from which we infer that $\#\mathbf{\Sigma} \lesssim \sum_{\mu \in P_0\breve{\mathbf{\Lambda}}} \#\breve{\mathbf{\Lambda}}_{1,\mu} = \#\breve{\mathbf{\Lambda}}$.

Considering (7.33), the range of $(\mathbf{U}_0 \otimes \mathrm{Id}) I_{\mathbf{\Lambda}}$ consists of vectors that can only have non-zero entries for indices $(\mu, \lambda) \in \breve{\vee}^0 \times P_1\mathbf{\Lambda}$ for which there exists a $\gamma \in \mathbf{\Lambda}_{0,\lambda}$ with $|\gamma| \geq |\mu|$ and $|\breve{S}^0(\mu) \cap S^0(\gamma)| > 0$. Since $\mathbf{\Lambda}_{0,\lambda}$ is a tree, and $S^0(\gamma') \supset S^0(\gamma)$ for any ancestor γ' of γ, equivalently $|\gamma| \geq |\mu|$ can be read as $|\gamma| = |\mu|$. Furthermore, in view of the subsequent application of $R_{\breve{\mathbf{\Lambda}}}(\mathrm{Id} \otimes \mathbf{A}_1)$, it suffices to consider those indices (μ, λ) with $\mu \in P_0\breve{\mathbf{\Lambda}}$. The set of (μ, λ) that we just described is given by the set $\mathbf{\Theta}$, and so we infer that

$$R_{\breve{\mathbf{\Lambda}}}(\mathrm{Id} \otimes \mathbf{A}_1)(\mathbf{U}_0 \otimes \mathrm{Id}) I_{\mathbf{\Lambda}} = R_{\breve{\mathbf{\Lambda}}}(\mathrm{Id} \otimes \mathbf{A}_1) I_{\mathbf{\Theta}} R_{\mathbf{\Theta}} (\mathbf{U}_0 \otimes \mathrm{Id}) I_{\mathbf{\Lambda}}.$$

Now let $(\mu, \lambda) \in \mathbf{\Theta}$. If λ' is an ancestor of λ, then from $P_0\mathbf{\Lambda}$ being a tree, and $\mathbf{\Lambda}_{0,\lambda} \subset \mathbf{\Lambda}_{0,\lambda'}$, we have $(\mu, \lambda') \in \mathbf{\Theta}$. If μ' is an ancestor of μ, then from $P_0\breve{\mathbf{\Lambda}}$ being a tree, and $\breve{S}^0(\mu') \supset \breve{S}^0(\mu)$, we infer that $(\mu', \lambda) \in \mathbf{\Theta}$, and thus that $\mathbf{\Theta}$ is a double-tree.

For any $\gamma \in \vee^0$, the number of $\mu \in \breve{\vee}^0$ with $|\mu| = |\gamma|$ and $|\breve{S}^0(\mu) \cap S^0(\gamma)| > 0$ is uniformly bounded, from which we infer that $\#\mathbf{\Theta} \lesssim \sum_{\lambda \in P_1\mathbf{\Lambda}} \#\mathbf{\Lambda}_{0,\lambda} = \#\mathbf{\Lambda}$. □

8 A parallel algorithm

8.1 Introduction

This chapter deals with solving parabolic evolution equations in a time-parallel fashion using tensor-product discretizations. Time-parallel algorithms for solving parabolic evolution equations have become a focal point following the enormous increase in parallel computing power. Spatial parallelism is a ubiquitous component in large-scale computations, but when spatial parallelism is exhausted, parallelization of the time axis is of interest.

Time-stepping methods first discretize the problem in space, and then solve the arising system of coupled ODEs sequentially, immediately revealing a primary source of difficulty for time-parallel computation.

Alternatively, one can solve simultaneously in space *and* time. Originally introduced in [BJ89, BJ90], these space-time methods are very flexible: some can guarantee quasi-best approximations, meaning that their error is proportional to that of the best approximation from the trial space [And13, DS18, FK21, SZ20], or drive adaptive routines [SY18, RS19]. Many are especially well-suited for time-parallel computation [GN16, NS19]. Since the first significant contribution to time-parallel algorithms [Nie64] in 1964, many methods suitable for parallel computation have surfaced; see the review [Gan15].

Parallel complexity

The (serial) complexity of an algorithm measures asymptotic runtime on a single processor in terms of the input size. *Parallel complexity* measures asymptotic runtime given *sufficiently many* parallel processors having access to a shared memory, i.e., assuming there are no communication costs.

In the current context of tensor-product discretizations of parabolic PDEs, we denote with N_t and $N_{\mathbf{x}}$ the number of unknowns in time and space respectively.

The parareal method [LMT01] aims at time-parallelism by alternating a serial coarse-grid solve with fine-grid computations in parallel. This way, each

iteration has a time-parallel complexity of $\mathcal{O}(\sqrt{N_t}N_{\mathbf{x}})$, and combined with parallel multigrid in space, a parallel complexity of $\mathcal{O}(\sqrt{N_t}\log N_{\mathbf{x}})$. The popular MGRIT algorithm extends these ideas to multiple levels in time; cf. [FFK+14].

Recently, Neumüller and Smears proposed an iterative algorithm that uses a Fast Fourier Transform in time. Each iteration runs serially in $\mathcal{O}(N_t\log(N_t)N_{\mathbf{x}})$ and parallel in time, in $\mathcal{O}(\log(N_t)N_{\mathbf{x}})$. By also incorporating parallel multigrid in space, its parallel runtime may be reduced to $\mathcal{O}(\log N_t + \log N_{\mathbf{x}})$.

Our contribution

We study a variational formulation introduced in [SW21b] which was based on work by Andreev [And13, And16]. Recently in [SvVW21, vVW21b], we studied this formulation in the context of space-time adaptivity and its efficient implementation in serial and on shared-memory parallel computers. The current chapter instead focuses on its massively parallel implementation and time-parallel performance.

Our method has remarkable similarities with the approach of [NS19], and the most essential difference is the substitution of their Fast Fourier Transform by our Fast Wavelet Transform. The strengths of both methods include a solid inf-sup theory that enables quasi-optimal approximate solutions from the trial space, ease of implementation, and excellent parallel performance in practice.

Our method has another strength: based on a wavelet transform, for fixed algebraic tolerance it runs serially in linear complexity. Parallel in time, it runs in complexity $\mathcal{O}(\log(N_t)N_{\mathbf{x}})$; parallel in *space and time*, in $\mathcal{O}(\log(N_tN_{\mathbf{x}}))$. Moreover, when solving to an algebraic error proportional to the discretization error, incorporating a *nested iteration* (cf. [Hac85, Ch. 5]) results in complexities $\mathcal{O}(N_tN_{\mathbf{x}})$, $\mathcal{O}(\log(N_t)N_{\mathbf{x}})$, and $\mathcal{O}(\log^2(N_tN_{\mathbf{x}}))$ respectively. This is on par with best-known results on parallel complexity for elliptic problems; see also [Bra81].

Organization of this chapter

In §8.2, we formally introduce the problem, derive a saddle-point formulation, and provide sufficient conditions for quasi-optimality of discrete solutions. In §8.3, we detail on the efficient computation of these discrete solutions. In §8.4 we take a concrete example—the reaction-diffusion equation—and analyze the serial and parallel complexity of our algorithm. In §8.5, we test these theoretical findings in practice. We conclude in §8.6.

Notations

For normed linear spaces U and V, in this work for convenience over $\mathbb{R}$, $\mathcal{L}(U,V)$ will denote the space of bounded linear mappings $U \to V$ endowed with the operator norm $\|\cdot\|_{\mathcal{L}(U,V)}$. The subset of invertible operators in $\mathcal{L}(U,V)$ with inverses in $\mathcal{L}(V,U)$ will be denoted as $\mathcal{L}\mathrm{is}(U,V)$.

Given a finite-dimensional subspace U^δ of a normed linear space U, we denote the trivial embedding $U^\delta \to U$ by E_U^δ. For a basis Φ^δ—viewed formally as a column vector—of U^δ, we define the *synthesis operator* as

$$\mathcal{F}_{\Phi^\delta} : \mathbb{R}^{\dim U^\delta} \to U^\delta : \boldsymbol{c} \mapsto \boldsymbol{c}^\top \Phi^\delta =: \sum_{\phi \in \Phi^\delta} c_\phi \phi.$$

Equip $\mathbb{R}^{\dim U^\delta}$ with the Euclidean inner product and identify $(\mathbb{R}^{\dim U^\delta})'$ with $\mathbb{R}^{\dim U^\delta}$ using the corresponding Riesz map. We find the adjoint of $\mathcal{F}_{\Phi^\delta}$, the *analysis operator*, to satisfy

$$(\mathcal{F}_{\Phi^\delta})' : (U^\delta)' \to \mathbb{R}^{\dim U^\delta} : f \mapsto f(\Phi^\delta) := [f(\phi)]_{\phi \in \Phi^\delta}.$$

For quantities f and g, by $f \lesssim g$, we mean that $f \leq C \cdot g$ with a constant that does not depend on parameters that f and g may depend on. By $f \eqsim g$, we mean that $f \lesssim g$ and $g \lesssim f$. For matrices $\boldsymbol{A}$ and $\boldsymbol{B} \in \mathbb{R}^{N \times N}$, by $\boldsymbol{A} \eqsim \boldsymbol{B}$ we will denote *spectral equivalence*, i.e. $\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} \eqsim \boldsymbol{x}^\top \boldsymbol{B} \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{R}^N$.

8.2 Quasi-optimal approximations to the parabolic problem

Let V, H be separable Hilbert spaces of functions on some spatial domain such that V is continuously embedded in H, i.e. $V \hookrightarrow H$, with dense compact embedding. Identifying H with its dual yields the Gelfand triple $V \hookrightarrow H \simeq H' \hookrightarrow V'$.

For a.e.

$$t \in I := (0, T),$$

let $a(t; \cdot, \cdot)$ denote a bilinear form on $V \times V$ so that for any $\eta, \zeta \in V$, $t \mapsto a(t; \eta, \zeta)$ is measurable on I, and such that for a.e. $t \in I$,

$$\begin{aligned} |a(t; \eta, \zeta)| &\lesssim \|\eta\|_V \|\zeta\|_V \quad (\eta, \zeta \in V) \quad \text{(\textit{boundedness})}, \\ a(t; \eta, \eta) &\gtrsim \|\eta\|_V^2 \qquad (\eta \in V) \qquad \text{(\textit{coercivity})}. \end{aligned}$$

With $(A(t)\cdot)(\cdot) := a(t; \cdot, \cdot) \in \mathcal{L}\mathrm{is}(V, V')$, given a forcing function g and initial value u_0, we want to solve the *parabolic initial value problem* of

$$\text{(8.1)} \quad \text{finding } u : I \to V \text{ such that } \quad \begin{cases} \frac{\mathrm{d}u}{\mathrm{d}t}(t) + A(t)u(t) &= g(t) \quad (t \in I), \\ u(0) &= u_0. \end{cases}$$

8.2.1 An equivalent self-adjoint saddle-point system

In a simultaneous space-time variational formulation, the parabolic problem reads as finding u from a suitable space of functions of time and space s.t.

$$(Bw)(v) := \int_I \langle \tfrac{\mathrm{d}w}{\mathrm{d}t}(t), v(t) \rangle_H + a(t; w(t), v(t)) \mathrm{d}t = \int_I \langle g(t), v(t) \rangle_H =: g(v)$$

for all v from another suitable space of functions of time and space. One possibility to enforce the initial condition is by testing against additional test functions.

Theorem 8.2.1 ([SS09]). *With $X := L_2(I;V)\cap H^1(I;V')$, $Y := L_2(I;V)$, we have*

$$\begin{bmatrix} B \\ \gamma_0 \end{bmatrix} \in \mathcal{L}\mathrm{is}(X, Y' \times H),$$

where for $t \in \bar{I}$, $\gamma_t \colon u \mapsto u(t,\cdot)$ denotes the trace map. In other words,

$$\text{(8.2)} \qquad \textit{finding } u \in X \textit{ s.t. } \quad (Bu, \gamma_0 u) = (g, u_0) \quad \textit{given} \quad (g, u_0) \in Y' \times H$$

is a well-posed simultaneous space-time variational formulation of (8.1).

We define $A \in \mathcal{L}\mathrm{is}(Y, Y')$ and $\partial_t \in \mathcal{L}\mathrm{is}(X, Y')$ as

$$(Au)(v) := \int_I a(t; u(t), v(t))\mathrm{d}t\,, \quad \text{and} \quad \partial_t := B - A.$$

Following [SW21b], we assume that A is *symmetric*. We can reformulate (8.2) as the self-adjoint saddle point problem

$$\text{(8.3)} \qquad \text{finding } (v,\sigma,u) \in Y \times H \times X \text{ s.t.} \quad \begin{bmatrix} A & 0 & B \\ 0 & \mathrm{Id} & \gamma_0 \\ B' & \gamma_0' & 0 \end{bmatrix} \begin{bmatrix} v \\ \sigma \\ u \end{bmatrix} = \begin{bmatrix} g \\ u_0 \\ 0 \end{bmatrix}.$$

By taking a Schur complement w.r.t. the H-block, we can reformulate this as

$$\text{(8.4)} \qquad \text{finding } (v,u) \in Y \times X \text{ s.t.} \quad \begin{bmatrix} A & B \\ B' & -\gamma_0'\gamma_0 \end{bmatrix} \begin{bmatrix} v \\ u \end{bmatrix} = \begin{bmatrix} g \\ -\gamma_0' u_0 \end{bmatrix}.$$

We equip Y and X with *'energy'-norms*

$$\|\cdot\|_Y^2 := (A\cdot)(\cdot), \quad \|\cdot\|_X^2 := \|\partial_t \cdot\|_{Y'}^2 + \|\cdot\|_Y^2 + \|\gamma_T \cdot\|_H^2,$$

which are equivalent to the canonical norms on Y and X.

8.2.2 Uniformly quasi-optimal Galerkin discretizations

Our numerical approximations will be based on the saddle-point formulation (8.4). Let $(Y^\delta, X^\delta)_{\delta\in\Delta}$ be a collection of closed subspaces of $Y \times X$ satisfying

$$\text{(8.5)} \qquad X^\delta \subset Y^\delta, \quad \partial_t X^\delta \subset Y^\delta \quad (\delta \in \Delta),$$

and

$$\text{(8.6)} \qquad 1 \geq \gamma_\Delta := \inf_{\delta\in\Delta} \inf_{0\neq u\in X^\delta} \sup_{0\neq v\in Y^\delta} \frac{(\partial_t u)(v)}{\|\partial_t u\|_{Y'}\|v\|_Y} > 0.$$

Remark 8.2.2. In [SW21b, §4], these conditions were verified for X^δ and Y^δ being tensor-products of (locally refined) finite element spaces in time and space. In Chapter 6, we relax these conditions to X_t^δ and Y^δ being *adaptive sparse grids*, allowing adaptive refinement locally in space *and* time simultaneously.

For $\delta \in \Delta$, let $(v^\delta, \overline{u}^\delta) \in Y^\delta \times X^\delta$ solve the Galerkin discretization of (8.4):

$$\begin{bmatrix} {E_Y^\delta}' A E_Y^\delta & {E_Y^\delta}' B E_X^\delta \\ {E_X^\delta}' B' E_Y^\delta & -{E_X^\delta}' \gamma_0' \gamma_0 E_X^\delta \end{bmatrix} \begin{bmatrix} v^\delta \\ \overline{u}^\delta \end{bmatrix} = \begin{bmatrix} {E_Y^\delta}' g \\ -{E_X^\delta}' \gamma_0' u_0 \end{bmatrix}. \tag{8.7}$$

The solution $(v^\delta, \overline{u}^\delta)$ of (8.7) exists uniquely, and exhibits *uniform quasi-optimality* in that $\|u - \overline{u}^\delta\|_X \leq \gamma_\Delta^{-1} \inf_{u_\delta \in X^\delta} \|u - u_\delta\|_X$ for all $\delta \in \Delta$.

Instead of solving a matrix representation of (8.7) using e.g. preconditioned MINRES, we will opt for a computationally more attractive method. By taking the Schur complement w.r.t. the Y^δ-block in (8.7), and replacing $({E_Y^\delta}' A E_Y^\delta)^{-1}$ in the resulting formulation by a *preconditioner* K_Y^δ that can be applied cheaply, we arrive at the *Schur complement formulation* of finding $u^\delta \in X^\delta$ s.t.

$$\underbrace{{E_X^\delta}'(B' E_Y^\delta K_Y^\delta {E_Y^\delta}' B + \gamma_0' \gamma_0) E_X^\delta}_{=:S^\delta} u^\delta = \underbrace{{E_X^\delta}'(B' E_Y^\delta K_Y^\delta {E_Y^\delta}' g + \gamma_0' u_0)}_{=:f^\delta}. \tag{8.8}$$

The resulting operator $S^\delta \in \mathcal{L}\mathrm{is}(X^\delta, {X^\delta}')$ is self-adjoint and elliptic. Given a self-adjoint operator $K_Y^\delta \in \mathcal{L}({Y^\delta}', Y^\delta)$ satisfying, for some $\kappa_\Delta \geq 1$,

$$\frac{\big((K_Y^\delta)^{-1} v\big)(v)}{(Av)(v)} \in [\kappa_\Delta^{-1}, \kappa_\Delta] \quad (\delta \in \Delta,\ \ v \in Y^\delta), \tag{8.9}$$

the solution u^δ of (8.8) exists uniquely as well. In fact, the following holds.

Theorem 8.2.3 ([SW21b, Rem. 3.8])**.** *Take* $(Y^\delta \times X^\delta)_{\delta \in \Delta}$ *satisfying* (8.5)–(8.6)*, and* K_Y^δ *satisfying* (8.9)*. Solutions* $u^\delta \in X^\delta$ *of* (8.8) *are uniformly quasi-optimal, i.e.*

$$\|u - u^\delta\|_X \leq \frac{\kappa_\Delta}{\gamma_\Delta} \inf_{u_\delta \in X^\delta} \|u - u_\delta\|_X \quad (\delta \in \Delta).$$

8.3 Solving efficiently on tensor-product discretizations

From now on, we assume that $X^\delta := X_t^\delta \otimes X_{\mathbf{x}}^\delta$ and $Y^\delta := Y_t^\delta \otimes Y_{\mathbf{x}}^\delta$ are *tensor-products*, and for ease of presentation, we assume that the spatial discretizations on X^δ and Y^δ coincide, i.e. $X_{\mathbf{x}}^\delta = Y_{\mathbf{x}}^\delta$, reducing (8.5) to $X_t^\delta \subset Y_t^\delta$ and $\frac{\mathrm{d}}{\mathrm{d}t} X_t^\delta \subset Y_t^\delta$.

We equip X_t^δ with a basis Φ_t^δ, $X_{\mathbf{x}}^\delta$ with $\Phi_{\mathbf{x}}^\delta$, and Y_t^δ with Ξ^δ.

8.3.1 Construction of K_Y^δ

Define $\boldsymbol{O} := \langle \Xi^\delta, \Xi^\delta \rangle_{L_2(I)}$ and $\boldsymbol{A}_{\mathbf{x}} := \langle \Phi_{\mathbf{x}}^\delta, \Phi_{\mathbf{x}}^\delta \rangle_V$. Given $\boldsymbol{K}_{\mathbf{x}} \eqsim \boldsymbol{A}_{\mathbf{x}}^{-1}$ uniformly in $\delta \in \Delta$, define

$$\boldsymbol{K}_Y := \boldsymbol{O}^{-1} \otimes \boldsymbol{K}_{\mathbf{x}}.$$

Then, the preconditioner $K_Y^\delta := \mathcal{F}_{\Xi^\delta \otimes \Phi_{\mathbf{x}}^\delta} \boldsymbol{K}_Y (\mathcal{F}_{\Xi^\delta \otimes \Phi_{\mathbf{x}}^\delta})' \in \mathcal{L}(Y^{\delta\prime}, Y^\delta)$ satisfies (8.9); cf. §6.5.6.

When Ξ^δ is orthogonal, $\boldsymbol{O}$ is diagonal and can be inverted exactly. For standard finite element bases $\Phi_{\mathbf{x}}^\delta$, suitable $\mathbf{K}_{\mathbf{x}}$ that can be applied efficiently (at cost linear in the discretization size) are provided by symmetric multigrid methods.

8.3.2 Preconditioning the Schur complement formulation

We will solve a matrix representation of (8.8) with an iterative solver, thus requiring a preconditioner. Inspired by the constructions of [And16, NS19], we build an *optimal* self-adjoint coercive preconditioner $K_X^\delta \in \mathcal{L}(X^{\delta\prime}, X^\delta)$ as a wavelet-in-time block-diagonal matrix with multigrid-in-space blocks.

