FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Petr Vacek

Multilevel methods

Department of Numerical Mathematics

Supervisor of the master thesis: prof. Ing. Zdeněk Strakoš, DrSc.
Study programme: Mathematics
Study branch: Numerical and Computational Mathematics

Prague 2020

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In Prague, 6.1.2020 signature of the author

Title: Multilevel methods

Author: Petr Vacek

Department: Department of Numerical Mathematics

Supervisor: prof. Ing. Zdeněk Strakoš, DrSc., Department of Numerical Mathematics

Abstract: The analysis of the convergence behavior of the multilevel methods is in the literature typically carried out under the assumption that the problem on the coarsest level is solved exactly. The aim of this thesis is to present a description of the multilevel methods which allows inexact solve on the coarsest level and to revisit selected results presented in literature using these weaker assumptions. In particular, we focus on the derivation of the uniform bound on the rate of convergence. Moreover, we discuss the possible dependence of the convergence behavior on the mesh size of the initial triangulation.

Keywords: multilevel methods, convergence, numerical stability, linear algebraic systems

I would like to thank my supervisor, Zdeněk Strakoš, for his many helpful comments and suggestions. I am grateful to Jan Papež and Tomáš Gergelits for helping me overcome many difficulties that occurred during writing this thesis. Most of all, I would like to thank my family and my girlfriend for their lasting support. Without their love, care, and encouragement this thesis would not have been written.

Contents

Introduction

Numerical solution of real-world problems is a process that typically consists of several stages. First, the problem is described using a mathematical model. After investigating the existence and uniqueness of its solution, the model is discretized, resulting in a finite-dimensional problem, which can be formulated as a system of algebraic equations. Finally, the solution of the algebraic system is numerically approximated using an appropriate numerical method implemented on a computer. In the case of direct methods the inaccuracy is determined by the input data, machine precision and the numerical stability of the method. For large algebraic systems the computation is usually carried out using an iterative solver that is stopped (truncated) using stopping criteria that ideally allow to control the distance of the computed approximate solution to the exact solution of the algebraic problem.

The individual stages should not be seen in isolation but rather as parts of the whole process. For example, in order to analyze the convergence behavior of an iterative solver and derive its reliable stopping criteria one should combine the aspects of all stages in the process. This matter is expressed, e.g., in the Introduction of the book [17, p. 6] by Liesen and Strakoš:

> *Proper understanding of the interaction between all stages in solving any real-world problem is fundamental for keeping the right perspective while working within any specific stage. Isolating the algebraic stage (...) leads to misunderstandings and misconceptions. For example, sometimes the computational stage is considered a routine application of a solver from some software package, and (...) the computational errors (both the truncation and the roundoff parts) are completely ignored, as though computers would give accurate solutions. In other cases, stopping criteria and convergence evaluation of iterative solvers are based on unrealistic assumptions, which can not be applied to practical computations.*

Analogous discussion can be found, with references to literature published previously, in the introduction of [18].

In this thesis we consider numerical solution of boundary value problems (BVPs) described using linear partial differential equations (PDEs). After discretizing the equations using the Galerkin finite element method (FEM), the resulting system of linear algebraic equations is solved by a multilevel method.

A multilevel method can be considered as a sequence of iterations performed using a sequence of refined meshes. The approximate solution is typically computed using smoothing procedures and a solver on the coarsest level. For an introduction to multilevel (multigrid) methods see, e.g., [6], [11, Chapter 13]. A short historical survey describing the development of convergence analysis can be found, e.g., in the introduction of [27].

Convergence analysis of multilevel methods and derivation of its stopping criteria is in literature typically done under the assumption that the coarsest level problem is solved exactly; see e.g., the convergence analysis in [26, 27], and the derivation of the stopping criteria in [2, Section 5]. This assumption is, however, not satisfied in practical computation either due to the use of an iterative solver on the coarsest level, or due to the finite precision arithmetic, or both; see, e.g., [16, Section 2.5.1, Chapter 5–7]. The aim of this thesis is to revisit selected results presented in literature using weaker assumptions on the solver on the coarsest level. In particular, we focus on the convergence analysis and derivation of the uniform bound on the rate of convergence stated in [26, 27].