Let U be a separable Hilbert space of functions over some domain. A given collection $\Psi = \{\psi_\lambda\}_{\lambda \in \vee_\Psi}$ is a *Riesz basis* for U when

$$\overline{\operatorname{span} \Psi} = U, \quad \text{and} \quad \|\boldsymbol{c}\|_{\ell_2(\vee_\Psi)} \eqsim \|\boldsymbol{c}^\top \Psi\|_U \quad \text{for all} \quad \boldsymbol{c} \in \ell_2(\vee_\Psi).$$

Thinking of Ψ being a basis of wavelet-type, for indices $\lambda \in \vee_\Psi$, its *level* is denoted $|\lambda| \in \mathbb{N}_0$. We call Ψ *uniformly local* when for all $\lambda \in \vee_\Psi$,

$$\operatorname{diam}(\operatorname{supp} \psi_\lambda) \lesssim 2^{-|\lambda|} \text{ and } \#\{\mu \in \vee_\Psi : |\mu| = |\lambda|, |\operatorname{supp} \psi_\mu \cap \operatorname{supp} \psi_\lambda| > 0\} \lesssim 1.$$

Assume $\Sigma := \{\sigma_\lambda : \lambda \in \vee_\Sigma\}$ is a uniformly local Riesz basis for $L_2(I)$ with $\{2^{-|\lambda|}\sigma_\lambda : \lambda \in \vee_\Sigma\}$ Riesz for $H^1(I)$. Writing $w \in X$ as $\sum_{\lambda \in \vee_\Sigma} \sigma_\lambda \otimes w_\lambda$ for some $w_\lambda \in V$, we define the bounded, symmetric, and coercive bilinear form

$$(D_X \sum_{\lambda \in \vee_\Sigma} \sigma_\lambda \otimes w_\lambda)(\sum_{\mu \in \vee_\Sigma} \sigma_\mu \otimes v_\mu) := \sum_{\lambda \in \vee_\Sigma} \langle w_\lambda, v_\lambda \rangle_V + 4^{|\lambda|} \langle w_\lambda, v_\lambda \rangle_{V'}.$$

The operator $D_X^\delta := {E_X^\delta}' D_X E_X^\delta$ is in $\mathcal{L}\mathrm{is}(X^\delta, X^{\delta\prime})$. Its norm and that of its inverse are bounded uniformly in $\delta \in \Delta$. When $X^\delta = \operatorname{span} \Sigma^\delta \otimes \Phi_{\mathbf{x}}^\delta$ for some $\Sigma^\delta := \{\sigma_\lambda : \lambda \in \vee_{\Sigma^\delta}\} \subset \Sigma$, the matrix representation of D_X^δ w.r.t. $\Sigma^\delta \otimes \Phi_{\mathbf{x}}^\delta$ is

$$(\mathcal{F}_{\Sigma^\delta \otimes \Phi^\delta})' D_X^\delta \mathcal{F}_{\Sigma^\delta \otimes \Phi^\delta} =: \mathbf{D}_X^\delta = \operatorname{blockdiag}[\boldsymbol{A}_{\mathbf{x}} + 4^{|\lambda|} \langle \Phi_{\mathbf{x}}^\delta, \Phi_{\mathbf{x}}^\delta \rangle_{V'}]_{\lambda \in \vee_{\Sigma^\delta}}.$$

Theorem 8.3.1 (§6.5.6)**.** *Define $\boldsymbol{M}_{\mathbf{x}} := \langle \Phi_{\mathbf{x}}^\delta, \Phi_{\mathbf{x}}^\delta \rangle_H$. When we have matrices $\boldsymbol{K}_j \eqsim (\mathbf{A}_{\mathbf{x}} + 2^j \mathbf{M}_{\mathbf{x}})^{-1}$ uniformly in $\delta \in \Delta$ and $j \in \mathbb{N}_0$, it follows that*

$$\mathbf{D}_X^{-1} \eqsim \boldsymbol{K}_X := \operatorname{blockdiag}[\boldsymbol{K}_{|\lambda|} \mathbf{A}_{\mathbf{x}} \boldsymbol{K}_{|\lambda|}]_{\lambda \in \vee_{\Sigma^\delta}}.$$

This yields an optimal preconditioner $K_X^\delta := \mathcal{F}_{\Sigma^\delta \otimes \Phi^\delta} \boldsymbol{K}_X (\mathcal{F}_{\Sigma^\delta \otimes \Phi^\delta})' \in \mathcal{L}\mathrm{is}(X^{\delta\prime}, X^\delta)$.

In [OR00] it was shown that under a 'full-regularity' assumption, for quasi-uniform meshes, a multiplicative multigrid method yields $\boldsymbol{K}_j$ satisfying the conditions of Thm. 8.3.1, which can moreover be applied in linear time.

8.3.3 Wavelets in time

The preconditioner $\boldsymbol{K}_X$ requires X_t^δ to be equipped with a *wavelet* basis Σ^δ, whereas one typically uses a different (single-scale) basis Φ_t^δ on X_t^δ. To bridge this gap, a basis transformation from Σ^δ to Φ_t^δ is required. We define the wavelet transform as $\boldsymbol{W}_t := (\mathcal{F}_{\Phi_t^\delta})^{-1}\mathcal{F}_{\Sigma^\delta}$.[1]

Define $V_j := \operatorname{span}\{\sigma_\lambda \in \Sigma : |\lambda| \le j\}$. Equip each V_j with a (single-scale) basis Φ_j, and assume that $\Phi_t^\delta := \Phi_J$ for some J, so that $X_t^\delta := V_J$. Since $V_{j+1} = V_j \oplus \operatorname{span}\Sigma_j$ where $\Sigma_j := \{\sigma_\lambda : |\lambda| = j\}$, there exist matrices $\boldsymbol{P}_j$ and $\boldsymbol{Q}_j$ such that $\Phi_j^\top = \Phi_{j+1}^\top \boldsymbol{P}_j$ and $\Psi_j^\top = \Phi_{j+1}^\top \boldsymbol{Q}_j$, with $\boldsymbol{M}_j := [\boldsymbol{P}_j|\boldsymbol{Q}_j]$ invertible.

Writing $v \in V_J$ in both forms $v = \boldsymbol{c}_0^\top \Phi_0 + \sum_{j=0}^{J-1} \boldsymbol{d}_j^\top \Psi_j$ and $v = \boldsymbol{c}_J^\top \Phi_J$, the basis transformation $\boldsymbol{W}_t := \boldsymbol{W}_J$ mapping wavelet coordinates $(\boldsymbol{c}_0^\top, \boldsymbol{d}_0^\top, \dots, \boldsymbol{d}_{J-1}^\top)$ to single-scale coordinates $\boldsymbol{c}_J$ satisfies

$$\boldsymbol{W}_J = \boldsymbol{M}_{J-1} \begin{bmatrix} \boldsymbol{W}_{J-1} & \boldsymbol{0} \\ \boldsymbol{0} & \mathrm{Id} \end{bmatrix}, \quad \text{and} \quad \boldsymbol{W}_0 := \mathrm{Id}. \tag{8.10}$$

Uniform locality of Σ implies *uniform sparsity* of the $\boldsymbol{M}_j$, i.e. with $\mathcal{O}(1)$ nonzeros per row and column. Then, assuming a geometrical increase in $\dim V_j$ in terms of j, which is true in the concrete setting below, matrix-vector products $\boldsymbol{x} \mapsto \boldsymbol{W}_t\boldsymbol{x}$ can be performed (serially) in linear complexity; cf. [Ste03b].

8.3.4 Solving the system

The matrix representation of S^δ and f^δ from (8.8) w.r.t. a basis $\Phi_t^\delta \otimes \Phi_{\mathbf{x}}^\delta$ of X^δ is

$$\boldsymbol{S} := (\mathcal{F}_{\Phi_t^\delta \otimes \Phi_{\mathbf{x}}^\delta})' S^\delta \mathcal{F}_{\Phi_t^\delta \otimes \Phi_{\mathbf{x}}^\delta} \quad \text{and} \quad \boldsymbol{f} := (\mathcal{F}_{\Phi_t^\delta \otimes \Phi_{\mathbf{x}}^\delta})' f^\delta.$$

Envisioning an iterative solver, using §8.3.2 we have a preconditioner in terms of the wavelet-in-time basis $\Sigma^\delta \otimes \Phi_{\mathbf{x}}^\delta$, with which their matrix representation is

$$\hat{\boldsymbol{S}} := (\mathcal{F}_{\Sigma^\delta \otimes \Phi_{\mathbf{x}}^\delta})' S^\delta \mathcal{F}_{\pm^\delta \otimes \oplus_{\mathbf{x}}^\delta} \quad \text{and} \quad \hat{\boldsymbol{f}} := (\mathcal{F}_{\Sigma^\delta \otimes \Phi_{\mathbf{x}}^\delta})' f^\delta. \tag{8.11}$$

These two forms are related: with the wavelet transform $\boldsymbol{W} := \boldsymbol{W}_t \otimes \mathrm{Id}_{\mathbf{x}}$, we have $\hat{\boldsymbol{S}} = \boldsymbol{W}^\top \boldsymbol{S}\boldsymbol{W}$ and $\hat{\boldsymbol{f}} = \boldsymbol{W}^\top \boldsymbol{f}$, and the matrix representation of (8.8) becomes

$$\text{finding } \boldsymbol{w} \quad \text{s.t.} \quad \hat{\boldsymbol{S}}\boldsymbol{w} = \hat{\boldsymbol{f}}. \tag{8.12}$$

[1] In literature, this transform is typically called an *inverse wavelet transform*.

We can then recover the solution in single-scale coordinates as $\boldsymbol{u} = \boldsymbol{W}\boldsymbol{w}$.

We use Preconditioned Conjugate Gradients (PCG), with preconditioner $\boldsymbol{K}_X$, to solve (8.12). Given an algebraic error tolerance $\varepsilon > 0$ and current guess $\boldsymbol{w}_k$, we monitor $\boldsymbol{r}_k^\top \boldsymbol{K}_X \boldsymbol{r}_k \leq \varepsilon^2$ where $\boldsymbol{r}_k := \hat{\boldsymbol{f}} - \hat{\boldsymbol{S}}\boldsymbol{w}_k$. This data is available within PCG, and constitutes a stopping criterium: with $u_k^\delta := \mathcal{F}_{\Sigma^\delta \otimes \Phi_{\mathbf{x}}^\delta} \boldsymbol{w}_k \in X^\delta$, we see

$$\boldsymbol{r}_k^\top \boldsymbol{K}_X \boldsymbol{r}_k = (f^\delta - S^\delta u_k^\delta)(K_X^\delta (f^\delta - S^\delta u_k^\delta)) \eqsim \|u^\delta - u_k^\delta\|_X^2 \tag{8.13}$$

where $\eqsim$ follows from (6.34), so that the algebraic error satisfies $\|u^\delta - u_k^\delta\|_X \lesssim \varepsilon$.

8.4 A concrete setting: the reaction-diffusion equation

On a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$, take $H := L_2(\Omega)$, $V := H_0^1(\Omega)$, and

$$a(t; \eta, \zeta) := \int_\Omega \boldsymbol{D}\nabla\eta \cdot \nabla\zeta + c\eta\zeta \mathrm{d}\mathbf{x}$$

where $\boldsymbol{D} = \boldsymbol{D}^\top \in \mathbb{R}^{d\times d}$ is positive definite, and $c \geq 0$.[2] We note that $A(t)$ is symmetric and coercive. W.l.o.g. we take $I := (0,1)$, i.e. $T := 1$.

Fix $p_t, p_{\mathbf{x}} \in \mathbb{N}$. With $\{\mathcal{T}_I\}$ the family of quasi-uniform partitions of I into subintervals, and $\{\mathcal{T}_\Omega\}$ that of conforming quasi-uniform triangulations of Ω, we define Δ as the collection of pairs $(\mathcal{T}_I, \mathcal{I}_\Omega)$. We construct our trial- and test spaces as

$$X^\delta := X_t^\delta \otimes X_{\mathbf{x}}^\delta, \quad Y^\delta := Y_t^\delta \otimes X_{\mathbf{x}}^\delta,$$

where, with $\mathbb{P}_p^{-1}(\mathcal{T})$ denoting the space of piecewise degree-p polynomials on $\mathcal{T}$,

$$X_t^\delta := H^1(I) \cap \mathbb{P}_{p_t}^{-1}(\mathcal{T}_I), \quad X_{\mathbf{x}}^\delta := H_0^1(\Omega) \cap \mathbb{P}_{p_{\mathbf{x}}}^{-1}(\mathcal{T}_\Omega), \quad Y_t^\delta := \mathbb{P}_{p_t}^{-1}(\mathcal{T}_I).$$

These spaces satisfy condition (8.5), with coinciding spatial discretizations on X^δ and Y^δ. For this choice of Δ, inf-sup condition (8.6) follows from [SW21b, Thm. 4.3].

For X_t^δ, we choose Φ_t^δ to be the Lagrange basis of degree p_t on $\mathcal{T}_I$; for $X_{\mathbf{x}}^\delta$, we choose $\Phi_{\mathbf{x}}^\delta$ to be that of degree $p_{\mathbf{x}}$ on $\mathcal{T}_\Omega$. An orthogonal basis Ξ^δ for Y_t^δ may be built as piecewise shifted Legendre polynomials of degree p_t w.r.t. $\mathcal{T}_I$.

For $p_t = 1$, one finds a suitable wavelet basis Σ in [Ste98]. For $p_t > 1$, one can either split the system into lowest- and higher-order parts and perform the transform on the lowest-order part only, or construct higher-order wavelets directly; cf. [Dij09].

Owing to the tensor-product structure of X^δ and Y^δ and of the operators A and ∂_t, the matrix representation of our formulation becomes remarkably simple.

[2] This is easily generalized to variable coefficients, but notation becomes more obtuse.

Lemma 8.4.1. *Define* $\boldsymbol{g} := (\mathcal{F}_{\Xi^\delta \otimes \Phi^\delta_{\mathbf{x}}})' g$, $\boldsymbol{u}_0 := \Phi^\delta_t(0) \otimes \langle u_0, \Phi^\delta_{\mathbf{x}} \rangle_{L_2(\Omega)}$, *and*

$$\begin{aligned} \boldsymbol{T} &:= \langle \tfrac{\mathrm{d}}{\mathrm{d}t} \Phi^\delta_t, \Xi^\delta \rangle_{L_2(I)}, & \boldsymbol{N} &:= \langle \Phi^\delta_t, \Xi^\delta \rangle_{L_2(I)}, \\ \boldsymbol{\Gamma}_0 &:= \Phi^\delta_t(0)[\Phi^\delta_t(0)]^\top, & \boldsymbol{M}_{\mathbf{x}} &:= \langle \Phi^\delta_{\mathbf{x}}, \Phi^\delta_{\mathbf{x}} \rangle_{L_2(\Omega)}, \\ \boldsymbol{A}_{\mathbf{x}} &:= \langle \boldsymbol{D} \nabla \Phi^\delta_{\mathbf{x}}, \nabla \Phi^\delta_{\mathbf{x}} \rangle_{L_2(\Omega)} + c\boldsymbol{M}_{\mathbf{x}}, & \boldsymbol{B} &:= \boldsymbol{T} \otimes \boldsymbol{M}_{\mathbf{x}} + \boldsymbol{N} \otimes \boldsymbol{A}_{\mathbf{x}}. \end{aligned}$$

With $\boldsymbol{K}_Y := \boldsymbol{O}^{-1} \otimes \boldsymbol{K}_{\mathbf{x}}$ *from §8.3.1, we can write* $\boldsymbol{S}$ *and* $\boldsymbol{f}$ *from §8.3.4 as*

$$\boldsymbol{S} = \boldsymbol{B}^\top \boldsymbol{K}_Y \boldsymbol{B} + \boldsymbol{\Gamma}_0 \otimes \boldsymbol{M}_{\mathbf{x}}, \quad \boldsymbol{f} = \boldsymbol{B}^\top \boldsymbol{K}_Y \boldsymbol{g} + \boldsymbol{u}_0.$$

Note that $\boldsymbol{N}$ *and* $\boldsymbol{T}$ *are non-square,* $\boldsymbol{\Gamma}_0$ *is very sparse, and* $\boldsymbol{T}$ *is bidiagonal.*

In fact, assumption (8.5) allows us to write $\boldsymbol{S}$ in an even simpler form.

Lemma 8.4.2. *The matrix* $\boldsymbol{S}$ *can be written as*

$$\begin{aligned} \boldsymbol{S} = \boldsymbol{A}_t \otimes (\boldsymbol{M}_{\mathbf{x}} \boldsymbol{K}_{\mathbf{x}} \boldsymbol{M}_{\mathbf{x}}) + \boldsymbol{M}_t \otimes (\boldsymbol{A}_{\mathbf{x}} \boldsymbol{K}_{\mathbf{x}} \boldsymbol{A}_{\mathbf{x}}) + \boldsymbol{L}^\top \otimes (\boldsymbol{M}_{\mathbf{x}} \boldsymbol{K}_{\mathbf{x}} \boldsymbol{A}_{\mathbf{x}}) \\ + \boldsymbol{L} \otimes (\boldsymbol{A}_{\mathbf{x}} \boldsymbol{K}_{\mathbf{x}} \boldsymbol{M}_{\mathbf{x}}) + \boldsymbol{\Gamma}_0 \otimes \boldsymbol{M}_{\mathbf{x}} \end{aligned}$$

where

$$\boldsymbol{L} := \langle \tfrac{\mathrm{d}}{\mathrm{d}t} \Phi^\delta_t, \Phi^\delta_t \rangle_{L_2(I)}, \quad \boldsymbol{M}_t := \langle \Phi^\delta_t, \Phi^\delta_t \rangle_{L_2(I)}, \quad \boldsymbol{A}_t := \langle \tfrac{\mathrm{d}}{\mathrm{d}t} \Phi^\delta_t, \tfrac{\mathrm{d}}{\mathrm{d}t} \Phi^\delta_t \rangle_{L_2(I)}.$$

This matrix representation does not depend on Y^δ_t *or* Ξ^δ *at all.*

Proof. The expansion of $\boldsymbol{B} := \boldsymbol{T} \otimes \boldsymbol{M}_{\mathbf{x}} + \boldsymbol{N} \otimes \boldsymbol{A}_{\mathbf{x}}$ in $\boldsymbol{S}$ yields a sum of five Kronecker products, one of which is

$$(\boldsymbol{T}^\top \otimes \boldsymbol{M}_{\mathbf{x}}) \boldsymbol{K}_Y (\boldsymbol{T} \otimes \boldsymbol{A}_{\mathbf{x}}) = (\boldsymbol{T}^\top \boldsymbol{O}^{-1} \boldsymbol{N}) \otimes (\boldsymbol{M}_{\mathbf{x}} \boldsymbol{K}_{\mathbf{x}} \boldsymbol{A}_{\mathbf{x}}).$$

We will show that $\boldsymbol{T}^\top \boldsymbol{O}^{-1} \boldsymbol{N} = \boldsymbol{L}^\top$; similar arguments hold for the other terms. Thanks to $X^\delta_t \subset Y^\delta_t$, we can define the trivial embedding $F^\delta_t : X^\delta_t \to Y^\delta_t$. Defining

$$\begin{aligned} T^\delta &\colon X^\delta_t \to Y^{\delta\prime}_t, \quad (T^\delta u)(v) := \langle \tfrac{\mathrm{d}}{\mathrm{d}t} u, v \rangle_{L_2(I)}, \\ M^\delta &\colon Y^\delta_t \to Y^{\delta\prime}_t, \quad (M^\delta u)(v) := \langle u, v \rangle_{L_2(I)}, \end{aligned}$$

we find $\boldsymbol{O} = (\mathcal{F}_{\Xi^\delta})' M^\delta \mathcal{F}_{\Xi^\delta}$, $\boldsymbol{N} = (\mathcal{F}_{\Xi^\delta})' M^\delta \mathcal{F}^\delta_t \mathcal{F}_{\Phi^\delta_t}$ and $\boldsymbol{T} = (\mathcal{F}_{\Xi^\delta})' T^\delta \mathcal{F}_{\Phi^\delta_t}$, so

$$\boldsymbol{T}^\top \boldsymbol{O}^{-1} \boldsymbol{N} = (\mathcal{F}_{\Phi^\delta_t})' T^{\delta\prime} F^\delta_t \mathcal{F}_{\Phi^\delta_t} = \langle \Phi_t, \tfrac{\mathrm{d}}{\mathrm{d}t} \Phi_t \rangle_{L_2(I)} = \boldsymbol{L}^\top. \qquad \square$$

8.4.1 Parallel complexity

The *parallel complexity* of our algorithm is the asymptotic runtime of solving (8.12) for $\boldsymbol{u} \in \mathbb{R}^{N_t N_{\mathbf{x}}}$ in terms of $N_t := \dim X_t^\delta$ and $N_{\mathbf{x}} := \dim X_{\mathbf{x}}^\delta$, given sufficiently many parallel processors and assuming no communication cost.

We understand the serial (resp. parallel) cost of a matrix $\boldsymbol{B}$, denoted $C_{\boldsymbol{B}}^s$ (resp. $C_{\boldsymbol{B}}^p$), as the asymptotic runtime of performing $\boldsymbol{x} \mapsto \boldsymbol{B}\boldsymbol{x} \in \mathbb{R}^N$ in terms of N, on a single (resp. sufficiently many) processors at no communication cost. For *uniformly sparse* matrices, i.e. with $\mathcal{O}(1)$ nonzeros per row and column, the serial cost is $\mathcal{O}(N)$, and the parallel cost is $\mathcal{O}(1)$ by computing each cell of the output concurrently.