The text is organized as follows. In the first chapter we introduce the model BVP problem described using second order linear elliptic PDE and its discretization using the Galerkin finite element method with piecewise-linear basis functions. Chapter 2 contains an abstract description of multilevel methods. After formulating an infinite-dimensional abstract problem and introducing its discretization, we describe the multilevel framework and the multilevel V-cycle schemes in both the operator and matrix-vector formulations. A bound on the rate of convergence of the multilevel V-cycle methods that does not dependent of the number of levels is present in Chapter 3. In Chapter 4 we consider application of the multilevel methods described in Chapters 2 and 3 to the model problem from Chapter 1. We discuss both exact and inexact solvers on the coarsest level and investigate the possibly different convergence behavior. Conclusions formulate the main points. The Appendix contains the detailed proof of the contraction property of the error propagation operator, which corresponds to the V-cycle scheme method.

Throughout this thesis we assume exact arithmetic computations, i.e., the inexactness that is considered is due to truncation of the algorithmic operations and it is not affected by rounding errors.

1. Model problem and its discretization

In this chapter we briefly introduce the function spaces used in the text. Further, we state the model problem and describe its discretization using the Galerkin finite element method (FEM).

1.1 Function spaces

This section gives an overview of the notation and function spaces used in the text. For general references see, e.g., [5, 13, 10, 1]. Our presentation is based on [18, Chapter 1] and [5].

For any real Hilbert space V, with the scalar product and the associated norm

$$(\cdot,\cdot)_V : V \times V \to \mathbb{R}, \quad \text{and} \quad \|\cdot\|_V = \sqrt{(\cdot,\cdot)_V},$$

let $V^{\#}$ denotes its *dual space*, i.e., the space of bounded linear functionals on V. Further let

$$\langle \cdot,\cdot \rangle : V^{\#} \times V \to \mathbb{R},$$

stand for the duality pairing and let $\|\cdot\|_{V^{\#}}$ be the *dual norm*, i.e.,

$$\|f\|_{V^{\#}} = \sup_{v \in V; \|v\|_V = 1} |\langle f, v \rangle| \quad \text{for all } f \in V^{\#}.$$

Let $\Omega \subset \mathbb{R}^d$ be an open, connected, bounded, polyhedral set with *Lipschitz boundary*. The boundary $\partial\Omega$ is said to be *Lipschitz*, cf. [18, p.10] and [13, Definition 1.2.1.1], if there is $\ell \in \mathbb{N}$ and the numbers $\alpha_1 > 0$ and $\alpha_2 > 0$, such that, the boundary is described by ℓ mutually overlapping Lipschitz maps $\varrho_1, \ldots, \varrho_\ell$, such that, for each map $\varrho \in \{\varrho_1, \ldots, \varrho_\ell\}$, upon appropriately reorienting the coordinate axis, the sets

$$\{x \in \mathbb{R}^d; \quad \max_{i=1,\ldots,d-1} |x_i| \le \alpha_1, \quad \varrho(x_1, \ldots, x_{d-1}) < x_d \le \varrho(x_1, \ldots, x_{d-1}) + \alpha_2\}$$

are subsets of Ω and the sets

$$\{x \in \mathbb{R}^d; \quad \max_{i=1,\ldots,d-1} |x_i| \le \alpha_1, \quad \varrho(x_1, \ldots, x_{d-1}) - \alpha_2 < x_d \le \varrho(x_1, \ldots, x_{d-1})\}$$

are contained in $\mathbb{R}^d \setminus \overline{\Omega}$.