From Theorem 8.3.1, we see that $\boldsymbol{K}_X$ is such that $\kappa_2(\boldsymbol{K}_X \hat{\boldsymbol{S}}) \lesssim 1$ uniformly in $\delta \in \Delta$. Therefore, for a given algebraic error tolerance ε, we require $\mathcal{O}(\log \varepsilon^{-1})$ PCG iterations. Assuming that the parallel cost of matrices dominates that of vector addition and inner products, the parallel complexity of a single PCG iteration is dominated by the cost of applying $\boldsymbol{K}_X$ and $\hat{\boldsymbol{S}}$. As $\hat{\boldsymbol{S}} = \boldsymbol{W}^\top \boldsymbol{S} \boldsymbol{W}$, our algorithm runs in complexity

$$\mathcal{O}(\log \varepsilon^{-1}[C_{\boldsymbol{K}_X}^\circ + C_{\boldsymbol{W}^\top}^\circ + C_{\boldsymbol{S}}^\circ + C_{\boldsymbol{W}}^\circ]) \quad (\circ \in \{s, p\}). \tag{8.14}$$

Theorem 8.4.3. *For fixed algebraic error tolerance $\varepsilon > 0$, our algorithm runs in*

- *serial complexity $\mathcal{O}(N_t N_{\mathbf{x}})$;*
- *time-parallel complexity $\mathcal{O}(\log(N_t) N_{\mathbf{x}})$;*
- *space-time-parallel complexity $\mathcal{O}(\log(\mathcal{N}_\sqcup \mathcal{N}_{\mathbf{x}}))$.*

Proof. We absorb the constant factor $\log \varepsilon^{-1}$ of (8.14) into $\mathcal{O}$. We analyse the cost of every matrix separately.

The (inverse) wavelet transform

As $\boldsymbol{W} = \boldsymbol{W}_t \otimes \mathrm{Id}_{\mathbf{x}}$, its serial cost equals $\mathcal{O}(C_{\boldsymbol{W}_t}^s N_{\mathbf{x}})$. The choice of wavelet allows performing $\boldsymbol{x} \mapsto \boldsymbol{W}_t \boldsymbol{x}$ at linear serial cost (cf. §8.3.3), so that $C_{\boldsymbol{W}}^s = \mathcal{O}(N_t N_{\mathbf{x}})$.

Using (8.10), we write $\boldsymbol{W}_t$ as the composition of J matrices, each uniformly sparse and hence at parallel cost $\mathcal{O}(1)$. Because the mesh in time is quasi-uniform, we have $J \eqsim \log N_t$. We find that $C_{\boldsymbol{W}_t}^p = \mathcal{O}(J) = \mathcal{O}(\log N_t)$, so that the time-parallel cost of $\boldsymbol{W}$ equals $\mathcal{O}(\log(N_t) N_{\mathbf{x}})$. By exploiting spatial parallelism as well, we find $C_{\boldsymbol{W}}^p = \mathcal{O}(\log N_t)$. Analogous arguments hold for $\boldsymbol{W}_t^\top$ and $\boldsymbol{W}^\top$.

The preconditioner

Recall that $\boldsymbol{K}_X := \mathrm{blockdiag}[\boldsymbol{K}_{|\lambda|}\mathbf{A}_\mathbf{x}\boldsymbol{K}_{|\lambda|}]_\lambda$. Since the cost of $\boldsymbol{K}_j$ is independent of j, we see that

$$C^s_{\boldsymbol{K}_X} = \mathcal{O}\big(N_t \cdot (2C^s_{\boldsymbol{K}_j} + C^s_{\boldsymbol{A}_\mathbf{x}})\big) = \mathcal{O}(2N_t C^s_{\boldsymbol{K}_j} + N_t N_\mathbf{x}).$$

Implementing the $\boldsymbol{K}_j$ as typical multiplicative multigrid solvers with linear serial cost, we find $C^s_{\boldsymbol{K}_X} = \mathcal{O}(N_t N_\mathbf{x})$.

Through temporal parallelism, we can apply each block of $\boldsymbol{K}_X$ concurrently, resulting in a time-parallel cost of $\mathcal{O}(2C^s_{\boldsymbol{K}_j} + C^s_{\boldsymbol{A}_\mathbf{x}}) = \mathcal{O}(N_\mathbf{x})$.

By parallelizing in space as well, we reduce the cost of the uniformly sparse $\boldsymbol{A}_\mathbf{x}$ to $\mathcal{O}(1)$. The parallel cost of multiplicative multigrid on quasi-uniform triangulations is $\mathcal{O}(\log N_\mathbf{x})$; cf. [MFL+91]. It follows that $C^p_{\boldsymbol{K}_X} = \mathcal{O}(\log N_\mathbf{x})$.

The Schur matrix

Using Lemma 8.4.1, we write $\boldsymbol{S} = \boldsymbol{B}^\top \boldsymbol{K}_Y \boldsymbol{B} + \boldsymbol{\Gamma}_0 \otimes \boldsymbol{M}_\mathbf{x}$ where $\boldsymbol{B} = \boldsymbol{T} \otimes \boldsymbol{M}_\mathbf{x} + \boldsymbol{N} \otimes \boldsymbol{A}_\mathbf{x}$, which immediately reveals that

$$\begin{aligned} C^s_{\boldsymbol{S}} &= C^s_{\boldsymbol{B}^\top} + C^s_{\boldsymbol{K}_Y} + C^s_{\boldsymbol{B}} + C^s_{\boldsymbol{\Gamma}_0} \cdot C^s_{\boldsymbol{M}} = \mathcal{O}(N_t N_\mathbf{x} + C^s_{\boldsymbol{K}_Y}), \quad \text{and} \\ C^p_{\boldsymbol{S}} &= \max\big\{C^p_{\boldsymbol{B}^\top} + C^p_{\boldsymbol{K}_Y} + C^p_{\boldsymbol{B}},\ C^p_{\boldsymbol{\Gamma}_0} \cdot C^p_{\boldsymbol{M}}\big\} = \mathcal{O}(C^p_{\boldsymbol{K}_Y}) \end{aligned}$$

because every matrix except $\boldsymbol{K}_Y$ is uniformly sparse. With arguments similar to the previous paragraph, we see that $\boldsymbol{K}_Y$ (and hence $\boldsymbol{S}$) has serial cost $\mathcal{O}(N_t N_\mathbf{x})$, time-parallel cost $\mathcal{O}(N_\mathbf{x})$, and space-time-parallel cost $\mathcal{O}(\log N_\mathbf{x})$. □

8.4.2 Solving to higher accuracy

Instead of *fixing* the algebraic error tolerance, maybe more realistic is is to desire a solution $\tilde{u}^\delta \in X^\delta$ for which the error is proportional to the discretization error, i.e. $\|u - \tilde{u}^\delta\|_X \lesssim \inf_{u_\delta \in X^\delta} \|u - u_\delta\|_X$.

Assuming that this error decays with a (problem-dependent) rate $s > 0$, i.e. $\inf_{u_\delta \in X^\delta} \|u - u_\delta\|_X \lesssim (N_t N_\mathbf{x})^{-s}$, then the same holds for the solution u^δ of (8.8); cf. Thm. 8.2.3. When the algebraic error tolerance decays as $\varepsilon \lesssim (N_t N_\mathbf{x})^{-s}$, a triangle inequality and (8.13) show that the error of our solution $\tilde{u}^\delta$ obtained by PCG decays at rate s too.

In this case, $\log \epsilon^{-1} = \mathcal{O}(\log(N_t N_\mathbf{x}))$. From (8.14) and the proof of Theorem 8.4.3, we find our algorithm to run in superlinear serial complexity $\mathcal{O}(N_t N_\mathbf{x} \log(N_t N_\mathbf{x}))$, time-parallel complexity $\mathcal{O}(\log^2(N_t) \log(N_\mathbf{x}) N_\mathbf{x})$, and polylogarithmic complexity $\mathcal{O}(\log^2(N_t N_\mathbf{x}))$ parallel in space and time.

For elliptic PDEs, algorithms are available that offer quasi-optimal solutions, serially in linear complexity $\mathcal{O}(N_\mathbf{x})$—the cost of a serial solve to *fixed* algebraic error—and in parallel in $\mathcal{O}(\log^2 N_\mathbf{x})$, by combining a *nested iteration* with parallel multigrid; cf. [Hac85, Ch. 5] and [Bra81].

In [HVW95], the question is posed whether "good serial algorithms for parabolic PDEs are intrinsically as parallel as good serial algorithms for elliptic PDEs", basically asking if the lower bound of $\mathcal{O}(\log^2(N_t N_{\mathbf{x}}))$ can be attained by an algorithm that runs serially in $\mathcal{O}(N_t N_{\mathbf{x}})$; see [Wor91, §2.2] for a formal discussion.

Nested iteration drives down the serial complexity of our algorithm to a linear $\mathcal{O}(N_t N_{\mathbf{x}})$, and also improves the time-parallel complexity to $\mathcal{O}(\log(N_t) N_{\mathbf{x}})$.[3] This is on par with the best-known results for elliptic problems, so we answer the question posed in [HVW95] in the affirmative.

8.5 Numerical experiments

We take the simple heat equation, i.e. $D = \mathrm{Id}_{\mathbf{x}}$ and $c = 0$. We select $p_t = p_{\mathbf{x}} = 1$, i.e. lowest order finite elements in space and time. We will use the 3-point wavelet introduced in [Ste98].

We implemented our algorithm in Python using the open source finite element library `NGSolve` [Sch14] for meshing and discretization of the bilinear forms in space and time, `MPI` through `mpi4py` [DPS05] for distributed computations, and `SciPy` [Vir20] for the sparse matrix-vector computations. The source code is available at [vVW21c].

8.5.1 Preconditioner calibration on a 2D problem

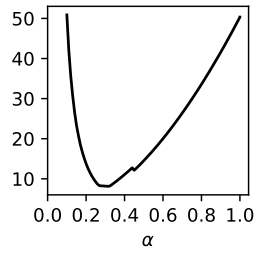
Our preconditioner is optimal, meaning that $\kappa_2(\boldsymbol{K}_X \hat{\boldsymbol{S}}) \lesssim 1$. Here we will investigate this condition number quantitatively.

As a model problem, we partition the temporal interval I uniformly into 2^J subintervals. We consider the domain $\Omega := [0,1]^2$, and triangulate it uniformly into 4^K triangles. We set $N_t := \dim X_t^\delta = 2^J + 1$ and $N_{\mathbf{x}} := \dim X_{\mathbf{x}}^\delta = (2^K - 1)^2$.

We start by using direct inverses $\boldsymbol{K}_j = (\boldsymbol{A}_{\mathbf{x}} + 2^j \boldsymbol{M}_{\mathbf{x}})^{-1}$ and $\boldsymbol{K}_{\mathbf{x}} = \boldsymbol{A}_{\mathbf{x}}^{-1}$ to determine the best possible condition numbers. We found that replacing $\boldsymbol{K}_j$ by $\boldsymbol{K}_j^\alpha = (\alpha \boldsymbol{A}_{\mathbf{x}} + 2^j \boldsymbol{M}_{\mathbf{x}})^{-1}$ for $\alpha = 0.3$ gave better conditioning; see also the left of Table 8.1. At the right of Table 8.1, we see that the condition numbers are very robust with respect to spatial refinements, but less so for refinements in time. Still, at $N_t = 16\,129$, we observe a modest $\kappa_2(\boldsymbol{K}_X \hat{\boldsymbol{S}})$ of 8.74.

Replacing the direct inverses with multigrid solvers, we found a good balance between speed and conditioning at 2 V-cycles with 3 Gauss-Seidel smoothing steps per grid. We decided to use these for our experiments.

[3] Interestingly, nested iteration offers no improvements parallel in space *and* time, with complexity still $\mathcal{O}(\log^2(N_t N_{\mathbf{x}}))$.



	$N_t = 65$	129	257	513	1 025	2 049	4 097	8 193
$N_{\mathbf{x}} = 49$	6.34	7.05	7.53	7.89	8.15	8.37	8.60	8.78
225	6.33	6.89	7.55	7.91	8.14	8.38	8.57	8.73
961	6.14	6.89	7.55	7.93	**8.15**	8.38	8.57	8.74
3 969	6.14	7.07	7.56	7.87	8.16	8.38	8.57	8.74
16 129	6.14	6.52	7.55	7.86	8.16	8.37	8.57	8.74

TABLE 8.1. Computed condition numbers $\kappa_2(\boldsymbol{K}_X \hat{\boldsymbol{S}})$. Left: fixed $N_t = 1025$, $N_{\mathbf{x}} = 961$ for varying α. Right: fixed $\alpha = 0.3$ for varying N_t and $N_{\mathbf{x}}$.

8.5.2 Time-parallel results

We perform computations on Cartesius, the Dutch supercomputer. Each Cartesius node has 64GB of memory and 12 cores (at 2 threads per core) running at 2.6GHz. Using the preconditioner detailed above, we iterate PCG on (8.12) with $\boldsymbol{S}$ computed as in Lemma 8.4.2, until achieving an algebraic error of $\varepsilon = 10^{-6}$; see also §8.3.4. For the spatial multigrid solvers, we use 2 V-cycles with 3 Gauss-Seidel smoothing steps per grid.

Memory-efficient time-parallel implementation

For $\boldsymbol{X} \in \mathbb{R}^{N_{\mathbf{x}} \times N_t}$, we define $(\boldsymbol{X}) \in \mathbb{R}^{N_t N_{\mathbf{x}}}$ as the vector obtained by stacking columns of $\boldsymbol{X}$ vertically. For memory efficency, we do not build matrices of the form $\boldsymbol{B}_t \otimes \boldsymbol{B}_{\mathbf{x}}$ appearing in Lemma 8.4.2 directly, but instead perform matrix-vector products using the identity

$$(\boldsymbol{B}_t \otimes \boldsymbol{B}_{\mathbf{x}})(\boldsymbol{X}) = (\boldsymbol{B}_{\mathbf{x}}(\boldsymbol{B}_t \boldsymbol{X}^\top)^\top) = (\mathrm{Id}_t \otimes \boldsymbol{B}_{\mathbf{x}})(\boldsymbol{B}_t \boldsymbol{X}^\top). \tag{8.15}$$

Each parallel processor stores only a subset of the temporal degrees of freedom, e.g. a subset of columns of $\boldsymbol{X}$. When $\boldsymbol{B}_t$ is uniformly sparse, which holds true for all of our temporal matrices, using (8.15) we can evaluate $(\boldsymbol{B}_t \otimes \boldsymbol{B}_{\mathbf{x}})(\boldsymbol{X})$ in $\mathcal{O}(C^s_{\boldsymbol{B}_{\mathbf{x}}})$ operations parallel in time: on each parallel processor, we compute 'our' columns of $\boldsymbol{Y} := \boldsymbol{B}_t \boldsymbol{X}^\top$ by receiving the necessary columns of $\boldsymbol{X}$ from neighbouring processors, and then compute $\boldsymbol{B}_{\mathbf{x}} \boldsymbol{Y}^\top$ without communication.

The preconditioner $\boldsymbol{K}_X$ is block-diagonal, making its time-parallel application trivial. Representing the wavelet transform of §8.3.3 as the composition of J Kronecker products allows a time-parallel implementation using the above.

2D problem

We select $\Omega := [0,1]^2$ with a uniform triangulation $\mathcal{T}_\Omega$, and we triangulate I uniformly into $\mathcal{T}_I$. We prescribe the smooth solution $u(t,x,y) := \exp(-2\pi^2 t)\sin(\pi x)\sin(\pi y)$, so the problem has vanishing forcing data g.

Table 8.2 details the strong scaling results, i.e. fixing the problem size and increasing the number of processors P. We triangulate I into 2^{14} time slabs, yielding $N_t = 16\,385$ temporal degrees of freedom, and Ω into 4^8 triangles, yielding a $X_{\mathbf{x}}^{\delta}$ of dimension $N_{\mathbf{x}} = 65\,025$. The resulting system contains $1\,065\,434\,625$ degrees of freedom and our solver reaches the algebraic error tolerance after 16 iterations. In perfect strong scaling, the total number of CPU-hours remains constant. Even at 2 048 processors, we observe a parallel efficiency of around 92.9%, solving this system in a modest 11.7 CPU-hours. Acquiring strong scaling results on a single node was not possible due to memory limitations.

Table 8.3 details the weak scaling results, i.e. fixing the problem size per processor and increasing the number of processors. In perfect weak scaling, the time per iteration should remain constant. We observe a slight increase in time per iteration on a single node, but when scaling to multiple nodes, we observe a near-perfect parallel efficiency of around 96.7%, solving the final system with $4\,278\,467\,585$ degrees of freedom in a mere 109 seconds.

3D problem

We select $\Omega := [0,1]^3$ with $u(t,x,y,z) := \exp(-3\pi^2 t)\sin(\pi x)\sin(\pi y)\sin(\pi z)$, so the problem has vanishing forcing data g.

Table 8.4 shows the strong scaling results. We triangulate I uniformly into 2^{14} time slabs, and Ω uniformly into 8^6 tetrahedra. The arising system has $N = 4\,097\,020\,095$ unknowns, which we solve to tolerance in 18 iterations. The results are comparable to those in two dimensions, albeit a factor two slower at similar problem sizes.

Table 8.5 shows the weak scaling results for the 3D problem. As in the two-dimensional case, we observe excellent scaling properties, and see that the time per iteration is nearly constant.

8.6 Conclusion

We have presented a framework for solving linear parabolic evolution equations massively in parallel. Based on earlier ideas [And16, NS19, SW21b], we found a remarkably simple symmetric Schur-complement equation. With a tensor-product discretization of the space-time cylinder using standard finite elements in time and space together with a wavelet-in-time multigrid-in-space preconditioner, we were able to solve the arising systems to fixed accuracy in a uniformly bounded number of PCG steps.

We found that our algorithm runs in linear complexity on a single processor. Moreover, when *sufficiently many* parallel processors are available and communication is free, its runtime scales *logarithmically* in the discretization size. These complexity results translate to a highly efficient algorithm in practice.

P	N_t	$N_{\mathbf{x}}$	$N = N_t N_{\mathbf{x}}$	its	time (s)	time/it (s)	CPU-hrs
1–16	16 385	65 025	1 065 434 625	——— out of memory ———			
32	16 385	65 025	1 065 434 625	16	1224.85	76.55	10.89
64	16 385	65 025	1 065 434 625	16	615.73	38.48	10.95
128	16 385	65 025	1 065 434 625	16	309.81	19.36	11.02
256	16 385	65 025	1 065 434 625	16	163.20	10.20	11.61
512	16 385	65 025	1 065 434 625	16	96.54	6.03	13.73
512	16 385	65 025	1 065 434 625	16	96.50	6.03	13.72
1 024	16 385	65 025	1 065 434 625	16	45.27	2.83	12.88
2 048	16 385	65 025	1 065 434 625	16	20.59	1.29	11.72

TABLE 8.2. Strong scaling results for the 2D problem.

	P	N_t	$N_{\mathbf{x}}$	$N = N_t N_{\mathbf{x}}$	its	time (s)	time/it (s)	CPU-hrs
single node	1	9	261 121	2 350 089	8	33.36	4.17	0.01
	2	17	261 121	4 439 057	11	46.66	4.24	0.03
	4	33	261 121	8 616 993	12	54.60	4.55	0.06
	8	65	261 121	16 972 865	13	65.52	5.04	0.15
	16	129	261 121	33 684 609	13	86.94	6.69	0.39
multiple nodes	32	257	261 121	67 108 097	14	93.56	6.68	0.83
	64	513	261 121	133 955 073	14	94.45	6.75	1.68
	128	1 025	261 121	267 649 025	14	93.85	6.70	3.34
	256	2 049	261 121	535 036 929	15	101.81	6.79	7.24
	512	4 097	261 121	1 069 812 737	15	101.71	6.78	14.47
	1 024	8 193	261 121	2 139 364 353	16	108.32	6.77	30.81
	2 048	16 385	261 121	4 278 467 585	16	109.59	6.85	62.34

TABLE 8.3. Weak scaling results for the 2D problem.

P	N_t	$N_{\mathbf{x}}$	$N = N_t N_{\mathbf{x}}$	its	time (s)	time/it (s)	CPU-hrs
1–64	16 385	250 047	4 097 020 095	——— out of memory ———			
128	16 385	250 047	4 097 020 095	18	3 308.49	174.13	117.64
256	16 385	250 047	4 097 020 095	18	1 655.92	87.15	117.75
512	16 385	250 047	4 097 020 095	18	895.01	47.11	127.29
1 024	16 385	250 047	4 097 020 095	18	451.59	23.77	128.45
2 048	16 385	250 047	4 097 020 095	18	221.12	12.28	125.80

TABLE 8.4. Strong scaling results for the 3D problem.

P	N_t	$N_{\mathbf{x}}$	$N = N_t N_{\mathbf{x}}$	its	time (s)	time/it (s)	CPU-hrs
16	129	250 047	32 256 063	15	183.65	12.24	0.82
32	257	250 047	64 262 079	16	196.26	12.27	1.74
64	513	250 047	128 274 111	16	197.55	12.35	3.51
128	1 025	250 047	256 298 175	17	210.21	12.37	7.47
256	2 049	250 047	512 346 303	17	209.56	12.33	14.90
512	4 097	250 047	1 024 442 559	17	210.14	12.36	29.89
1 024	8 193	250 047	2 048 635 071	18	221.77	12.32	63.08
2 048	16 385	250 047	4 097 020 095	18	221.12	12.28	125.80

Table 8.5. Weak scaling results for the 3D problem.

The numerical experiments serve as a showcase for the described space-time method, and exhibit its excellent time-parallelism by solving a linear system with over 4 billion unknowns in just 109 seconds, using just over 2 thousand parallel processors. By incorporating spatial parallelism as well, we expect these results to scale well to much larger problems.

Although performed in the rather restrictive setting of the heat equation discretized using piecewise linear polynomials on uniform triangulations, the parallel framework already allows solving more general linear parabolic PDEs using polynomials of varying degree on locally refined (tensor-product) meshes. In this more general setting, we envision load balancing to become the main hurdle in achieving good scaling results.

9 Adaptive BEM for the heat equation

9.1 Introduction

This chapter is about the adaptive numerical approximation of the heat equation using a simultaneous space-time boundary element method (BEM). In the last years, there has been a growing interest in space-time BEM for the heat equation [CS13, MST14, MST15, HT18, CR19, DNS19, DZO$^+$19, Tau19, ZWOM21]. In contrast to the differential operator based variational formulation on the space-time cylinder, the variational formulation corresponding to space-time BEM is coercive [AN87, Cos90] so that the discretized version always has a unique solution regardless of the chosen trial space which is even quasi-optimal in the natural energy norm. Moreover, it is naturally applicable on unbounded domains and only requires a mesh of the lateral boundary of the space-time cylinder resulting in a dimension reduction. The potential disadvantage that discretizations lead to dense matrices due to the nonlocality of the boundary integral operators has been tackled, e.g., in [MST14, MST15, HT18] via the fast multipole method and $\mathcal{H}$-matrices.

Two often mentioned advantages of simultaneous space-time methods are their potential for massive parallelization as well as their potential for fully adaptive refinement to resolve singularities local in both space and time. While the first advantage has been investigated in, e.g., [DZO$^+$19, ZWOM21], the latter requires suitable *a posteriori* computable error estimators, which have not been developed yet for the heat equation. Indeed, concerning *a posteriori* error estimation as well as adaptive refinement for BEM for time-dependent problems, we are only aware of the works [Glä12, GÖSS20] for the wave equation in two and three space dimensions, respectively.

In the present manuscript, we generalize the results [Fae00, Fae02] from Faermann for stationary PDEs to the heat equation: Let $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, be a Lipschitz domain with boundary $\Gamma := \partial\Omega$ and $T > 0$ a given end time point with corresponding time interval $I := (0, T)$. We abbreviate the space-time cylinder $Q := I \times \Omega$ with lateral boundary $\Sigma := I \times \Gamma$ and corresponding outer

normal vector $\boldsymbol{n} \in \mathbb{R}^d$. With the heat kernel

$$G(t,\boldsymbol{x}) := \begin{cases} \frac{1}{(4\pi t)^{d/2}}\, e^{-\frac{|\boldsymbol{x}|^2}{4t}} & \text{for } (t,\boldsymbol{x}) \in (0,\infty)\times\mathbb{R}^d, \\ 0 & \text{else,} \end{cases}$$

and a given function $f : \Sigma \to \mathbb{R}$, we consider the boundary integral equation,

$$(\mathcal{V}\phi)(t,\boldsymbol{x}) := \int_\Sigma G(t-s,\boldsymbol{x}-\boldsymbol{y})\phi(t-s,\boldsymbol{x}-\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}s = f(t,\boldsymbol{x}), \tag{9.1}$$

for a.e. $(t,\boldsymbol{x}) \in \Sigma$. Here, $\mathcal{V}$ is the single-layer operator. For given initial condition $u_0 : \Omega \to \mathbb{R}$ and Dirichlet data $u_D : \Sigma \to \mathbb{R}$, such equations arise from the heat equation

$$\begin{aligned} \partial_t u - \Delta u &= 0 && \text{on } Q, \\ u &= u_D && \text{on } \Sigma, \\ u(0,\cdot) &= u_0 && \text{on } \Omega. \end{aligned} \tag{9.2}$$

Let $\mathcal{P}$ be a mesh of the space-time boundary Σ consisting of prismatic elements $J\times K$ with $J \subseteq \overline{I}$ and $K \subseteq \Gamma$, and let Φ be an associated approximation of ϕ. Typically, Φ is a piecewise polynomial with respect to $\mathcal{P}$. As $\mathcal{V}$ is an isomorphism from the dual space $H^{-1/2,-1/4}(\Sigma) := H^{1/2,1/4}(\Sigma)'$ to the anisotropic Sobolev space $H^{1/2,1/4}(\Sigma)$, the discretization error $\|\phi-\Phi\|_{H^{-1/2,-1/4}(\Sigma)}$ is equivalent to the norm of the residual $\|f-\mathcal{V}\Phi\|_{H^{1/2,1/4}(\Sigma)}$. We show that the residual norm can be localized up to weighted L_2-terms, i.e.,

$$\sum_{J\times K\in\mathcal{P}} \eta_{\mathcal{P}}(\Phi,J\times K)^2 \lesssim \|f-\mathcal{V}\Phi\|^2_{H^{1/2,1/4}(\Sigma)} \lesssim \sum_{J\times K\in\mathcal{P}} \eta_{\mathcal{P}}(\Phi,J\times K)^2 + \zeta_{\mathcal{P}}(\Phi,J\times K)^2,$$

where $\eta_{\mathcal{P}}(\Phi,J\times K)^2$ measures the $H^{1/2,1/4}$-seminorm of the residual in a neighborhood of $J\times K$ and $\zeta_{\mathcal{P}}(\Phi) := (\operatorname{diam}(K)^{-1}+|J|^{-1/2})\|f-\mathcal{V}\Phi\|^2_{L_2(J\times K)}$. The hidden constants depend only on the regularity of the of the meshes found by fixing either the temporal or the spatial coordinate in $\mathcal{P}$. In particular, we do not require any assumption on the relation between the spatial and temporal size of the mesh elements, making anisotropically refined meshes possible.

If the elements satisfy the scaling $|J| \eqsim \operatorname{diam}(K)^2$ and if Φ is the Galerkin approximation of ϕ in a discrete space $\mathcal{X}$ that contains at least all $\mathcal{P}$-piecewise constant functions, then we can additionally prove that

$$\zeta_{\mathcal{P}}(\Phi,J\times K) \lesssim \eta_{\mathcal{P}}(\Phi,J\times K).$$

Indeed, numerical experiments (with $d=2$) suggest that this is not the case in general: If the scaling condition is not enforced, we observe situations where the L_2-terms ζ do not decay under mesh-refinement.

That being said, the estimator η does not only behave efficiently but also reliably in all considered examples. Moreover, anisotropic refinement steered by the space- and time-components of the estimator always yield the optimal algebraic convergence rate of both the estimator and the error.

Outline

The remainder of this chapter is organized as follows: Section 9.2 summarizes the general principles of the space-time boundary element method for the heat equation. Section 9.3 recalls the localization argument of [Fae00, Fae02] and applies it to anisotropic Sobolev spaces (Theorem 9.3.3). This result is then invoked in Corollary 9.3.5 for the residual, resulting in efficient and reliable *a posteriori* computable error bounds. In particular, a Poincaré-type inequality (Lemma 9.3.4) allows to estimate the weighted L_2-terms that are still present in the upper bound from Theorem 9.3.3. Finally, Section 9.4 introduces an adaptive algorithm for $d = 2$ which is based on the derived error estimator. Different marking and refinement strategies are presented. The adaptive algorithm is subsequently applied to several concrete examples with typical singularities in space and time.

9.2 Preliminaries

9.2.1 General notation

Throughout and without any ambiguity, $|\cdot|$ denotes the absolute value of scalars, the Euclidean norm of vectors in $\mathbb{R}^n$, or the the measure of a set in $\mathbb{R}^n$, e.g., the length of an interval or the area of a surface in $\mathbb{R}^3$. We write $A \lesssim B$ to abbreviate $A \leq CB$ with some generic constant $C > 0$, which is clear from the context. Moreover, $A \eqsim B$ abbreviates $A \lesssim B \lesssim A$.

9.2.2 Anisotropic Sobolev spaces

For d-dimensional $\omega \subseteq \Omega$ or $(d-1)$-dimensional $\omega \subseteq \Gamma$, and $\mu \in (0,1]$, we first recall the Sobolev space $H^\mu(\omega) := \big\{v \in L_2(\omega) : \|v\|_{H^\mu(\omega)} < \infty\big\}$ associated with the Sobolev–Slobodeckij norm $\|v\|^2_{H^\mu(\omega)} := \|v\|^2_{L_2(\omega)} + |v|^2_{H^\mu(\omega)}$, with

$$|v|^2_{H^\mu(\omega)} := \begin{cases} \int_\omega \int_\omega \frac{|v(\boldsymbol{x})-v(\boldsymbol{y})|^2}{|\boldsymbol{x}-\boldsymbol{y}|^{\dim(\omega)+2\mu}} \,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{x} & \text{if } \mu \in (0,1), \\ \|\nabla_\omega v\|^2_{L_2(\omega)} & \text{if } \mu = 1, \end{cases}$$

where $\dim(\omega)$ denotes the dimension of ω, i.e., d or $d-1$, and ∇_ω denotes the (weak) gradient on ω, i.e., the standard gradient or the surface gradient.

Moreover, we define for any subinterval $J \subseteq \overline{I}$, $\nu \in (0,1]$, and any Banach space X,

$$H^\nu(J;X) := \big\{v \in L_2(J;X) : \|v\|_{H^\nu(J;X)} < \infty\big\}$$

associated with the norm $\|v\|^2_{H^\nu(J;X)} := \|v\|^2_{L_2(J;X)} + |v|^2_{H^\nu(J;X)}$, with

$$|v|^2_{H^\nu(J;X)} := \begin{cases} \int_J \int_J \frac{\|v(t)-v(s)\|^2_X}{|t-s|^{1+2\nu}} \,\mathrm{d}s\,\mathrm{d}t & \text{if } \nu \in (0,1), \\ \|\partial_t v\|^2_{L_2(\omega)} & \text{if } \nu = 1, \end{cases}$$

where ∂_t denotes the (weak) time derivative. If $X = \mathbb{R}$, we simply write $H^\nu(J)$, $\|v\|_{H^\nu(J)}$, and $|v|_{H^\nu(J)}$. Finally, we recall the anisotropic Sobolev space

$$H^{\mu,\nu}(J \times \omega) := L_2(J; H^\mu(\omega)) \cap H^\nu(J; L_2(\omega))$$

with corresponding norm

$$\|v\|^2_{H^{\mu,\nu}(J\times\omega)} := \|v\|^2_{L_2(J;H^\mu(\omega))} + \|v\|^2_{H^\nu(J;L_2(\omega))} \quad \big(v \in H^{\mu,\nu}(J \times \omega)\big).$$

We will sometimes use the abbreviation

$$|v|^2_{L_2(J;H^\mu(\omega))} := \int_J |v(t,\cdot)|^2_{H^\mu(\omega)} \,\mathrm{d}t \quad \big(v \in L_2(J; H^\mu(\omega))\big).$$

For $\omega \in \{\Omega, \Gamma\}$, we denote $H^{-\mu,-\nu}(I \times \omega)$ for the dual of $H^{\mu,\nu}(I \times \omega)$ with duality pairing $\langle \cdot\,,\, \cdot \rangle_{I\times\omega}$. We view $L_2(I \times \omega)$ as subspace of $H^{-\mu,-\nu}(I \times \omega)$ via

$$\langle v\,,\,\psi\rangle_{I\times\omega} := \int_I \int_\omega v(t,\boldsymbol{x})\psi(t,\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \quad \big(v \in H^{\mu,\nu}(I \times \omega), \psi \in L_2(I \times \omega)\big).$$

9.2.3 Boundary integral equations

It is well-known that for $u_0 \in L^2(\Omega)$ and $u_D \in H^{1/2,1/4}(\Sigma)$, the heat equation (9.2) admits a unique solution $u \in H^{1,1/2}(Q)$. With the normal derivative $\phi_N := \partial_{\boldsymbol{n}} u \in H^{-1/2,-1/4}(\Sigma)$, u satisfies the representation formula

$$u = \widetilde{\mathscr{M}_0} u_0 + \widetilde{\mathscr{V}} \phi_N - \widetilde{\mathscr{K}} u_D, \tag{9.3}$$

where

$$(\widetilde{\mathscr{M}_0} u_0)(t,\boldsymbol{x}) := \int_\Omega G(t, \boldsymbol{x}-\boldsymbol{y}) u_0(\boldsymbol{y}) \,\mathrm{d}\boldsymbol{y} \quad \big((t,\boldsymbol{x}) \in Q\big) \tag{9.4}$$

denotes the initial potential,

$$(\widetilde{\mathscr{V}}\phi_N)(t,\boldsymbol{x}) := \int_\Sigma G(t-s, \boldsymbol{x}-\boldsymbol{y}) \phi_N(\boldsymbol{y}) \,\mathrm{d}\boldsymbol{y}\,\mathrm{d}s \quad \big((t,\boldsymbol{x}) \in Q\big) \tag{9.5}$$

denotes the single-layer potential, and

$$(\widetilde{\mathscr{K}} u_D)(t,\boldsymbol{x}) := \int_\Sigma \partial_{\boldsymbol{n}(\boldsymbol{y})} G(t-s, \boldsymbol{x}-\boldsymbol{y}) u_D(\boldsymbol{y}) \,\mathrm{d}\boldsymbol{y}\,\mathrm{d}s \quad \big((t,\boldsymbol{x}) \in Q\big) \tag{9.6}$$

denotes the double-layer potential. These linear operators satisfy the mapping properties $\widetilde{\mathscr{M}_0} : L^2(\Omega) \to H^{1,1/2}(Q)$, $\widetilde{\mathscr{V}_0} : H^{-1/2,-1/4}(\Sigma) \to H^{1,1/2}(Q)$, and $\widetilde{\mathscr{K}_0} : H^{1/2,1/4}(\Sigma) \to H^{1,1/2}(Q)$. The lateral trace $(\cdot)|_\Sigma$ of these potentials is given by

$$(\widetilde{\mathscr{M}_0} u_0)|_\Sigma = \mathscr{M}_0 u_0, \quad (\widetilde{\mathscr{V}}\phi_N)|_\Sigma = \mathscr{V}\phi_N, \quad (\widetilde{\mathscr{K}} u_D)|_\Sigma = (\mathscr{K} - 1/2) u_D,$$

where the initial operator $\mathscr{M}_0$, the single-layer operator $\mathscr{V}$, and the double-layer operator $\mathscr{K}$ are defined as in (9.4)–(9.6) for $(t, \boldsymbol{x}) \in \Sigma$. Applying the lateral trace to (9.3) thus results in

$$\mathscr{V}\phi_N = (\mathscr{K} + 1/2)u_D - \mathscr{M}_0 u_0, \tag{9.7}$$

i.e., (9.1) with $f := (\mathscr{K} + 1/2)u_D - \mathscr{M}_0 u_0$. As the single-layer operator $\mathscr{V}$ is also coercive, i.e.,

$$\langle \mathscr{V}\psi\,,\,\psi\rangle_\Sigma \geq c_{\mathrm{coe}}\|\psi\|^2_{H^{-1/2,-1/4}(\Sigma)} \quad \big(\psi \in H^{-1/2,-1/4}(\Sigma)\big) \tag{9.8}$$

with some constant $c_{\mathrm{coe}} > 0$, (9.7) is uniquely solvable and the solution ϕ_N is just the missing normal derivative $\partial_{\boldsymbol{n}} u$ to compute u via the representation formula (9.3).

Alternatively, one can make the ansatz $u = \widetilde{\mathscr{M}_0} u_0 + \widetilde{\mathscr{V}}\phi$. Indeed, both $\widetilde{\mathscr{M}_0} u_0$ and $\widetilde{\mathscr{V}}\phi$ satisfy the heat equation, where $\widetilde{\mathscr{M}_0} u_0$ restricted to $\{0\} \times \Omega$ coincides with u_0 and $\widetilde{\mathscr{V}}\phi$ vanishes there. To satisfy the Dirichlet boundary conditions, one has to solve

$$\mathscr{V}\phi = u_D - \mathscr{M}_0 u_0, \tag{9.9}$$

i.e., (9.1) with $f := u_D - \mathscr{M}_0 u_0$. While (9.7) is called direct method as it directly provides the physically relevant quantity $\phi_N = \partial_{\boldsymbol{n}} u$, (9.9) is called indirect method.

For more details and proofs, we refer to the seminal works [AN87, Noo88, Cos90], which considered $u_0 = 0$, and to [DNS19, Doh19] for the general case.

9.2.4 Boundary meshes

Throughout this work, we consider prismatic meshes $\mathcal{P}$ of Σ:

- $\mathcal{P}$ is a finite set of prisms of the form $P = J \times K$, where $J \subseteq \overline{I} = [0, T]$ is some non-empty compact interval and $K \subseteq \Gamma$ is the image of some compact Lipschitz domain[1] $\hat{K} \subset \mathbb{R}^{d-1}$ under some bi-Lipschitz mapping;
- for all $P, \tilde{P} \in \mathcal{P}$ with $P \neq \tilde{P}$, the intersection has measure zero on Σ;
- $\mathcal{P}$ is a partition of Σ, i.e., $\Sigma = \bigcup_{P\in\mathcal{P}} P$.

For arbitrary $t \in \overline{I}$ and $\boldsymbol{x} \in \Gamma$, we abbreviate the induced sets

$$\mathcal{P}|_t := \big\{K \subseteq \Gamma \,:\, (\{t\} \times \Gamma) \cap (J \times K) \neq \emptyset \text{ for some } J \times K \in \mathcal{P}\big\}$$

[1] A compact Lipschitz domain is the closure of a bounded Lipschitz domain. For $d = 2$, it is a compact interval with non-empty interior.

and

$$\mathcal{P}|_{\boldsymbol{x}} := \big\{ J \subseteq \overline{I} \,:\, (\overline{I} \times \{\boldsymbol{x}\}) \cap (J \times K) \neq \emptyset \text{ for some } J \times K \in \mathcal{P} \big\}.$$

For almost all $t \in \overline{I}$, $\mathcal{P}|_t$ is a mesh of Γ, i.e., a partition of Γ into finitely many compact Lipschitz domains such that the intersection of two distinct elements has measure zero on Γ. Similarly, for almost all $\boldsymbol{x} \in \Gamma$, $\mathcal{P}|_{\boldsymbol{x}}$ is a mesh of $\overline{I}$, i.e., a partition of $\overline{I}$ into finitely many non-empty compact intervals such that the intersection of two different intervals is at most a point. Note that for one fixed prismatic mesh $\mathcal{P}$ there exist constants $C_{\mathrm{nei}} \geq 1$, $C_{\mathrm{dist}} \geq 1$, $C_{\mathrm{shape}} \geq 1$, and $C_{\mathrm{lqu}} \geq 1$ such that:

- for almost all $t \in \overline{I}$, the number of neighbors of an element in $\mathcal{P}|_t$ is bounded, i.e.,

$$\#\big\{ \tilde{K} \in \mathcal{P}|_t \,:\, K \cap \tilde{K} \neq \emptyset \big\} \leq C_{\mathrm{nei}} \quad \text{for all } K \in \mathcal{P}|_t. \tag{9.10}$$

- for almost all $t \in \overline{I}$, the elements of $\mathcal{P}|_t$ are uniformly away from non-neighboring elements, i.e.,

$$\operatorname{diam}(K) \leq C_{\mathrm{dist}} \operatorname{dist}(K, \tilde{K}) \quad \text{for all } K, \tilde{K} \in \mathcal{P}|_t \text{ with } K \cap \tilde{K} = \emptyset; \tag{9.11}$$

- for almost all $t \in \overline{I}$, the elements of $\mathcal{P}|_t$ are shape-regular, i.e.,

$$C_{\mathrm{shape}}^{-1} |K|^{d-1} \leq \operatorname{diam}(K)^{d-1} \leq C_{\mathrm{shape}} |K| \quad \text{for all } K \in \mathcal{P}|_t; \tag{9.12}$$

- for almost all $\boldsymbol{x} \in \Gamma$, $\mathcal{P}|_{\boldsymbol{x}}$ is locally quasi-uniform, i.e.,

$$|J| \leq C_{\mathrm{lqu}} |\tilde{J}| \quad \text{for all } J, \tilde{J} \in \mathcal{P}|_{\boldsymbol{x}} \text{ with } J \cap \tilde{J} \neq \emptyset. \tag{9.13}$$

In the remainder of this work, we will always indicate the dependence of estimates on these particular constants.