For an integer $k \ge 0$, $C^k(\Omega)$ denotes the space of k times continuously differentiable functions in Ω, i.e., the space of functions u such that, for any multiindex $\alpha = (\alpha_1, \ldots, \alpha_d)$ with integers $\alpha_i \ge 0$, $|\alpha| := \sum_{i=1}^d \alpha_i \le k$,

$$D^\alpha u := \frac{\partial^{|\alpha|} u}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}$$

is a continuous function in Ω; $C^\infty(\Omega) := \cap_{k=1}^\infty C^k(\Omega)$ and $C_c^\infty(\Omega)$ consist of functions from $C^\infty(\Omega)$ with compact support in Ω. $C^k(\overline{\Omega})$ is the space of functions u from $C^k(\Omega)$ such that, for any multiindex α, $|\alpha| \le k$, the function $D^\alpha u$ admits a continuous extension to $\overline{\Omega}$.

Lebesgue spaces

For $1 \le p < \infty$, the *Lebesgue space* $L^p(\Omega)$ with the norm $\|\cdot\|_{L^p}$ is defined as (see, e.g., [5, Chapter 4])

$$L^p(\Omega) := \{u : \Omega \to \mathbb{R}; u \text{ is measurable and } \|u\|_{L^p(\Omega)} := \Big(\int_\Omega |u|^p\Big)^{1/p} < \infty\},$$
$$L^\infty(\Omega) := \{u : \Omega \to \mathbb{R}; u \text{ is measurable and } \|u\|_{L^\infty} < \infty\},$$

where $\|\cdot\|_{L^\infty(\Omega)}$ is the proper generalization of the maximum norm to measurable functions. The technical difference is that the values of a function on a set of measure zero do not affect the value of the $\|\cdot\|_{L^\infty(\Omega)}$ norm, i.e., (with $|\Upsilon|_d$ denoting the d-dimensional Lebesgue measure of $\Upsilon \subset \Omega$)

$$\|u\|_{L^\infty} := \inf_{\{\Upsilon\subset\Omega;|\Upsilon|_d=0\}} \sup_{\{x\in\Omega\setminus\Upsilon\}} \{|u(x)| < \infty\}.$$

Sobolev spaces

For an integer $k \ge 0$, the *Sobolev space* $H^k(\Omega)$ with the norm $\|\cdot\|_{H^k}$ consists of functions $u \in L^2(\Omega)$, for which any *weak derivative* up to the order k belongs to $L^2(\Omega)$, i.e., for any multiindex α, $|\alpha| \le k$ there is $g_\alpha \in L^2(\Omega)$ such that

$$\int_\Omega u D^\alpha \varphi = (-1)^{|\alpha|} \int_\Omega g_\alpha \varphi \quad \text{for all } \varphi \in C_c^\infty(\Omega);$$

see, e.g., [5, Section 9.1]. It is usual to write $D^\alpha u$ instead of g_α, ∇u then denotes the row[1] vector of the first weak partial derivatives. In this formalism

$$H^k(\Omega) := \{u \in L^2(\Omega); D^\alpha u \in L^2(\Omega), \text{for all } \alpha; |\alpha| \le k\}$$

and

$$\|u\|_{H^k} := \Bigg(\sum_{|\alpha|\le k} \|D^\alpha u\|^2_{L^2(\Omega)}\Bigg)^{\frac{1}{2}}. \tag{1.1}$$

The spaces $L^2(\Omega)$ and $H^1(\Omega)$ are Hilbert spaces with the inner products

$$(u,v)_{L^2} := \int_\Omega uv \quad \text{and} \quad (u,v)_{H^1} := \int_\Omega (uv + \nabla u \cdot \nabla v) = \int_\Omega \Bigg(uv + \sum_{i=1}^d \frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_i}\Bigg);$$

see, e.g., [5, Proposition 9.1]. We shall also make use of the H^1-*seminorm* (see, e.g., [1, Section 4.29])

$$|u|_{H^1} := \Big(\int_\Omega \nabla u \cdot \nabla u\Big)^{\frac{1}{2}} = \|\nabla u\|_{L^2}.$$

The mapping (see, e.g., [13, Theorem 1.5.1.3])

$$u \mapsto u|_{\partial\Omega},$$

[1] Following the notation in [18], vectors with components corresponding to the individual dimensions in $\mathbb{R}^d$ are *row vectors*. On the contrary, algebraic vectors associated with the discrete algebraic formulations of various problems using matrix representations are *column vectors*.

which is defined for continuous functions on $\overline{\Omega}$, has a unique continuous extension as an operator from $H^1(\Omega)$ to $L^2(\partial\Omega)$ called the *trace operator*. Moreover the *trace inequality* (cf. [13, Theorem 1.5.1.10]) says that there is a positive constant C depending only on Ω such that

$$\|u\|_{L^2(\partial\Omega)} \leq C\|u\|_{H^1}, \quad \text{for all } u \in H^1(\Omega). \tag{1.2}$$

Further, $H_0^1(\Omega)$ denotes the space consisting of functions from $H^1(\Omega)$ having a zero trace on the boundary, i.e.,

$$H_0^1(\Omega) := \left\{v \in H^1(\Omega); u|_{\partial\Omega} = 0\right\}.$$

There is a positive constant C depending only on Ω, such that

$$\|u\|_{L^2} \leq C|u|_{H^1}, \quad \text{for all } u \in H_0^1(\Omega); \tag{1.3}$$

see, e.g., [5, Corollary 9.19]. This inequality is known as the *Poincaré–Fridrichs inequality* and it implies that $|\cdot|_{H^1}$ is a norm on $H_0^1(\Omega)$ topologically equivalent to $\|\cdot\|_{H^1}$.

1.2 Model problem

Following [26], we consider the subsequent second order elliptic boundary-value problem. Given $\Omega \subset \mathbb{R}^2$, $\tilde{a} : \overline{\Omega} \to \mathbb{R}$ and $f : \Omega \to \mathbb{R}$ find $u : \overline{\Omega} \to \mathbb{R}$ satisfying

$$\begin{aligned} -\nabla \cdot (\tilde{a}\nabla u) = f \quad &\text{in} \quad \Omega, \\ u = 0 \quad &\text{on} \quad \partial\Omega, \end{aligned} \tag{1.4}$$

where

PDE 1: $\Omega \subset \mathbb{R}^2$, is an open, bounded, connected, polygonal set with Lipschitz boundary,

PDE 2: $\tilde{a} \in C^1(\overline{\Omega})$ is *uniformly positive*, i.e., there is a positive constant $c_{\tilde{a}}$ such that

$$0 < c_{\tilde{a}} \leq \tilde{a}(x), \quad \text{for all } x \in \Omega,$$

PDE 3: $f \in L^2(\Omega)$.

Defining the linear operator

$$\mathcal{A} : H_0^1(\Omega) \to \left(H_0^1(\Omega)\right)^{\#}, \qquad \langle \mathcal{A}u, v\rangle = \int_\Omega (\tilde{a}\nabla u) \cdot \nabla v,$$

and the linear functional

$$b \in \left(H_0^1(\Omega)\right)^{\#}, \qquad \langle b, v\rangle := \int_\Omega fv,$$

the *weak formulation* of (1.4) reads as: find $u \in H_0^1(\Omega)$ such that

$$\langle \mathcal{A}u, v\rangle = \langle b, v\rangle \quad \text{for all } v \in H_0^1(\Omega), \quad \text{i.e.,} \quad \mathcal{A}u = b \quad \text{in } (H_0^1(\Omega))^{\#}; \tag{1.5}$$

cf., [26, Section 5].

Further we consider the space $H_0^1(\Omega)$ with the norm $|\cdot|_{H^1}$ and discuss the properties of the operator $\mathcal{A}$ and the functional b. The definition of $\mathcal{A}$ together with the assumptions **PDE 1-2** implies that $\mathcal{A}$ is *self-adjoint (w.r.t. the duality pairing)*, *bounded* and *coercive*, i.e.,

$$\langle \mathcal{A}u, v\rangle = \langle \mathcal{A}v, u\rangle \qquad \text{for all } u, v \in H_0^1(\Omega), \tag{1.6}$$

$$\text{there is a } C_{\mathcal{A}} > 0 : \quad \sup_{v\in H_0^1(\Omega);|v|_{H^1}=1} \|\mathcal{A}v\|_{(H_0^1)^\#} \le C_{\mathcal{A}}, \tag{1.7}$$

$$\text{there is a } c_{\mathcal{A}} > 0 : \quad \inf_{v\in H_0^1(\Omega);|v|_{H^1}=1} \langle \mathcal{A}v, v\rangle \ge c_{\mathcal{A}}. \tag{1.8}$$