Remark 9.2.1. If, for $d = 2$, the meshes $\mathcal{P}|_t$ are found by iteratively bisecting some initial mesh and the level difference of neighboring elements is bounded by 1, then the constants from (9.10)–(9.12) depend only on the initial mesh; cf. [AFF+13]. For $d = 3$, the same holds true if the initial mesh is for instance a conforming (curvilinear) triangulation of Γ and one iteratively applies newest vertex bisection. The arguments for (9.10)–(9.11) are found in [AFF+17, Section 2.3 and 4.1].

9.2.5 Boundary element method

Given a prismatic boundary mesh $\mathcal{P}$ and an associated finite-dimensional trial space $\mathcal{X} \subset H^{-1/2,-1/4}(\Sigma)$, e.g., the space of all $\mathcal{P}$-piecewise polynomials of

some fixed degree in space and time, let $\Phi \in \mathcal{X}$ denote the Galerkin discretization of the solution ϕ of the boundary integral equation (9.1), i.e.,

$$\langle \mathscr{V}\Phi\,,\,\Psi\rangle_\Sigma = \langle f\,,\,\Psi\rangle_\Sigma \quad \big(\Psi \in \mathcal{X}\big),$$

which is equivalent to the Galerkin orthogonality

$$\langle \mathscr{V}(\phi-\Phi)\,,\,\Psi\rangle_\Sigma = 0 \quad \big(\Psi \in \mathcal{X}\big). \tag{9.14}$$

Note that coercivity (9.8) guarantees unique solvability of the latter equations, and the Céa lemma applies

$$\|\phi-\Phi\|_{H^{-1/2,-1/4}(\Sigma)} \le \frac{C_{\mathrm{cont}}}{c_{\mathrm{coe}}} \min_{\Psi\in\mathcal{X}} \|\phi-\Psi\|_{H^{-1/2,-1/4}(\Sigma)},$$

where C_{cont} is the operator norm of $\mathscr{V} : H^{-1/2,-1/4}(\Sigma) \to H^{1/2,1/4}(\Sigma)$.

Suppose $\mathcal{P} = \big\{J\times K \,:\, J\in\mathcal{P}_{\overline{I}}, K\in\mathcal{P}_\Gamma\big\}$ is a full tensor-mesh corresponding to a mesh $\mathcal{P}_\Gamma$ of Γ with uniform mesh-size $h_{\boldsymbol{x}} \eqsim \operatorname{diam}(K)$ for all $K\in\mathcal{P}_\Gamma$ and a mesh $\mathcal{P}_{\overline{I}}$ of $\overline{I}$ with uniform step-size $h_t \eqsim h_{\boldsymbol{x}}^\sigma$ for some $\sigma>0$. Using $\mathcal{P}$-piecewise polynomials of some degree $p_{\boldsymbol{x}}\in\mathbb{N}_0$ in space- and some degree $p_t\in\mathbb{N}_0$ in time-direction as trial space $\mathcal{X}$, then gives the error decay rate

$$\min_{\Psi\in\mathcal{X}} \|\phi-\Psi\|_{H^{-1/2,-1/4}(\Sigma)} \lesssim N^{-\frac{\min\{p_{\boldsymbol{x}}+3/2,(p_t+5/4)\sigma\}}{d-1+\sigma}} \quad \text{for all smooth } \phi; \tag{9.15}$$

see [CR19, Theorem 3.3]. Here, $N \eqsim h_{\boldsymbol{x}}^{-(d-1)} h_t^{-1} = h_{\boldsymbol{x}}^{d-1+\sigma}$ denotes the number of degrees of freedom. The optimal grading parameter is thus given by $\sigma = (p_{\boldsymbol{x}}+\frac{3}{2})/(p_t+\frac{5}{4})$ with resulting rate $\mathcal{O}\big(N^{-\frac{p_{\boldsymbol{x}}+3/2}{d-1+\sigma}}\big)$.

9.3 A posteriori error estimation

As $\mathscr{V}$ is an isomorphism, it holds that

$$\|\phi-\Phi\|_{H^{-1/2,-1/4}(\Sigma)} \eqsim \|f-\mathscr{V}\Phi\|_{H^{1/2,1/4}(\Sigma)}. \tag{9.16}$$

Here, $\Phi \in H^{-1/2,-1/4}(\Sigma)$ can be an arbitrary approximation of the solution ϕ of (9.1). While the right-hand side is in principle *a posteriori* computable, the computation of the Sobolev–Slobodeckij norm over the full space-time boundary Σ is expensive, and it does not provide any information on where to locally refine the given mesh to increase the accuracy of the approximation. According to (9.16), it is sufficient to derive suitable estimate for the residual $f-\mathscr{V}\Phi$ in the $H^{1/2,1/4}(\Sigma)$-norm. Recall that this term is $L_2(\Sigma)$-orthogonal to all functions $\Psi\in\mathcal{X}$ provided that Φ is the Galerkin approximation of ϕ in $\mathcal{X}$; see (9.14).

9.3.1 Localization of the anisotropic Sobolev–Slobodeckij norm

The following proposition provides the key argument for our *a posteriori* error estimation. While the first inequality is trivial, the original version of the second one already goes back to [Fae00, Fae02]. We make use of the slightly generalized version from [GP20, Lemma 4.5]; see [Gan17, Lemma 5.3.2] for a detailed proof.

Proposition 9.3.1. *Let $\mu \in (0,1)$ and $\mathcal{P}_\Gamma$ be a mesh of Γ. Then, there exist constants $C_1, C_2 > 0$ such that for all $v \in H^\mu(\Gamma)$, there holds that*

$$C_1^{-1} \sum_{K\in\mathcal{P}_\Gamma} \sum_{\substack{\tilde{K}\in\mathcal{P}_\Gamma\\ K\cap\tilde{K}\neq\emptyset}} |v|^2_{H^\mu(K\cup\tilde{K})} \le \|v\|^2_{H^\mu(\Gamma)} \le \sum_{K\in\mathcal{P}_\Gamma} \sum_{\substack{\tilde{K}\in\mathcal{P}_\Gamma\\ K\cap\tilde{K}\neq\emptyset}} |v|^2_{H^\mu(K\cup\tilde{K})} + C_2 \sum_{K\in\mathcal{P}_\Gamma} \operatorname{diam}(K)^{-2\mu}\|v\|^2_{L_2(K)}.$$

The constant C_1 is given as $C_1 = 2(C_{\mathrm{nei}}+1)^2$ with C_{nei} from (9.10) *(with $\mathcal{P}|_t$ replaced by $\mathcal{P}_\Gamma$), and C_2 depends only on the dimension d, μ, Γ, and the constant C_{dist} from* (9.11) *(with $\mathcal{P}|_t$ replaced by $\mathcal{P}_\Gamma$).* □

Note that local quasi-uniformity (9.13) (with $\mathcal{P}|_{\boldsymbol{x}}$ replaced by $\mathcal{P}_{\overline{I}}$) of a time mesh $\mathcal{P}_{\overline{I}}$ is actually equivalent to

$$\operatorname{diam}(J) = |J| \le C_{\mathrm{lqu}} \operatorname{dist}(J,\tilde{J}) \quad \text{for all } J,\tilde{J}\in\mathcal{P}_{\overline{I}} \text{ with } J\cap\tilde{J}=\emptyset.$$

Moreover, for any element $J \in \mathcal{P}_{\overline{I}}$, there are at most three $\tilde{J} \in \mathcal{P}_{\overline{I}}$ with $J \cap \tilde{J} \neq \emptyset$. In particular, the same reference as before applies and we also obtain the following proposition.

Proposition 9.3.2. *Let $\nu \in (0,1)$ and $\mathcal{P}_{\overline{I}}$ be a mesh of $\overline{I}$. Then, there exist constants $C_1, C_2 > 0$ such that for all $v \in H^\nu(I)$, there holds that*

$$C_1^{-1} \sum_{J\in\mathcal{P}_{\overline{I}}} \sum_{\substack{\tilde{J}\in\mathcal{P}_{\overline{I}}\\ J\cap J\neq\emptyset}} |v|^2_{H^\nu(J\cup\tilde{J})} \le \|v\|^2_{H^\nu(I)} \le \sum_{J\in\mathcal{P}_{\overline{I}}} \sum_{\substack{\tilde{J}\in\mathcal{P}_{\overline{I}}\\ J\cap J\neq\emptyset}} |v|^2_{H^\nu(J\cup\tilde{J})} + C_2 \sum_{J\in\mathcal{P}_{\overline{I}}} |J|^{-2\nu}\|v\|^2_{L_2(J)}.$$

The constant C_1 is given as $C_1 = 32$, and C_2 depends only on ν, $|I|$, and the constant C_{lqu} from (9.13) *(with $\mathcal{P}|_{\boldsymbol{x}}$ replaced by $\mathcal{P}_{\overline{I}}$).* □

The latter two propositions allow to derive the following *a posteriori error estimation*, which can be employed for arbitrary approximations Φ.

Theorem 9.3.3. *Let $\mu, \nu \in (0,1)$ and $\mathcal{P}$ be a prismatic mesh of Σ. Then, there exist constants $C'_{\mathrm{eff}}, C''_{\mathrm{rel}} > 0$ such that for all $v \in H^{\mu,\nu}(\Sigma)$, there holds that*

$$\sum_{J\times K\in\mathcal{P}} \Big(\sum_{\substack{\tilde{J}\times\tilde{K}\in\mathcal{P}\\ |J\cap\tilde{J}|>0\\ K\cap\tilde{K}\neq\emptyset}} |v|^2_{L_2(J\cap\tilde{J};H^\mu(K\cup\tilde{K}))} + \sum_{\substack{\tilde{J}\times\tilde{K}\in\mathcal{P}\\ J\cap\tilde{J}\neq\emptyset\\ |K\cap\tilde{K}|>0}} |v|^2_{H^\nu(J\cup\tilde{J};L_2(K\cap\tilde{K}))} \Big) \leq (C'_{\mathrm{eff}})^2 \|v\|^2_{H^{\mu,\nu}(\Sigma)}$$

as well as

$$\begin{aligned}(C'_{\mathrm{rel}})^{-2}\|v\|^2_{H^{\mu,\nu}(\Sigma)} \leq & \sum_{J\times K\in\mathcal{P}} \Big(\sum_{\substack{\tilde{J}\times\tilde{K}\in\mathcal{P}\\ |J\cap\tilde{J}|>0\\ K\cap\tilde{K}\neq\emptyset}} |v|^2_{L_2(J\cap\tilde{J};H^\mu(K\cup\tilde{K}))} + \sum_{\substack{\tilde{J}\times\tilde{K}\in\mathcal{P}\\ J\cap\tilde{J}\neq\emptyset\\ |K\cap\tilde{K}|>0}} |v|^2_{H^\nu(J\cup\tilde{J};L_2(K\cap\tilde{K}))} \Big) \\ & + \sum_{J\times K\in\mathcal{P}} \big(\operatorname{diam}(K)^{-2\mu} + |J|^{-2\nu}\big)\|v\|^2_{L_2(J\times K)}. \end{aligned} \tag{9.17}$$

The constant C'_{eff} is given as $C'_{\mathrm{eff}} = \max(2(C_{\mathrm{nei}}+1)^2, 32)$ with C_{nei} from (9.10)*, and C'_{rel} depends only on d, μ, ν, Γ, $|I|$ and the constants C_{dist} from* (9.11) *as well as C_{lqu} from* (9.13)*.*

Proof. We split the proof into four steps.
Step 1: In this step, we bound $\|v\|_{L_2(I;H^\mu(\Gamma))}$ from below. Proposition 9.3.1 gives that

$$\|v\|^2_{L_2(I;H^\mu(\Gamma))} = \int_I \|v(t,\cdot)\|^2_{H^\mu(\Gamma)}\,\mathrm{d}t \gtrsim \int_I \sum_{K\in\mathcal{P}|_t} \sum_{\substack{\tilde{K}\in\mathcal{P}|_t\\ K\cap\tilde{K}\neq\emptyset}} |v(t,\cdot)|^2_{H^\mu(K\cup\tilde{K})}\,\mathrm{d}t.$$

Note that $K \in \mathcal{P}|_t$ is equivalent to $J\times K \in \mathcal{P}$ for some J with $t \in J$. With the indicator function $\mathbb{1}_S$ of a set S, the last term thus is equal to

$$\begin{aligned}\int_I \sum_{K\in\mathcal{P}|_t} \sum_{\substack{\tilde{K}\in\mathcal{P}|_t\\ K\cap\tilde{K}\neq\emptyset}} |v(t,\cdot)|^2_{H^\mu(K\cup\tilde{K})}\,\mathrm{d}t &= \int_I \sum_{J\times K\in\mathcal{P}} \mathbb{1}_J(t) \sum_{\substack{\tilde{J}\times\tilde{K}\in\mathcal{P}\\ K\cap\tilde{K}\neq\emptyset}} \mathbb{1}_{\tilde{J}}(t)|v(t,\cdot)|^2_{H^\mu(K\cup\tilde{K})}\,\mathrm{d}t \\ &= \sum_{J\times K\in\mathcal{P}} \sum_{\substack{\tilde{J}\times\tilde{K}\in\mathcal{P}\\ |J\cap\tilde{J}|>0\\ K\cap\tilde{K}\neq\emptyset}} |v|^2_{L_2(J\cap\tilde{J};H^\mu(K\cup\tilde{K}))}.\end{aligned}$$

Step 2: In this step, we bound $\|v\|_{L_2(I;H^\mu(\Gamma))}$ from above. Proposition 9.3.1 gives that

$$\begin{aligned}&\|v\|^2_{L_2(I;H^\mu(\Gamma))} = \int_I \|v(t,\cdot)\|^2_{H^\mu(\Gamma)}\,\mathrm{d}t \\ &\lesssim \int_I \sum_{K\in\mathcal{P}|_t} \sum_{\substack{\tilde{K}\in\mathcal{P}|_t\\ K\cap\tilde{K}\neq\emptyset}} |v(t,\cdot)|^2_{H^\mu(K\cup\tilde{K})} + \sum_{K\in\mathcal{P}|_t} \operatorname{diam}(K)^{-2\mu}\|v(t,\cdot)\|^2_{L_2(K)}\,\mathrm{d}t.\end{aligned}$$

The first term in this expression has already been treated in Step 1. As $K \in \mathcal{P}|_t$ is equivalent to $J \times K \in \mathcal{P}$ for some J with $t \in J$, the second term reads

$$\begin{aligned}\int_I \sum_{K\in\mathcal{P}|_t} \operatorname{diam}(K)^{-2\mu}\|v(t,\cdot)\|_{L_2(K)}^2 \,\mathrm{d}t &= \int_I \sum_{J\times K\in\mathcal{P}} \mathbb{1}_J(t)\, \operatorname{diam}(K)^{-2\mu}\|v(t,\cdot)\|_{L_2(K)}^2 \,\mathrm{d}t \\ &= \sum_{J\times K\in\mathcal{P}} \operatorname{diam}(K)^{-2\mu}\|v\|_{L_2(J\times K)}^2.\end{aligned}$$

Step 3: In this step, we bound $\|v\|_{H^\nu(I;L_2(\Gamma))}$ from below. The Fubini theorem, Proposition 9.3.2, and the same argument as in Step 1 give that

$$\begin{aligned}\|v\|_{H^\nu(I;L_2(\Gamma))}^2 &\gtrsim \int_\Gamma \sum_{J\in\mathcal{P}|_{\boldsymbol{x}}} \sum_{\substack{\tilde J\in\mathcal{P}|_{\boldsymbol{x}}\\ J\cap\tilde J\neq\emptyset}} |v(\cdot,\boldsymbol{x})|_{H^\nu(J\cup\tilde J)}^2 \,\mathrm{d}\boldsymbol{x} \\ &= \int_\Gamma \sum_{J\times K\in\mathcal{P}} \mathbb{1}_K(\boldsymbol{x}) \sum_{\substack{\tilde J\times\tilde K\in\mathcal{P}\\ J\cap\tilde J\neq\emptyset}} \mathbb{1}_{\tilde K}(\boldsymbol{x}) |v(\cdot,\boldsymbol{x})|_{H^\nu(J\cup\tilde J)}^2 \,\mathrm{d}\boldsymbol{x} \\ &= \sum_{J\times K\in\mathcal{P}} \sum_{\substack{\tilde J\times\tilde K\in\mathcal{P}\\ J\cap\tilde J\neq\emptyset\\ |K\cap\tilde K|>0}} |v|_{H^\nu(J\cup\tilde J;L_2(K\cap\tilde K))}^2.\end{aligned}$$

Step 4: In this step, we bound $\|v\|_{H^\nu(I;L_2(\Gamma))}$ from above. The Fubini theorem and Proposition 9.3.2 give that

$$\begin{aligned}\|v\|_{H^\nu(I;L_2(\Gamma))}^2 &= \int_\Gamma \|v(\cdot,\boldsymbol{x})\|_{H^\nu(I)}^2 \,\mathrm{d}\boldsymbol{x} \\ &\lesssim \int_\Gamma \sum_{J\in\mathcal{P}|_{\boldsymbol{x}}} \sum_{\substack{\tilde J\in\mathcal{P}|_{\boldsymbol{x}}\\ J\cap\tilde J\neq\emptyset}} |v(\cdot,\boldsymbol{x})|_{H^\nu(J\cup\tilde J)}^2 + \sum_{J\in\mathcal{P}|_{\boldsymbol{x}}} |J|^{-2\nu}\|v(\cdot,\boldsymbol{x})\|_{L_2(J)}^2 \,\mathrm{d}\boldsymbol{x}.\end{aligned}$$

The first term has already been treated in Step 3. The second term reads

$$\begin{aligned}\int_\Gamma \sum_{J\in\mathcal{P}|_{\boldsymbol{x}}} |J|^{-2\nu}\|v(\cdot,\boldsymbol{x})\|_{L_2(J)}^2 \,\mathrm{d}\boldsymbol{x} &= \int_\Gamma \sum_{J\times K\in\mathcal{P}} \mathbb{1}_K(\boldsymbol{x})\, |J|^{-2\nu}\|v(\cdot,\boldsymbol{x})\|_{L_2(J)}^2 \,\mathrm{d}\boldsymbol{x} \\ &= \sum_{J\times K\in\mathcal{P}} |J|^{-2\nu}\|v\|_{L_2(J\times K)}^2.\end{aligned}$$

This concludes the proof. □

9.3.2 Poincaré-type inequality

Assuming the grading $|J| \eqsim \operatorname{diam}(K)^{\mu/\nu}$ as well as $L_2(\Sigma)$-orthogonality of v to piecewise constants, the following local Poincaré-type inequality allows to get rid of the weighted L_2-terms in (9.17). The proof works essentially as in [Cos90, Proposition 5.3], where a global version on uniform meshes is considered.