The boundedness and coercivity constants can be taken as

$$C_{\mathcal{A}} = \max_{x\in\overline{\Omega}} \tilde{a}(x) \quad \text{and} \quad c_{\mathcal{A}} = c_{\tilde{a}}. \tag{1.9}$$

The assumption **PDE 3** yields that b is *bounded*, i.e.,

$$\text{there is a } C_b > 0 : \quad \|b\|_{(H_0^1)^\#} = \sup_{v\in H_0^1(\Omega);|v|_{H^1}=1} |\langle b, v\rangle| \le C_b. \tag{1.10}$$

For the proof of the properties (1.7) - (1.10) see, e.g., the discussions in [18, Chapter 2] or the author's bachelor thesis [23, Section 1.2], where it is proven in more general setting.

The properties (1.6)-(1.8) yield that $\mathcal{A}$ defines an $\mathcal{A}$-inner product on $H_0^1(\Omega)$

$$(\cdot,\cdot)_{\mathcal{A}} : H_0^1(\Omega) \times H_0^1(\Omega) \to \mathbb{R}, \quad (\cdot,\cdot)_{\mathcal{A}} := \langle \mathcal{A}\cdot, \cdot\rangle$$

and the associated $\mathcal{A}$-norm $\|\cdot\|_{\mathcal{A}} := \sqrt{(\cdot,\cdot)_{\mathcal{A}}}$ is topologically equivalent to the norm $|\cdot|_{H^1}$; see, e.g., [18, Chapters 2-3].

The Lax–Milgram theorem (see, e.g., [18, Section 3.3]) says that there is a unique weak solution $u \in H_0^1(\Omega)$ of (1.5) such that

$$|u|_{H^1} \le \frac{1}{c_{\mathcal{A}}} \|b\|_{(H_0^1)^\#}.$$

The *regularity* of the weak solution u of (1.5) depends on the assumptions **PDE 1-3**; see, e.g., [13, 5, 19]. Assuming, for example, in addition to **PDE 1-3** that Ω is convex, the weak solution u belongs also to $H^2(\Omega)$; see, e.g., the result in [13, Theorem 3.2.1.2].

1.3 Finite element method

In this section we briefly, describe the Galerkin discretization using the piecewise linear finite element method (FEM). For a general introduction on FEM see, e.g., [7, 4, 11].

Triangulation and the finite element spaces

Consider the domain Ω defined in the assumption **PDE 1**. Let T be a *triangulation* of Ω, i.e., a finite partition of Ω which satisfies the following assumptions (see, e.g., [7, 4]).

FEM 1: Every element in T is a triangle.

FEM 2: The closure of Ω is the union of all elements in T, i.e., $\overline{\Omega} = \cup_{K \in T} K$.

FEM 3: Any two elements in T are either disjoint or share a common edge or vertex.

For any element $K \in T$, h_K denotes its diameter and ρ_K the diameter of the largest disc inscribed into K. The *mesh size parameter* h_T and the *shape parameter* C_T are defined as (see, e.g., [25, 7, 4])

$$h_T := \max_{K \in T} h_K \quad \text{and} \quad C_T := \max_{K \in T} \frac{h_K}{\rho_K}. \tag{1.11}$$

Let $\mathcal{N}$ denote the set of all nodes (i.e. the vertices of the elements of T) and $\mathcal{N}_{\text{int}} := \mathcal{N} \setminus \partial\Omega$ the set of free nodes. A continuous function on $\overline{\Omega}$ is said to be *piecewise linear* on T, if its restriction to any element in T is a linear polynomial; see, e.g., [7, 4]. For every node $z \in \mathcal{N}$, let ϕ_z be the continuous piecewise linear function on T that has a value one at node z and vanishes at all other nodes.

Let S denote the *finite element space* of continuous piecewise linear functions on T, i.e., the space spanned by the functions $\{\phi_z,\, z \in \mathcal{N}\}$ and let S_0 denote the subspace of S containing all functions from S that vanish on the boundary $\partial\Omega$, i.e., the space spanned by $\{\phi_z,\, z \in \mathcal{N}_{\text{int}}\}$; see, e.g., [7, 4]. The basis of S_0 can be written as $\Phi := (\phi_1, \ldots, \phi_N)$, where N is the number of free nodes. The spaces S and S_0 are finite dimensional subspaces of $H^1(\Omega)$ and $H_0^1(\Omega)$, respectively; see, e.g., [7, Theorem 2.1.1.].

Galerkin discretization

The *Galerkin discretization* of (1.5) reads as

$$\text{find } u_T \in S_0 : \quad \langle \mathcal{A} u_T, v \rangle = \langle b, v \rangle \quad \text{for all } v \in S_0; \tag{1.12}$$

see, e.g., [7, 4]. The solution u_T is in the literature called the *Galerkin solution.* The formulation of (1.12) yields that the residual

$$r(u_T) := b - \mathcal{A} u_T \in (H_0^1(\Omega))^{\#}$$

is orthogonal (w.r.t. the duality pairing) to the subspace S_0, i.e.,

$$\langle r(v_h), v \rangle = \langle b - \mathcal{A} u_T, v \rangle = 0 \quad \text{for all } v \in S_0.$$

This property is in the literature known as the *Galerkin orthogonality*; see, e.g., [4, Proposition 2.5.9].

Exploiting the linearity of $\mathcal{A}$ and b, the problem (1.12) can be formulated as a system of linear algebraic equations for the coordinates of u_T in the basis Φ

$$\text{find } \mathbf{u} \in \mathbb{R}^N : \quad \mathbf{A}\mathbf{u} = \mathbf{b}, \tag{1.13}$$

where

$$\begin{aligned} &\mathbf{A} \in \mathbb{R}^{N \times N}, && (\mathbf{A})_{ij} = \langle \mathcal{A}\phi_j, \phi_i \rangle, \quad i, j = 1, \ldots, N, \\ &\mathbf{b} \in \mathbb{R}^N, && (\mathbf{b})_i = \langle b, \phi_i \rangle, \qquad i = 1, \ldots, N. \end{aligned}$$

The self-adjointness and coercivity of $\mathcal{A}$ yield the symmetry and positive definiteness of the matrix $\mathbf{A}$; see Section 2.4 below, where it is shown in more general setting. The local support of the basis functions $\{\phi_z, z \in \mathcal{N}_{\text{int}}\}$ and the definition of the matrix $\boldsymbol{A}$ imply that $\boldsymbol{A}$ is *sparse*, i.e., it has only few nonzero entries; see, e.g., the discussion in [11, Section 4.1.2].

2. Multilevel methods: Abstract description

The exposition starts with formulation of an abstract infinite-dimensional problem. After discussing its properties and the existence of its solution the problem is discretized. The multilevel methods are further described as methods for computing the approximate solution of the discretized problem using a finite sequence of finite-dimensional subspaces.

As mentioned in the Introduction, even though the multilevel methods are formulated in the finite-dimensional framework, it is important to have the underlying infinite-dimensional problem in mind. Especially, the properties of the infinite-dimensional problem will be used when studying the convergence of the multilevel methods in the next chapter.

This chapter contains the description of the two-level scheme and the V-cycle schemes in both the operator and matrix-vector formulations and it concludes with comments on connections with the subspace correction methods. The exposition of this chapter is motivated by and loosely follows the exposition in [26, 27].

2.1 Infinite-dimensional problem

In this section we follow the exposition in [18, Chapters 2-3] and [15, Section 2].