Lemma 9.3.4. *Let $\mu, \nu \in (0,1)$ and $\mathcal{P}$ be a prismatic mesh of Σ. Then, there holds for all $v \in H^{\mu,\nu}(\Sigma)$ and all $J \times K \in \mathcal{P}$ with $\langle v\,,\,1\rangle_{L_2(J\times K)} = 0$ that*

$$\|v\|^2_{L_2(J\times K)} \leq C_{\mathrm{shape}}\big(\operatorname{diam}(K)^{2\mu}|v|^2_{L_2(J;H^\mu(K))} + |J|^{2\nu}|v|^2_{H^\nu(J;L_2(K))}\big).$$

Here, $C_{\mathrm{shape}} \geq 1$ is the constant from (9.12)*.*

Proof. Let Π_J, Π_K, and $\Pi_{J\times K}$ denote the L_2-orthogonal projection onto the space of constants on J, K, and $J\times K$, respectively. Note that $\Pi_{J\times K} = \Pi_J \otimes \Pi_K$ and thus

$$\begin{aligned}\|v\|_{L_2(J\times K)} &= \|(1-\Pi_{J\times K})v\|_{L_2(J\times K)}\\ &\leq \|(1-\Pi_J\otimes\mathrm{Id})v\|_{L_2(J\times K)} + \|(\Pi_J\otimes\mathrm{Id} - \Pi_J\otimes\Pi_K)v\|_{L_2(J\times K)}.\end{aligned}$$

As Π_J has operator norm 1, a standard Poincaré-type inequality, see, e.g., [Fae02, Lemma 3.4] for the elementary proof, shows for the second term that

$$\begin{aligned}\|(\Pi_J\otimes\mathrm{Id} - \Pi_J\otimes\Pi_K)v\|^2_{L_2(J\times K)} &\leq \|(1-\mathrm{Id}\otimes\Pi_K)v\|^2_{L_2(J\times K)}\\ &= \int_J \|(1-\Pi_K)v(t,\cdot)\|^2_{L_2(K)}\,\mathrm{d}t\\ &\leq \frac{\operatorname{diam}(K)^{d-1+2\mu}}{2|K|}\int_J |v(t,\cdot)|^2_{H^\mu(K)}\,\mathrm{d}t\\ &\leq \frac{C_{\mathrm{shape}}}{2}\operatorname{diam}(K)^{2\mu}|v|^2_{L_2(J;H^\mu(K)}.\end{aligned}$$

The first term can be estimated similarly

$$\|(1-\Pi_J\otimes\mathrm{Id})v\|^2_{L_2(J\times K)} \leq \frac{1}{2}|J|^{2\nu}|v|^2_{H^\nu(J;L_2(K))},$$

which concludes the proof. □

9.3.3 A posteriori error estimators

For arbitrary prismatic meshes $\mathcal{P}$ of Σ with some associated trial space $\mathcal{X} \subset H^{-1/2,-1/4}(\Sigma)$ and $\Phi \in \mathcal{X}$, we define the following error indicators for all $J\times K \in \mathcal{P}$,

$$\begin{aligned}\eta^{\boldsymbol{x}}_{\mathcal{P}}(\Phi, J\times K) &:= \sum_{\substack{\tilde{J}\times\tilde{K}\in\mathcal{P}\\ |J\cap\tilde{J}|>0\\ K\cap\tilde{K}\neq\emptyset}} |f-\mathscr{V}\Phi|^2_{L_2(J\cap\tilde{J};H^{1/2}(K\cup\tilde{K}))},\\ \eta^{t}_{\mathcal{P}}(\Phi, J\times K)^2 &:= \sum_{\substack{\tilde{J}\times\tilde{K}\in\mathcal{P}\\ J\cap\tilde{J}\neq\emptyset\\ |K\cap\tilde{K}|>0}} |f-\mathscr{V}\Phi|^2_{H^{1/4}(J\cup\tilde{J};L_2(K\cap\tilde{K}))},\\ \zeta^{\boldsymbol{x}}_{\mathcal{P}}(\Phi, J\times K) &:= \operatorname{diam}(K)^{-1}\|f-\mathscr{V}\Phi\|^2_{L_2(J\times K)},\\ \zeta^{t}_{\mathcal{P}}(\Phi, J\times K)^2 &:= |J|^{-1/2}\|f-\mathscr{V}\Phi\|^2_{L_2(J\times K)}.\end{aligned}$$

The corresponding error estimators read as

$$\eta_{\mathcal{P}}(\Phi, J\times K)^2 := \eta_{\mathcal{P}}^{\boldsymbol{x}}(\Phi, J\times K)^2 + \eta_{\mathcal{P}}^{t}(\Phi, J\times K)^2,$$
$$\zeta_{\mathcal{P}}(\Phi, J\times K)^2 := \zeta_{\mathcal{P}}^{\boldsymbol{x}}(\Phi, J\times K)^2 + \zeta_{\mathcal{P}}^{t}(\Phi, J\times K)^2,$$
$$\eta_{\mathcal{P}}(\Phi)^2 := \sum_{J\times K\in\mathcal{P}} \eta_{\mathcal{P}}(\Phi, J\times K)^2, \quad \zeta_{\mathcal{P}}(\Phi)^2 := \sum_{J\times K\in\mathcal{P}} \zeta_{\mathcal{P}}(\Phi, J\times K)^2.$$

With (9.16), we overall obtain the following *a posteriori* estimates.

Corollary 9.3.5. *Let ϕ be the solution of* (9.1) *and $\mathcal{P}$ be a prismatic mesh of Σ with some associated discrete trial space $\mathcal{X} \subset H^{-1/2,-1/4}(\Sigma)$. Then, there exist constants $C_{\mathrm{eff}}, \tilde{C}_{\mathrm{rel}} > 0$ such that for arbitrary $\Phi \in \mathcal{X}$, there holds that*

$$C_{\mathrm{eff}}^{-1}\eta_{\mathcal{P}}(\Phi) \le \|\phi - \Phi\|_{H^{-1/2,-1/4}(\Sigma)} \le \tilde{C}_{\mathrm{rel}}\big(\eta_{\mathcal{P}}(\Phi)^2 + \zeta_{\mathcal{P}}(\Phi)^2\big)^{1/2}.$$

If the space $\mathcal{X}$ contains all $\mathcal{P}$-piecewise constant functions and $\Phi \in \mathcal{X}$ is the Galerkin approximation of ϕ, there further holds that

$$\zeta_{\mathcal{P}}(\Phi, J\times K)^2 \le C_{\mathrm{shape}}\big(\operatorname{diam}(K)^{-1} + |J|^{-1/2}\big)\big(\operatorname{diam}(K) + |J|^{1/2}\big)\eta_{\mathcal{P}}(\Phi, J\times K)^2$$

for all $J \times K \in \mathcal{P}$. If $C_{\mathrm{grad}}^{-1}\operatorname{diam}(K) \le |J|^{1/2} \le C_{\mathrm{grad}}\operatorname{diam}(K)$ is satisfied for all $J \times K \in \mathcal{P}$ and a uniform constant $C_{\mathrm{grad}} \ge 1$, this implies the existence of a constant $C_{\mathrm{rel}} > 0$ such that

$$\|\phi - \Phi\|_{H^{-1/2,-1/4}(\Sigma)} \le C_{\mathrm{rel}}\eta_{\mathcal{P}}(\Phi).$$

The constants C_{eff} and $\tilde{C}_{\mathrm{rel}}$ are given as $C_{\mathrm{eff}} = C'_{\mathrm{eff}}\,C_{\mathrm{cont}}$ and $C_{\mathrm{rel}} = C'_{\mathrm{rel}}/c_{\mathrm{coe}}$ with C'_{eff} and C'_{rel} from Theorem 9.3.3, the operator norm C_{cont} of $\mathcal{V}$, and c_{coe} from (9.8). *The constant C_{rel} is given as $C_{\mathrm{rel}} = \tilde{C}_{\mathrm{rel}}\sqrt{2C_{\mathrm{shape}}(1 + C_{\mathrm{grad}})}$.* □

Remark 9.3.6. According to (9.15), the required scaling $|J| \eqsim \operatorname{diam}(K)^2$, i.e., $\sigma = 2$, is the optimal scaling for approximating smooth solutions ϕ if the polynomial degrees of $\mathcal{X}$ satisfy $p_{\boldsymbol{x}} = 2p_t + 1$.

9.4 Numerical experiments

In this section, we employ the error estimator η within an adaptive algorithm using different refinement strategies, and investigate the resulting convergence rates. We restrict ourselves to the case $d = 2$, with $\Gamma = \partial\Omega$ being the boundary of a polygonal domain $\Omega \subset \mathbb{R}^2$, and set the time domain to be $I = (0, 1)$.

For $\mathcal{P}$ a prismatic mesh of the space-time boundary, i.e., a quadrilateral mesh as $d = 2$, we consider the trial space $\mathcal{X}$ of piecewise constants with respect to $\mathcal{P}$. In particular, this allows us to perform integration in time analytically for all integrals that are involved in the computation of the Galerkin matrix and the evaluation of the single-layer operator $\mathcal{V}$; see, e.g., [Cos90]. The remaining

integrals over Γ have a logarithmic singularity, for which we use the quadrature rules from [Smi00]. For the computation of the Sobolev–Slobodeckij seminorm in the Faermann estimator $\eta_{\mathcal{P}}(\Phi)$, we use Duffy transformations and Gauss quadrature for the regularized integrands.

9.4.1 Adaptive algorithm

In our numerical experiments below, we employ the following adaptive algorithm with $\theta = 0.9$.

Algorithm. *Let $0 < \theta \leq 1$ be a marking parameter and $\mathcal{P} = \big\{J \times K : J \in \mathcal{P}_{\overline{I}}, K \in \mathcal{P}_\Gamma\big\}$ be an initial tensor-mesh corresponding to a mesh $\mathcal{P}_\Gamma$ of Γ and a mesh $\mathcal{P}_{\overline{I}}$ of $\overline{I} = [0,T]$. For each $\ell = 0, 1, 2, \dots$, iterate the following steps:*

(i) *Compute Galerkin approximation Φ_ℓ of ϕ in the space $\mathcal{X}_\ell$ of all $\mathcal{P}_\ell$-piecewise constant functions on Σ.*

(ii) *Compute indicators $\eta^{\boldsymbol{x}}_{\mathcal{P}_\ell}(\Phi_\ell, J\times K)$ and $\eta^{t}_{\mathcal{P}_\ell}(\Phi_\ell, J\times K)$ for all elements $J\times K \in \mathcal{P}_\ell$.*

(iii) *Determine two minimal sets of marked elements $\mathcal{M}^{\boldsymbol{x}}_\ell, \mathcal{M}^t_\ell \subseteq \mathcal{P}_\ell$ such that*

$$\theta^2 \eta_{\mathcal{P}_\ell}(\Phi_\ell)^2 \leq \sum_{J\times K\in\mathcal{M}^{\boldsymbol{x}}_\ell} \eta^{\boldsymbol{x}}_{\mathcal{P}_\ell}(\Phi_\ell, J\times K)^2 + \sum_{J\times K\in\mathcal{M}^t_\ell} \eta^{t}_{\mathcal{P}_\ell}(\Phi_\ell, J\times K)^2. \tag{9.18}$$

(iv) *Refine at least all marked elements to obtain a new mesh $\mathcal{P}_{\ell+1}$.*

We will focus on *isotropic* and *anisotropic* adaptive strategies:

- In isotropic refinement, we require $\mathcal{M}^{\boldsymbol{x}}_\ell = \mathcal{M}^t_\ell$ in the marking step (iii), so that (9.18) simplifies to $\theta^2\eta_{\mathcal{P}_\ell}(\Phi_\ell)^2 \leq \sum_{J\times K\in\mathcal{M}_\ell} \eta_{\mathcal{P}_\ell}(\Phi_\ell, J\times K)^2$. In the refinement step (iv), we iteratively mark additional elements to ensure that, after subdividing all marked elements into four congruent rectangles, the new mesh $\mathcal{P}_{\ell+1}$ has only one hanging node per edge.
- In anisotropic refinement, we bisect the elements $\mathcal{M}^{\boldsymbol{x}}_\ell \setminus \mathcal{M}^t_\ell$ in space, bisect the elements $\mathcal{M}^t_\ell \setminus \mathcal{M}^{\boldsymbol{x}}_\ell$ in time, and subdivide all elements $\mathcal{M}^{\boldsymbol{x}}_\ell \cap \mathcal{M}^t_\ell$ into four congruent rectangles. Then, we iteratively bisect additional elements in space and/or time to ensure that the level difference in space and in time between elements sharing an edge in the new mesh $\mathcal{P}_{\ell+1}$ is bounded by 1. Here, the level in space and the level in time of elements are defined as the number of bisections in space and time, respectively, to obtain the element from the initial mesh $\mathcal{P}_0$.

For comparison, we also include uniform refinement, where $\mathcal{P}_{\ell+1}$ is obtained from $\mathcal{P}_\ell$ by subdividing each element into four congruent rectangles. For all considered refinement strategies, it is easy to see that the mesh constants from (9.10)–(9.13) for to $(\mathcal{P}_\ell)_{\ell\in\mathbb{N}_0}$ depend only on the initial mesh $\mathcal{P}_0$.

9.4.2 Reference for exact error

As the exact error $\|\phi - \Phi\|_{H^{-1/2,-1/4}(\Sigma)}$ cannot be readily computed in the examples below, we compare the error estimator $\eta_{\mathcal{P}}$ and the weighted L_2-terms $\zeta_{\mathcal{P}}$ from Section 9.3.3 with the following $(h - h/2)$-estimator: For a mesh $\mathcal{P}$, define the uniformly refined mesh as $\widehat{\mathcal{P}}$. With the the Galerkin approximation $\widehat{\Phi}$ from the refined trial space, we define the $(h - h/2)$-estimator as

$$\|\Phi - \widehat{\Phi}\|_{\mathscr{V}}^2 := \langle \mathscr{V}(\Phi - \widehat{\Phi})\,,\, \Phi - \widehat{\Phi}\rangle_{\Sigma}.$$

Under the saturation assumption $\|\phi - \widehat{\Phi}\|_{\mathscr{V}} \leq q_{\mathrm{sat}}\|\phi - \Phi\|_{\mathscr{V}}$, the triangle inequality shows that this estimator is equivalent to $\|\phi - \Phi\|_{\mathscr{V}}$, and therefore to the error $\|\phi - \Phi\|_{H^{-1/2,-1/4}(\Sigma)}$ by coercivity of $\mathscr{V}$. Note that the saturation assumption is indeed satisfied under the realistic (asymptotic) assumption that $\|\phi - \Phi\|_{\mathscr{V}} = \mathcal{O}\big((\#\mathcal{P})^{-s}\big)$ for some arbitrary rate $s > 0$.

9.4.3 Smooth problem

Let $\Omega = (0,1)^2$ and prescribe $u(t, x_1, x_2) := \exp(-2\pi^2 t)\sin(\pi x_1)\sin(\pi x_2)$ with initial condition $u_0(x_1, x_2) := \sin(\pi x_1)\sin(\pi x_2)$ and Dirichlet data $u_D \equiv 0$. We choose $\mathcal{P}_0 := \big\{[0,1] \times K \,:\, K \in \mathcal{P}_\Gamma\big\}$ with the uniform mesh $\mathcal{P}_\Gamma$ of Γ being aligned with the corners and consisting of four elements, as initial mesh of the space-time boundary Σ.

Figure 9.1 displays the results in double-logarithmic plots so that the slopes of the lines indicate the corresponding convergence rates. With the number of degrees of freedom $N = \#\mathcal{P}$, we see the expected rate $\mathcal{O}(N^{-5/8}) = \mathcal{O}(N^{-0.625})$ from (9.15) for both uniform refinement and isotropic refinement (with still slightly worse rate for the L_2-terms $\zeta_{\mathcal{P}}(\Phi)$ for uniform refinement), albeit adaptive isotropic refinement offers quantitively better results. For anisotropic refinement refinement, the rate is improved to $\mathcal{O}(N^{-15/22}) \approx \mathcal{O}(N^{-0.68})$. According to (9.15), this coincides with the best possible rate that can be achieved with uniform tensor-meshes, where the optimal scaling parameter in $h_t \eqsim h_{\boldsymbol{x}}^\sigma$ is given by $\sigma = 6/5$. Note that we do not require setting an explicit scaling in our anisotropic adaptive algorithm, it recovers the optimal rate automatically.

9.4.4 Mildly singular problem

Let $\Omega = (0,1)^2$, with initial condition $u_0 \equiv 0$ and Dirichlet data $u_D(t, x_1, x_2) := t^2$. We expect the solution here to be only singular in the four corners of the unit square as the initial condition is compatible with the Dirichlet data. The initial mesh $\mathcal{P}_0$ is chosen as in Section 9.4.3. Figure 9.2 displays the results. The assymptotic decay rate for all estimators under uniform refinement seems to be $\mathcal{O}(N^{-1/3})$, which is improved to $\mathcal{O}(N^{-1/2})$ for isotropic refinement, and finally, under anistriopic refinement this becomes the optimal rate $\mathcal{O}(N^{-15/22})$.

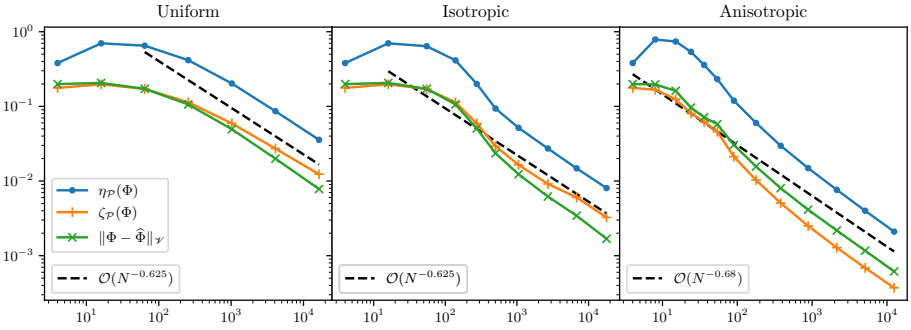


FIGURE 9.1. Error estimators for the smooth problem of Section 9.4.3 plotted double-logarithmically over the degrees of freedom $N = \#\mathcal{P}$: uniform refinement (left), isotropic refinement (middle), and anisotropic refinement (right).

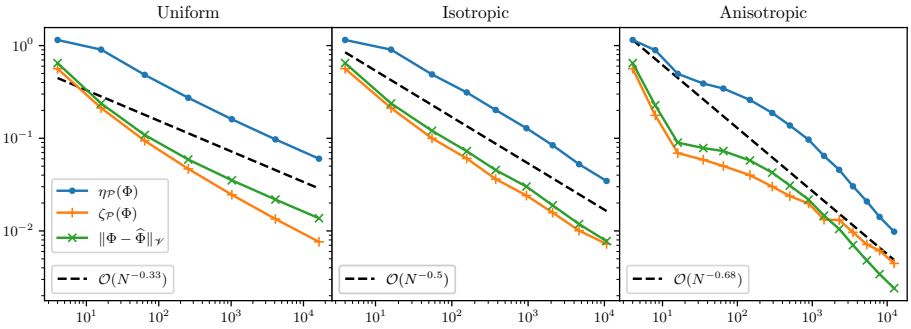


FIGURE 9.2. Error estimators for the mildly singular problem of Section 9.4.4 plotted double-logarithmically over the degrees of freedom $N = \#\mathcal{P}$: uniform refinement (left), isotropic refinement (middle), and anisotropic refinement (right).

9.4.5 Singular problem

Let $\Omega = (0,1)^2$ with initial condition $u_0 \equiv 0$ and Dirichlet data $u_D \equiv 1$. The solution to this problem is known to have a strong singularity for $t = 0$ due to the incompatibility of initial and boundary conditions, in addition to singularities in the four corners of the unit square. The initial mesh $\mathcal{P}_0$ is chosen as in Section 9.4.3.

Figure 9.3 displays the results. The Faermann estimator $\eta_{\mathcal{P}}(\Phi)$ and the $(h-h/2)$-estimator $\|\Phi - \widehat{\Phi}\|_{\mathcal{Y}}$ show both the same sensible convergence behavior for this problem. For uniform refinement, they display the rate $\mathcal{O}(N^{-1/8})$, which is then improved by isotropic refinement to $\mathcal{O}(N^{-1/4})$. Finally, for anistropic

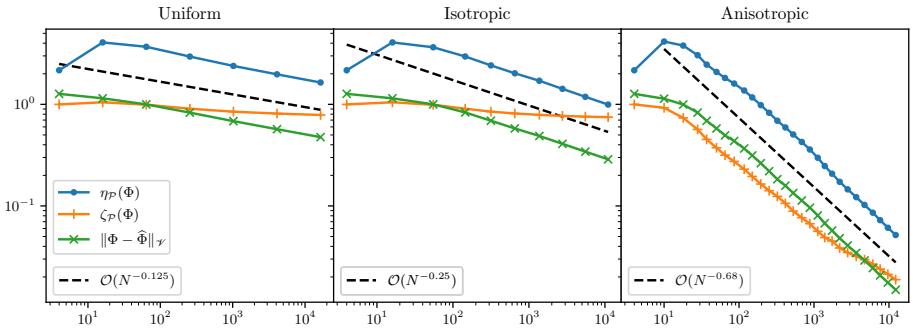


Figure 9.3. Error estimators for the singular problem of Section 9.4.5 plotted double-logarithmically over the degrees of freedom $N = \#\mathcal{P}$: uniform refinement (left), isotropic refinement (middle), and anisotropic refinement (right).

refinement, they achieve the best possible rate $\mathcal{O}(N^{-15/22})$, recovering the rate for a smooth problem. Looking at Figure 9.4, we see strong anisotropic refinement towards $t = 0$ with elements of size $h_{\boldsymbol{x}} = 1, h_t = 2^{-18}$, and some mild refinement towards the corners of the unit square.

On the other hand, the weighted L_2-terms $\zeta_{\mathcal{P}}(\Phi)$ do not seem to decay for uniform or isotropic refinement, and seem to degenerate for anisotropic refinement. This is problematic for the reliability bound in Corollary 9.3.5. Further inspection suggests that this is a theoretical shortcoming rather than a practical one. This is hinted by the $(h - h/2)$-estimator, which one generally assumes to be reliable. Note that this does not contradict the theoretical results from Corollary 9.3.5, which states $\zeta_{\mathcal{P}}(\Phi) \lesssim \eta_{\mathcal{P}}(\Phi) \lesssim \|\phi - \Phi\|_{H^{-1/2,-1/4}(\Sigma)}$ only under the additional parabolic scaling assumption $h_t \eqsim h_{\boldsymbol{x}}^2$ for all space-time elements.

Under this parabolic scaling assumption, the optimal error decay rate for smooth problems becomes $\mathcal{O}(N^{-1/2})$; see (9.15). Figure 9.5 displays the results of uniform and adaptive refinement, with meshes that satisfy this scaling constraint[2], providing convergence rates $\mathcal{O}(N^{-0.18})$ and $\mathcal{O}(N^{-0.4})$, respectively, for *all* considered estimators.

9.4.6 Singular L-shape problem

We consider the L-shaped domain $\Omega := (-1, 1)^2 \setminus [-1, 0]^2$ with data $u_0 \equiv 1$ and $u_D \equiv 0$. The solution has a strong singularity for $t = 0$, in addition to

[2] For uniform refinement, all elements are bisected once in space-direction and three times in time-direction. For adaptive refinement, we assume $\mathcal{M}_\ell^{\boldsymbol{x}} = \mathcal{M}_\ell^t$. All these marked elements are subdivided into four congruent rectangles, where we use additional bisections in space and/or time to guarantee that the level differences between elements sharing an edge is bounded by 1 and that $\frac{1}{2} h_{\boldsymbol{x}}^2 \le h_t \le 2 h_{\boldsymbol{x}}^2$.

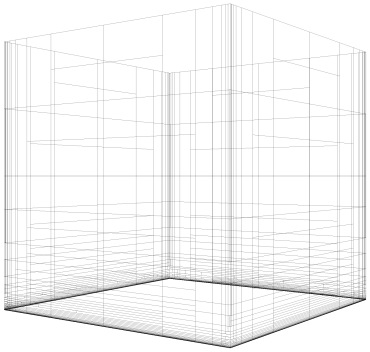


FIGURE 9.4. Mesh with $N = 1391$ elements, generated by anisotropic refinement for the singular problem of Section 9.4.5.

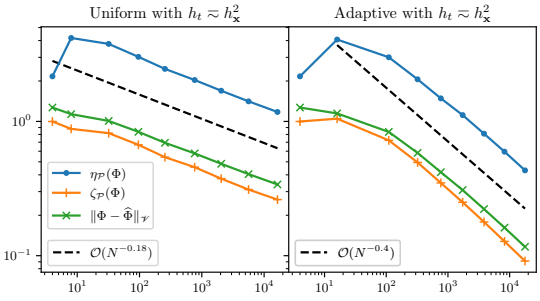


FIGURE 9.5. Error estimators for the singular problem of Section 9.4.5 plotted double-logarithmically over the degrees of freedom $N = \#\mathcal{P}$: uniform refinement (left) and adaptive refinement (right) with parabolic scaling $h_t \eqsim h_{\boldsymbol{x}}^2$.

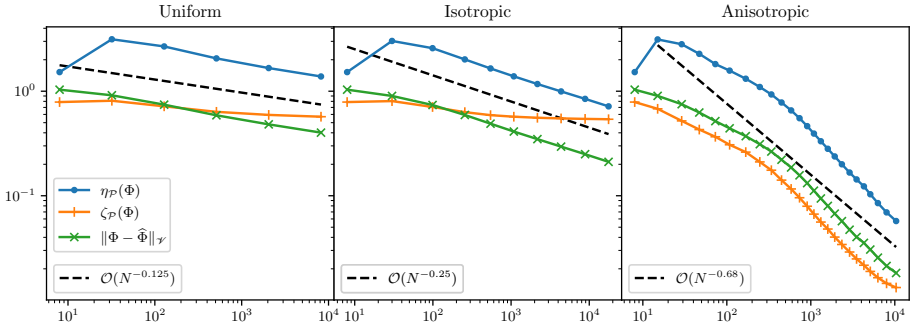


FIGURE 9.6. Error estimators for the singular L-shape problem of Section 9.4.6 plotted double-logarithmically over the degrees of freedom $N = \#\mathcal{P}$: uniform refinement (left), isotropic refinement (middle), and anisotropic refinement (right).

a singularity at the re-entrant corner $(0, 0)$. We choose $\mathcal{P}_0 := \big\{[0, 1] \times K : K \in \mathcal{P}_\Gamma\big\}$, with the uniform mesh $\mathcal{P}_\Gamma$ of Γ being aligned with the corners and consisting of eight elements, as initial mesh of the space-time boundary Σ. Figure 9.6 displays the results, which are similar to those of Section 9.4.5 with a better behavior of the L_2-terms $\zeta_{\mathcal{P}}(\Phi)$ for anisotropic refinement.

Enforcing the parabolic scaling $h_t \eqsim h_{\boldsymbol{x}}^2$ as in Section 9.4.5, *all* estimators converge again with the same rates, being $\mathcal{O}(N^{-0.18})$ for uniform refinement and $\mathcal{O}(N^{-0.45})$ for adaptive refinement (not displayed).

Summary

Partial differential equations are used for the modeling of many (natural) phenomena such as fluid flow, heat dissipation, sound propagation, chemical reactions and the weather. Typically, it is impossible to find closed-form solutions to these differential equations. Instead, one may use numerical methods to construct approximations of the solution to these differential equations. Ideally, these numerical methods provide us with *high quality* approximations and do so *quickly*.

In this thesis we look at such numerical methods for approximating linear partial differential equations. This thesis is structured into two parts. In the first part we study *preconditioning*, a technique used to accelerate iterative solvers for systems of linear equations arising in our approximation schemes. In the second part we focus on (adaptive) numerical methods for parabolic evolution equations in a simultaneous *space-time* approach.

In Part I we study the construction of preconditioners for discretized operators using the concept of *operator preconditioning*. The idea is to precondition the discretized operator by a discretized operator of opposite order. It turns out that in order to get a uniformly well-conditioned system, as well as a preconditioner that can be implemented efficiently, the second discretization has to be carefully chosen dependent on the first one.

In Chapter 2 we apply this idea to construct uniform preconditioners for operators of *negative order* discretized by (dis)continuous piecewise polynomials, of any degree, with respect to some mesh. The application cost of the preconditioner is the cost of the opposite order operator discretized by continuous piecewise linears on the same mesh, plus minor cost that scales linearly in the number of mesh cells. Compared to earlier proposals, our approach has the advantages that it does not require a (barycentric) refinement of the mesh, nor the inversion of a non-diagonal matrix, and that it applies without any mildly grading assumption on the mesh.

In Chapter 3 we propose a multi-level type operator that both fulfills the role of the opposite order operator *and* can be applied in optimal *linear complexity*. So, when it is used within the framework from Chapter 2, it provides a uniform

preconditioner for operators of negative order that can be applied in linear complexity.

In Chapter 4 we investigate the operator preconditioning framework for *positive order* operators, discretized by continuous piecewise polynomials with respect to some mesh. We obtain uniform preconditioners with similar advantages as those found in Chapter 2. That is, we construct uniform preconditioners whose application cost equals the cost of the opposite order operator discretized with discontinuous or continuous finite elements on the same mesh, plus minor cost of linear complexity.

In Chapter 5 we study the preconditioning framework in a more restrictive setting. We build uniform preconditioners for operators discretized by continuous piecewise polynomials as the composition of an opposite order operator, discretized on the same ansatz space, and two identical diagonal scaling operators, whose matrix representation is the lumped mass matrix.

In Part II of this thesis, we discuss the (adaptive) numerical solution of parabolic evolution equations, e.g. the heat equation, written in a simultaneous space-time variational formulation. The 'natural' weak formulation of these problems is not coercive, making it hard to discretize the problem. We consider a minimal residual discretization, which leads to quasi-optimal approximations under an inf-sup stability condition on the pairs of trial- and test spaces.

In Chapter 6 we propose an r-linearly converging adaptive method for parabolic evolution equations that is able to resolve singularities locally in space and time. We achieve this by using trial- and test spaces that are given as the spans of wavelets-in-time tensorized with (locally refined) finite element spaces-in-space. Thanks to this tensor product ansatz, we are able to solve the whole time evolution at the cost of solving the corresponding stationary problem. We also introduce optimal preconditioners, allowing one to solve the discrete problem efficiently. Numerical results illustrate our theoretical findings.

In Chapter 7 we discuss an implementation of the aforementioned adaptive method in which every step is of linear complexity. In particular, we propose a matrix-free algorithm that can apply the system matrices in linear complexity, even though their matrix representation is not sparsely populated.

In Chapter 8 we consider a (time-)parallel algorithm for parabolic evolution equations using trial- and test spaces that are tensor products of temporal and spatial spaces. We present large parallel computations showing the effectiveness of the method in practice.

Finally, in Chapter 9 we propose an adaptive space-time boundary element method for the solution of the homogeneous heat equation with prescribed initial condition and Dirichlet data. We introduce an a posteriori error estimator, and use it to drive an adaptive loop with anisotropic refinement. In all our numerical experiments, the adaptive algorithm recovers the optimal error decay rate.

Samenvatting

Partiële differentiaalvergelijkingen worden gebruikt voor het modelleren van veel (natuurlijke) verschijnselen, zoals de stroming van vloeistoffen, de verspreiding van warmte, de voortplanting van geluid, het verloop van chemische reacties en het weer. In veel realistische gevallen is het onmogelijk om een expliciet functievoorschrift voor de oplossing van de differentiaalvergelijking te vinden. Daarom gebruikt men numerieke methoden om benaderingen van de oplossing uit te rekenen. Idealiter geven deze numerieke methoden *snel* een *goede* benadering.

In dit proefschrift doen we onderzoek naar dergelijke numerieke methoden. Het proefschrift is opgebouwd uit twee delen. In het eerste deel bestuderen we *preconditionering*, een techniek voor het versnellen van iteratieve methoden die gebruikt worden om de stelsels lineaire vergelijkingen in de numerieke methoden op te lossen. In het tweede deel bekijken we het (adaptief) benaderen van parabolische evolutieproblemen, zoals de warmtevergelijking, met ruimte-tijd methoden.

In deel I bekijken we *operatorpreconditionering*: een methode om een gediscretiseerde operator te preconditioneren met een gediscretiseerde operator van tegenovergestelde orde. Om te zorgen dat dit tot uniform goed-geconditioneerde systemen leidt en dat de preconditioneerder efficiënt kan worden toegepast, moet de tweede discretisatieruimte zorgvuldig gekozen worden.

In hoofdstuk 2 passen we dit concept toe om uniforme preconditioneerders te maken voor operatoren van negatieve orde gediscretiseerd met (dis)continue stuksgewijs polynomen, van een willekeurige graad, op een gegeven rooster. De kosten van het toepassen van de preconditioneerder zijn gelijk aan de kosten van de gediscretiseerde operator van tegenovergestelde orde op hetzelfde rooster, plus lage kosten die lineair schalen in het aantal roosterelementen. In vergelijking met methoden uit de literatuur, heeft onze aanpak de voordelen dat er geen (barycentrisch) verfijnd rooster nodig is, dat er geen niet-diagonaal matrix geïnverteerd moet worden en we geen aanname nodig hebben over dat het verschil in grootte van naastliggende roosterelementen voldoende klein is.

In hoofdstuk 3 construeren we een multischaaloperator die de rol van de tegenovergestelde operator vervult én die in optimale (lineaire) complexiteit

kan worden toegepast. In combinatie met de bevindingen uit hoofdstuk 2 geeft dit een uniforme preconditioneerder voor operators van negatieve orde die in lineaire complexiteit kan worden toegepast.

In hoofdstuk 4 bekijken we nogmaals het concept van operatorpreconditionering, ditmaal voor operatoren van positieve orde gediscretiseerd door continue stuksgewijs polynomen op een rooster. We construeren uniforme preconditioneerders met dezelfde voordelen als die uit hoofdstuk 2.

In hoofdstuk 5 bekijken we het concept van operatorpreconditionering in een beperkte setting. We maken uniforme preconditioneerders voor operatoren die gediscretiseerd zijn met continue stuksgewijze polynomen. De gevonden preconditioneerders bestaan uit de tegenovergestelde operator, gediscretiseerd op dezelfde deelruimte, en twee gelijke diagonale schaaloperatoren.

In deel II van dit proefschrift kijken we naar parabolische evolutieproblemen met ruimte-tijd methoden. De natuurlijke zwakke formulering die hoort bij zulke vergelijkingen is niet coercief, wat het lastig maakt om deze problemen te discretiseren. We bekijken een residu-minimalisatie formulering van het probleem die leidt tot quasi-optimale benaderingen onder een inf-sup aanname op de paren van zoek- en testruimten.

In hoofdstuk 6 introduceren we een adaptieve numerieke methode voor parabolische problemen die r-linear convergeert en die singulariteiten lokaal in plaats en tijd kan benaderen. We bereiken dit door gebruik te maken van zoek- en testruimten die opgespannen zijn door het tensorproduct van wavelets in tijd en (lokaal verfijnde) eindige elementenruimten in ruimte. Door deze ruimten te gebruiken, kunnen we de gehele tijdsevolutie oplossen met dezelfde kosten als het oplossen van het bijbehorende tijdsonafhankelijke probleem. We introduceren ook optimale preconditioneerders, zodat het gediscretiseerde probleem efficiënt opgelost kan worden. We illustreren onze theoretische bevindingen met numerieke experimenten.

In hoofdstuk 7 bestuderen we een implementatie van de bovengenoemde adaptieve methode, waarin elke stap in lineaire complexiteit kan worden uitgevoerd. In het bijzonder geven we een algoritme om de systeemmatrices in lineaire complexiteit toe te toepassen, hoewel deze matrices niet ijl zijn.

In hoofdstuk 8 geven we een parallel (in tijd) algoritme voor parabolische vergelijkingen met zoek- en testruimten die het tensorproduct zijn van ruimten in tijd en ruimte. We laten de effectiviteit van de methode in de praktijk zien middels grootschalige berekeningen op een supercomputer.

Tot slot, in hoofdstuk 9 introduceren we een adaptieve ruimte-tijd randelementenmethode voor de homogene warmtevergelijking met gegeven begin- en randvoorwaarden. We introduceren een a posteriori foutschatter, die we vervolgens gebruiken om een adaptief algoritme te sturen met anisotrope verfijning. In de door ons bekeken experimenten, behaalt deze methode de optimale convergentiesnelheid, ook voor problemen die singulier in tijd en ruimte zijn.

Dankwoord

Allereerst wil ik mijn promotor Rob Stevenson bedanken. Ik heb veel van je geleerd. Je hebt mij iedere keer opnieuw verbaasd met jouw (wiskundig) inzicht. Dank voor je geduld, advies, humor, het becommentariëren van talloze stukken tekst en de ruimte die je mij hebt geboden om een half jaar stage te lopen. Je bent een goede mentor en ik ben bevoorrecht dat jij mijn promotor was.

Ik ben dankbaar dat ik dit promotietraject samen met jou, Jan Westerdiep, heb mogen doorlopen. De band die we hebben opgebouwd als medestudenten, collega's en vrienden is bijzonder. Je was er altijd om advies te geven of mijn gedachtestroom aan te horen op zowel inhoudelijk vlak als daarbuiten. De samenwerking met jou heeft mij scherp gehouden, zeker in de coronaperiode. Je ongekende enthousiasme, behulpzaamheid en positiviteit maken je een mooi mens. Dank voor alles.

De collega's van het Korteweg-de Vries instituut bedank ik voor het creëren van een prettige werksfeer. Gregor, je was een fijne kamergenoot. Dank voor je belangstelling, de vele discussies, het kritisch lezen van mijn introductie en je toewijding aan het paper dat wij samen hebben gemaakt. Jan Brandts bedank ik voor zijn rol als copromotor en Marieke voor het mogelijk maken van werken op kantoor tijdens de coronapandemie. Mees, Ruben en Jan, de programmeerwedstrijden waar wij als team aan hebben meegedaan waren vermakelijke maar vaak ook frustrerende uitdagingen waar ik veel van heb geleerd, dank hiervoor.

Ik spreek ook mijn dank uit aan alle vrienden om mij heen. Ontspanning in de vorm van sporten, borrelen, vakantie vieren of bordspellen spelen heeft mij veel energie gegeven. In het bijzonder dank ik Floris en Rick voor alle etentjes, Maarten voor de squashpartijen, Danny en Niki voor de fietstochten en Roos voor het zijn van een fijne huisgenoot.

Mijn familie bedank ik voor alle steun door de jaren heen, met name mijn moeder Monique voor haar betrokkenheid en de aanmoedigende woorden wanneer deze nodig waren.

Tot slot bedank ik mijn vriendin Ingrid. Ik waardeer jouw liefdevolle steun enorm, dankjewel.

Bibliography

[AFF+13] M. Aurada, M. Feischl, T. Führer, M. Karkulik, and D. Praetorius. Efficiency and optimality of some weighted-residual error estimator for adaptive 2D boundary element methods. *Comput. Methods Appl. Math.*, 13(3):305–332, 2013.

[AFF+17] M. Aurada, M. Feischl, T. Führer, M. Karkulik, J.M. Melenk, and D. Praetorius. Local inverse estimates for non-local boundary integral operators. *Math. Comp.*, 86(308):2651–2686, 2017.

[Alp93] B.K. Alpert. A class of bases in L^2 for the sparse representation of integral operators. *SIAM J. Math. Anal.*, 24:246–262, 1993.

[AN87] D.N. Arnold and P.J. Noon. Boundary integral equations of the first kind for the heat equation. In *Boundary elements IX, Vol. 3 (Stuttgart, 1987)*, pages 213–229. Comput. Mech., Southampton, 1987.

[And13] R. Andreev. Stability of sparse space-time finite element discretizations of linear parabolic evolution equations. *IMA J. Numer. Anal.*, 33(1):242–260, 2013.

[And16] R. Andreev. Wavelet-in-time multigrid-in-space preconditioning of parabolic evolution equations. *SIAM J. Sci. Comput.*, 38(1):A216–A242, 2016.

[BC07] A. Buffa and S.H. Christiansen. A dual finite element complex on the barycentric refinement. *Math. Comp.*, 76(260):1743–1769, 2007.

[BD04] P. Binev and R. DeVore. Fast computation in adaptive tree approximation. *Numer. Math.*, 97(2):193 – 217, 2004.

[BDDP02] P. Binev, W. Dahmen, R. DeVore, and P. Petruchev. Approximation classes for adaptive methods. *Serdica Math. J.*, 28:391–416, 2002.

[Beb00] M. Bebendorf. Approximation of boundary element matrices. *Numer. Math.*, 86(4):565–589, 2000.

[BFV19] P. Binev, F. Fierro, and A. Veeser. Near-best adaptive approximation on conforming meshes. *Preprint*, 2019, arXiv:1912.13437.

[BJ89] I. Babuška and T. Janik. The h-p version of the finite element method for parabolic equations. Part I. The p-version in time. *Numer. Methods Partial Differential Equations*, 5(4):363–399, 1989.

[BJ90] I. Babuška and T. Janik. The h-p version of the finite element method for parabolic equations. II. The h-p version in time. *Numer. Methods Partial Differential Equations*, 6(4):343–369, 1990.

[BPV00] J.H. Bramble, J.E. Pasciak, and P.S. Vassilevski. Computational scales of Sobolev norms with application to preconditioning. *Math. Comp.*, 69(230):463–480, 2000.

[Bra81] A. Brandt. Multigrid solvers on parallel computers. In *Elliptic Problem Solvers*, pages 39–83. Elsevier, 1 1981.

[Bra01] D. Braess. *Finite Elements*. Cambridge University Press, 2001. Second edition.

[BRU20] N. Beranek, M.A. Reinhold, and K. Urban. A space-time variational method for optimal control problems. *Preprint*, 2020, arXiv:2010.00345.

[BY14] R.E. Bank and H. Yserentant. On the H^1-stability of the L_2-projection onto finite element spaces. *Numer. Math.*, 126(2):361–381, 2014.

[BZ96] R. Balder and C. Zenger. The solution of multidimensional real Helmholtz equations on sparse grids. *SIAM J. Sci. Comput.*, 17(3):631–646, 1996.

[Car02] C. Carstensen. Merging the Bramble-Pasciak-Steinbach and the Crouzeix-Thomée criterion for H^1-stability of the L^2-projection onto finite element spaces. *Math. Comp.*, 71(237):157–163, 2002.

[Car04] C. Carstensen. An adaptive mesh-refining algorithm allowing for an H^1 stable L^2 projection onto Courant finite element spaces. *Constr. Approx.*, 20(4):549–564, 2004.

[CDD01] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods for elliptic operator equations – Convergence rates. *Math. Comp*, 70:27–75, 2001.

[CN00] S.H. Christiansen and J.-C. Nédélec. Des préconditionneurs pour la résolution numérique des équations intégrales de frontière de l'acoustique. *C. R. Acad. Sci. Paris Sér. I Math.*, 330(7):617–622, 2000.

[Cos90] M. Costabel. Boundary integral operators for the heat equation. *Integral Equations Operator Theory*, 13(4):498–552, 1990.

[CR19] A. Chernov and A. Reinarz. Sparse grid approximation spaces for space–time boundary integral formulations of the heat equation. *Comput. Math. Appl.*, 78(11):3605–3619, 2019.

[CS11] N.G. Chegini and R.P. Stevenson. Adaptive wavelets schemes for parabolic problems: Sparse matrices and numerical results. *SIAM J. Numer. Anal.*, 49(1):182–212, 2011.

[CS13] A. Chernov and Ch. Schwab. Sparse space-time Galerkin BEM for the nonstationary heat equation. *ZAMM Z. Angew. Math. Mech.*, 93(6–7):403–413, 2013.

[Dev20] D. Devaud. Petrov-Galerkin space-time hp-approximation of parabolic equations in $H^{1/2}$. *IMA J. Numer. Anal.*, 40(4):2717–2745, 2020.

[DFG+04] W. Dahmen, B. Faermann, I.G. Graham, W. Hackbusch, and S.A. Sauter. Inverse inequalities on non-quasiuniform meshes and application to the mortar element method. *Math. Comp.*, 73:1107–1138, 2004.

[DGVdZ18] R. Dyja, B. Ganapathysubramanian, and K.G. Van der Zee. Parallel-in-space-time, adaptive finite element framework for nonlinear parabolic equations. *SIAM J. Sci. Comput.*, 40(3):C283–C304, 2018.

[Dij09] T.J. Dijkema. *Adaptive tensor product wavelet methods for the solution of PDEs*. PhD thesis, Utrecht University, 2009.

[DKS16] L. Diening, Ch. Kreuzer, and R.P. Stevenson. Instance optimality of the adaptive maximum strategy. *Found. Comput. Math.*, 16(1):33–68, 2016.

[DL92] R. Dautray and J.-L. Lions. *Mathematical analysis and numerical methods for science and technology. Vol. 5*. Springer-Verlag, Berlin, 1992.

[DNS19] S. Dohr, K. Niino, and O. Steinbach. Space-time boundary element methods for the heat equation. In *Space-Time Methods: Applications to Partial Differential Equations*, pages 1–60. De Gruyter, 2019.

[Doh19] S. Dohr. *Distributed and Preconditioned Space–Time Boundary Element Methods for the Heat Equation*. PhD thesis, TU Graz, 2019.

[DPS05] L. Dalcín, R. Paz, and M. Storti. MPI for Python. *J. Parallel Distrib. Comput.*, 65(9):1108–1115, 9 2005.

[DS18] D. Devaud and Ch. Schwab. Space-time hp-approximation of parabolic equations. *Calcolo*, 55(3):Paper No. 35, 23, 2018.

[DS20] L. Diening and J. Storn. A Space-Time DPG Method for the Heat Equation. *Preprint*, 2020, arXiv:2012.13229.

[DST20] L. Diening, J. Storn, and T. Tscherpel. On the Sobolev and L^p-stability of the L^2-projection. *Preprint*, 2020, arXiv:2008.01801.

[DSW21] W. Dahmen, R.P. Stevenson, and J. Westerdiep. Accuracy controlled data assimilation for parabolic problems. *Preprint*, 2021, arXiv:2105.05836.

[DZO+19] S. Dohr, J. Zapletal, G. Of, M. Merta, and M. Kravčenko. A parallel space–time boundary element method for the heat equation. *Comput. Math. Appl.*, 78(9):2852–2866, 2019.

[EG04] A. Ern and J.-L. Guermond. *Theory and practice of finite elements*, volume 159 of *Applied Mathematical Sciences*. Springer, New York, 2004.

[ESV17] A. Ern, I. Smears, and M. Vohralík. Guaranteed, locally space-time efficient, and polynomial-degree robust a posteriori error estimates for high-order discretizations of parabolic problems. *SIAM J. Numer. Anal.*, 55(6):2811–2834, 2017.

[Fae00] B. Faermann. Localization of the Aronszajn-Slobodeckij norm and application to adaptive boundary elements methods. Part I. The two-dimensional case. *IMA J. Numer. Anal.*, 20(2):203–234, 2000.

[Fae02] B. Faermann. Localization of the Aronszajn-Slobodeckij norm and application to adaptive boundary element methods. Part II. The three-dimensional case. *Numer. Math.*, 92(3):467–499, 2002.

[FFK+14] R. D. Falgout, S. Friedhoff, Tz. V. Kolev, S. P. MacLachlan, and J. B. Schroder. Parallel Time Integration with Multigrid. *SIAM J. Sci. Comput.*, 36(6):C635–C661, 1 2014.

[FH19] T. Führer and N. Heuer. Optimal quasi-diagonal preconditioners for pseudodifferential operators of order minus two. *J. Sci. Comput.*, 79(2):1161–1181, 2019.

[FHPS19] T. Führer, A. Haberl, D. Praetorius, and S. Schimanko. Adaptive BEM with inexact PCG solver yields almost optimal computational costs. *Numer. Math.*, 141(4):967–1008, 2019.

[FK21] T. Führer and M. Karkulik. Space–time least-squares finite elements for parabolic equations. *Comput. Math. Appl.*, 92:27–36, 6 2021.

[For77] M. Fortin. An analysis of the convergence of mixed finite element methods. *RAIRO Anal. Numér.*, 11(4):341–354, iii, 1977.

[Gan15] M.J. Gander. 50 years of time parallel time integration. In *Multiple shooting and time domain decomposition methods*, volume 9 of *Contrib. Math. Comput. Sci.*, pages 69–113. Springer, Cham, 2015.

[Gan17] G. Gantner. *Optimal adaptivity for splines in finite and boundary element methods*. PhD thesis, TU Wien, 2017.

[GHS05] I.G. Graham, W. Hackbusch, and S.A. Sauter. Finite elements on degenerate meshes: inverse-type inequalities and applications. *IMA J. Numer. Anal.*, 25(2):379–407, 2005.

[GHS16] F. D. Gaspoz, C.-J. Heine, and K. G. Siebert. Optimal grading of the newest vertex bisection and H^1-stability of the L_2-projection. *IMA J. Numer. Anal.*, 36(3):1217–1241, 2016.

[GK11] M.D. Gunzburger and A. Kunoth. Space-time adaptive wavelet methods for control problems constrained by parabolic evolution equations. *SIAM J. Contr. Optim.*, 49(3):1150–1170, 2011.

[Glä12] M. Gläfke. *Adaptive methods for time domain boundary integral equations*. PhD thesis, Brunel University London, 2012.

[GN16] M.J. Gander and M. Neumüller. Analysis of a new space-time parallel multigrid algorithm for parabolic problems. *SIAM J. Sci. Comput.*, 38(4):A2173–A2208, 2016.

[GO07] M. Griebel and D. Oeltz. A sparse grid space-time discretization scheme for parabolic problems. *Computing*, 81(1):1–34, 2007.

[GÖSS20] H. Gimperlein, C. Özdemir, D. Stark, and E.P. Stephan. A residual a posteriori error estimate for the time–domain boundary element method. *Numer. Math.*, 146(2):239–280, 2020.

[GP20] G. Gantner and D. Praetorius. Adaptive BEM for elliptic PDE systems, part I: abstract framework, for weakly-singular integral equations. *Appl. Anal.*, published online:1–34, 2020.

[GS19] H. Gimperlein and J. Stocek. Space-time adaptive finite elements for nonlocal parabolic variational inequalities. *Comput. Methods Appl. Mech. Engrg.*, 352:137–171, 2019.

[GS21] G. Gantner and R.P. Stevenson. Further results on a space-time FOSLS formulation of parabolic PDEs. *ESAIM Math. Model. Numer. Anal.*, 55:283–299, 2021.

[GvV21] G. Gantner and R. van Venetië. Adaptive space-time BEM for the heat equation. *Preprint*, 2021, arXiv:2108.03055.

[Hac85] W. Hackbusch. *Multi-Grid Methods and Applications*, volume 4 of *Springer Series in Computational Mathematics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1985.

[Hac99] W. Hackbusch. A sparse matrix arithmetic based on $\mathcal{H}$-matrices. Part i: Introduction to $\mathcal{H}$-matrices. *Computing*, 62(2):89–108, 1999.

[Hac12] W. Hackbusch. *Tensor Spaces and Numerical Tensor Calculus*, volume 42 of *Springer Series in Computational Mathematics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[Hip06] R. Hiptmair. Operator preconditioning. *Comput. Math. Appl.*, 52(5):699–706, 2006.

[HJHUT18] R. Hiptmair, C. Jerez-Hanckes, and C. Urzúa-Torres. Closed-form inverses of the weakly singular and hypersingular operators on disks. *Integral Equations Operator Theory*, 90(1):Art. 4, 14, 2018.

[HJHUT20] R. Hiptmair, C. Jerez-Hanckes, and C. Urzúa-Torres. Optimal operator preconditioning for Galerkin boundary element methods on 3-dimensional screens. *SIAM J. Numer. Anal.*, 58(1):834–857, 2020.

[HK12] R. Hiptmair and L. Kielhorn. BETL – A generic boundary element template library. Technical Report 2012-36, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2012.

[HLNS19] Ch. Hofer, U. Langer, M. Neumüller, and R. Schneckenleitner. Parallel and robust preconditioning for space-time isogeometric analysis of parabolic evolution problems. *SIAM J. Sci. Comput.*, 41(3):A1793–A1821, 2019.

[HT18] H. Harbrecht and J. Tausch. A fast sparse grid based space–time boundary element method for the nonstationary heat equation. *Numer. Math.*, 140(1):1–26, 2018.

[HUT16] R. Hiptmair and C. Urzúa-Torres. Dual mesh operator preconditioning on 3d screens: Low-order boundary element discretization. Technical Report 2016-14, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2016.

[HVW95] G. Horton, S. Vandewalle, and P. Worley. An Algorithm with Polylog Parallel Complexity for Solving Parabolic Partial Differential Equations. *SIAM J. Sci. Comput.*, 16(3):531–541, 5 1995.

[Kat60] T. Kato. Estimation of iterated matrices, with application to the von Neumann condition. *Numer. Math.*, 2:22–29, 1960.

[KS08] Y. Kondratyuk and R.P. Stevenson. An optimal adaptive finite element method for the Stokes problem. *SIAM J. Numer. Anal.*, 46(2):747–775, 2008.

[KS14] S. Kestler and R.P. Stevenson. Fast evaluation of system matrices w.r.t. multi-tree collections of tensor product refinable basis functions. *J. Comput. Appl. Math.*, 260:103–116, 2014.

[KSU16] S. Kestler, K. Steih, and K. Urban. An efficient space-time adaptive wavelet Galerkin method for time-periodic parabolic partial differential equations. *Math. Comp.*, 85(299):1309–1333, 2016.

[LM17] S. Larsson and M. Molteni. Numerical solution of parabolic problems based on a weak space-time formulation. *Comput. Methods Appl. Math.*, 17(1):65–84, 2017.

[LMN16] U. Langer, S.E. Moore, and M. Neumüller. Space-time isogeometric analysis of parabolic evolution problems. *Comput. Methods Appl. Mech. Engrg.*, 306:342–363, 2016.

[LMT01] J.-L. Lions, Y. Maday, and G. Turinici. Résolution d'EDP par un schéma en temps «pararéel ». *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 332(7):661–668, 4 2001.

[LS20] U. Langer and A. Schafelner. Adaptive Space-Time Finite Element Methods for Non-autonomous Parabolic Problems with Distributional Sources. *Comput. Methods Appl. Math.*, 20(4):677–693, 2020.

[Mau95] J. Maubach. Local bisection refinement for n-simplicial grids generated by reflection. *SIAM J. Sci. Comput.*, 16(1):210–227, 1995.

[MFL+91] O.A McBryan, P.O Frederickson, J. Lindenand, A. Schüller, K. Solchenbach, K. Stüben, C.-A. Thole, and U. Trottenberg. Multigrid methods on parallel computers—a survey of recent developments. *Impact Comput. Sci. Engrg.*, 3(1):1–75, 1991.

[MST14] M. Messner, M. Schanz, and J. Tausch. A fast Galerkin method for parabolic space–time boundary integral equations. *J. Comput. Phys.*, 258:15–30, 2014.

[MST15] M. Messner, M. Schanz, and J. Tausch. An efficient Galerkin boundary element method for the transient heat equation. *SIAM J. Sci. Comput.*, 37(3):A1554–A1576, 2015.

[Nie64] J. Nievergelt. Parallel methods for integrating ordinary differential equations. *Comm. ACM*, 7:731–733, 1964.

[Nit06] P.-A. Nitsche. Best N-term approximation spaces for tensor product wavelet bases. *Constr. Approx.*, 24(1):49–70, 2006.

[Noo88] P.J. Noon. *The single layer heat potential and Galerkin boundary element methods for the heat equation*. PhD thesis, University of Maryland, 1988.

[NS19] M. Neumüller and I. Smears. Time-parallel iterative solvers for parabolic evolution equations. *SIAM J. Sci. Comput.*, 41(1):C28–C51, 2019.

[OR00] M.A. Olshanskii and A. Reusken. On the convergence of a multigrid method for linear reaction-diffusion problems. *Computing*, 65(3):193–202, 2000.

[Osw94] P. Oswald. *Multilevel finite element approximation: Theory and applications*. B.G. Teubner, Stuttgart, 1994.

[Pab15] R. Pabel. *Adaptive Wavelet Methods for Variational Formulations of Nonlinear Elliptic PDEs on Tensor-Product Domains*. PhD thesis, Universität zu Köln, 2015.

[Pfl10] D. Pflüger. *Spatially Adaptive Sparse Grids for High-Dimensional Problems*. PhD thesis, Technische Universität München, 2010.

[PP20] C.-M. Pfeiler and D. Praetorius. Dörfler marking with minimal cardinality is a linear complexity problem. *Math. Comp.*, 89(326):2735–2752, 2020.

[PW12] J. W. Pearson and A. J. Wathen. A new approximation of the Schur complement in preconditioners for PDE-constrained optimization. *Numer. Linear Algebra Appl.*, 19(5):816–829, 2012.

[Rek18] N. Rekatsinas. *Optimal adaptive wavelet methods for solving first order system least squares*. PhD thesis, University of Amsterdam, 2018.

[RS18] N. Rekatsinas and R.P. Stevenson. A quadratic finite element wavelet Riesz basis. *Int. J. Wavelets Multiresolut. Inf. Process.*, 16(4):1850033, 17, 2018.

[RS19] N. Rekatsinas and R. Stevenson. An optimal adaptive tensor product wavelet solver of a space-time FOSLS formulation of parabolic evolution problems. *Adv. Comput. Math.*, 45(2):1031–1066, 4 2019.

[ŚBA+15] W. Śmigaj, T. Betcke, S. Arridge, J. Phillips, and M. Schweiger. Solving boundary integral problems with BEM++. *ACM Trans. Math. Software*, 41(2):Art. 6, 40, 2015.

[Sch14] Joachim Schöberl. C++11 Implementation of Finite Elements in NGSolve. Technical report, Institute for Analysis and Scientific Computing, Vienna University of Technology, 2014.

[Smi00] R.N.L. Smith. Direct Gauss quadrature formulae for logarithmic singularities on isoparametric elements. *Eng. Anal. Bound. Elem.*, 24(2):161–167, 2000.

[SS09] Ch. Schwab and R.P. Stevenson. Space-time adaptive wavelet methods for parabolic evolution problems. *Math. Comp.*, 78(267):1293–1318, 2009.

[Ste96] R.P. Stevenson. The frequency decomposition multi-level method: A robust additive hierarchical basis preconditioner. *Math. Comp.*, 65(215):983–997, July 1996.

[Ste98] R.P. Stevenson. Stable three-point wavelet bases on general meshes. *Numer. Math.*, 80(1):131–158, 1998.

[Ste02] O. Steinbach. On a generalized L_2 projection and some related stability estimates in Sobolev spaces. *Numer. Math.*, 90(4):775–786, 2002.

[Ste03a] O. Steinbach. *Stability estimates for hybrid coupled domain decomposition methods*, volume 1809 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2003.

[Ste03b] R.P. Stevenson. Locally supported, piecewise polynomial biorthogonal wavelets on nonuniform meshes. *Constr. Approx.*, 19(4):477–508, 2003.

[Ste08a] Olaf Steinbach. *Numerical approximation methods for elliptic boundary value problems*. Springer, New York, 2008. Finite and boundary elements, Translated from the 2003 German original.

[Ste08b] R.P. Stevenson. The completion of locally refined simplicial partitions created by bisection. *Math. Comp.*, 77:227–241, 2008.

[Ste15] O. Steinbach. Space-Time Finite Element Methods for Parabolic Problems. *Comput. Methods Appl. Math.*, 15(4):551–566, 2015.

[SU09] W. Sickel and T. Ullrich. Tensor products of Sobolev-Besov spaces and applications to approximation from the hyperbolic cross. *J. Approx. Theory*, 161:748–786, 2009.

[SvV20a] R.P. Stevenson and R. van Venetië. Uniform preconditioners for problems of negative order. *Math. Comp.*, 89(322):645–674, 2020.

[SvV20b] R.P. Stevenson and R. van Venetië. Uniform preconditioners for problems of positive order. *Comput. Math. Appl.*, 79(12):3516–3530, 2020.

[SvV21a] R.P. Stevenson and R. van Venetië. Uniform Preconditioners of Linear Complexity for Problems of Negative Order. *Comput. Methods Appl. Math.*, 21(2):469–478, 2021.

[SvV21b] R.P. Stevenson and R. van Venetië. Operator preconditioning: the simplest case. *Preprint*, 2021, arXiv:2102.11951.

[SvVW21] R.P. Stevenson, R. van Venetië, and J. Westerdiep. A wavelet-in-time, finite element-in-space adaptive method for parabolic evolution equations. *Preprint*, arXiv:2101.03956, 2021.

[SW98] O. Steinbach and W. L. Wendland. The construction of some efficient preconditioners in the boundary element method. *Adv. Comput. Math.*, 9(1-2):191–216, 1998.

[SW21a] R.P. Stevenson and J. Westerdiep. Minimal residual space-time discretizations of parabolic equations: Asymmetric spatial operators. *Preprint*, 2021, arXiv:2106.01090.

[SW21b] R.P. Stevenson and J. Westerdiep. Stability of Galerkin discretizations of a mixed space-time variational formulation of parabolic evolution equations. *IMA J. Numer. Anal.*, 41(1):28–47, 2021.

[SY18] O. Steinbach and H. Yang. Comparison of algebraic multigrid methods for an adaptive space-time finite-element discretization of the heat equation in 3D and 4D. *Numer. Linear Algebra Appl.*, 25(3):e2143, 17, 2018.

[SZ90] L. R. Scott and S. Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, 54(190):483–493, 1990.

[SZ20] O. Steinbach and M. Zank. Coercive space-time finite element methods for initial boundary value problems. *Electron. Trans. Numer. Anal.*, 52:154–194, 2020.

[Szy06] D.B. Szyld. The many proofs of an identity on the norm of oblique projections. *Numer. Algorithms*, 42(3-4):309–323, 2006.

[Tau19] J. Tausch. Nyström method for BEM of the heat equation with moving boundaries. *Adv. Comput. Math.*, 45(5):2953–2968, 2019.

[Tra97] C.T. Traxler. An algorithm for adaptive mesh refinement in n dimensions. *Computing*, 59(2):115–137, 1997.

[Vir20] P. Virtanen. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*, 17(3):261–272, 3 2020.

[VS18] F. J. Vermolen and A. Segal. On an integration rule for products of barycentric coordinates over simplexes in $\mathbb{R}^n$. *J. Comput. Appl. Math.*, 330:289–294, 2018.

[vVW21a] R. van Venetië and J. Westerdiep. A parallel algorithm for solving linear parabolic evolution equations. In *9th Parallel-in-Time Workshop*, 2021, arXiv:2009.08875.

[vVW21b] R. van Venetië and J. Westerdiep. Efficient space-time adaptivity for parabolic evolution equations using wavelets in time and finite elements in space. *Preprint*, 2021, arXiv:2104.08143.

[vVW21c] R. van Venetië and J. Westerdiep. Implementation of: A parallel algorithm for solving linear parabolic evolution equations, 2021, zenodo:4475959.

[vVW21d] R. van Venetië and J. Westerdiep. Implementation of: Efficient space-time adaptivity for parabolic evolution equations using wavelets in time and finite elements in space, 2021, zenodo:4697250.

[Wat87] A. J. Wathen. Realistic eigenvalue bounds for the Galerkin mass matrix. *IMA Journal of Numerical Analysis*, 7(4):449–457, 1987.

[Wlo82] J. Wloka. *Partielle Differentialgleichungen*. B. G. Teubner, Stuttgart, 1982. Sobolevräume und Randwertaufgaben.

[Wor91] P.H. Worley. Limits on parallelism in the numerical solution of linear partial differential equations. *SIAM J. Sci. Statist. Comput.*, 12(1):1–35, 1991.

[WZ17] J. Wu and H. Zheng. Uniform convergence of multigrid methods for adaptive meshes. *Appl. Numer. Math.*, 113:109–123, 2017.

[XZ03] J. Xu and L. Zikatanov. Some observations on Babuška and Brezzi theories. *Numer. Math.*, 94(1):195–202, 2003.

[ZMD+11] J. Zitelli, I. Muga, L. Demkowicz, J. Gopalakrishnan, D. Pardo, and V. M. Calo. A class of discontinuous Petrov-Galerkin methods. Part IV: the optimal test norm and time-harmonic wave propagation in 1D. *J. Comput. Phys.*, 230(7):2406–2432, 2011.

[ZWOM21] J. Zapletal, R. Watschinger, G. Of, and M. Merta. Semi-analytic integration for a parallel space-time boundary element method modeling the heat equation. *Preprint*, arXiv:2102.09811, 2021.