Let V be a real infinite-dimensional Hilbert space with the inner product and the associated norm

$$(\cdot,\cdot)_V : V \times V \to \mathbb{R} \quad \text{and} \quad \|\cdot\|_V := \sqrt{(\cdot,\cdot)_V}.$$

Let $V^{\#}$ denotes the dual space of V and let

$$\langle \cdot,\cdot \rangle : V^{\#} \times V \to \mathbb{R}, \tag{2.1}$$

stands for the duality pairing.

Let $\mathcal{A} : V \to V^{\#}$ be a linear operator that is self-adjoint (w.r.t. the duality pairing (2.1)), bounded and coercive, i.e.,

$$\langle \mathcal{A}u, v\rangle = \langle \mathcal{A}v, u\rangle \quad \text{for all } u, v \in V,$$

$$\text{there is a } C_{\mathcal{A}} > 0 : \qquad \sup_{v\in V;\|v\|_V=1} \|\mathcal{A}v\|_{V^{\#}} \leq C_{\mathcal{A}}, \tag{2.2}$$

$$\text{there is a } c_{\mathcal{A}} > 0 : \qquad \inf_{v\in V;\|v\|_V=1} \langle \mathcal{A}v, v\rangle \geq c_{\mathcal{A}}. \tag{2.3}$$

The inequalities (2.2) and (2.3) imply (see, e.g., [15, Section 2])

$$c_{\mathcal{A}}\|v\|_V^2 \leq \langle \mathcal{A}v, v\rangle \leq C_{\mathcal{A}}\|v\|_V^2 \quad \text{for all } v \in V. \tag{2.4}$$

The properties of $\mathcal{A}$ yield that $\mathcal{A}$ defines an inner product on V (see, e.g., [18, Section 3.2])

$$(\cdot,\cdot)_{\mathcal{A}} : V \times V \to \mathbb{R}, \quad (\cdot,\cdot)_{\mathcal{A}} := \langle \mathcal{A}\cdot, \cdot\rangle,$$

and the associated $\mathcal{A}$-norm $\|\cdot\|_{\mathcal{A}} := \sqrt{(\cdot,\cdot)_{\mathcal{A}}}$.

Finally, given $b \in V^{\#}$, consider the problem:

$$\text{find } u \in V: \quad \langle \mathcal{A}u, v\rangle = \langle b, v\rangle \quad \text{for all } v \in V, \quad \text{i.e.,} \quad \mathcal{A}u = b \quad \text{in } V^{\#}. \tag{2.5}$$

Using the Lax–Milgram theorem (see, e.g., [18, Section 3.3]), for each $b \in V^{\#}$ there is a unique solution $u \in V$ of (2.5), i.e., the inverse mapping $\mathcal{A}^{-1} : V^{\#} \to V$ exists.

2.2 Discretization and multilevel framework

In order to perform numerical computations, the infinite-dimensional problem (2.5) has to be discretized. Consider a finite sequence of finite-dimensional nested subspaces of V

$$V_0 \subset V_1 \subset \cdots \subset V_J. \tag{2.6}$$

Since V_j, $j = 0, 1, \ldots, J$, is a finite-dimensional space, all norms on V_j are topologically equivalent and all linear functionals are therefore bounded (w.r.t. any norm on V_j). The dual space $V_j^{\#}$ is thus a space consisting of all linear functionals on V_j and there holds

$$V_0^{\#} \supset V_1^{\#} \supset \cdots \supset V_J^{\#} \supset V^{\#}.$$

Discretizing the problem (2.5) on the finest[1] subspace V_J using the Galerkin method read as

$$\text{find } u_J \in V_J: \quad \langle \mathcal{A}u_J, v\rangle = \langle b, v\rangle \quad \text{for all } v \in V_J,$$

i.e.,

$$\text{find } u_J \in V_J: \quad \mathcal{A}u_J = b \quad \text{in } V_J^{\#}, \tag{2.7}$$

The concept of the multilevel methods is to compute an approximation to the solution of the discrete problem (2.7) using *smoothing* on levels $1, \ldots, J$, and a *solver* on the coarsest level. For the introduction to the multilevel methods with explanation of the concept of smoothing on concrete examples see, e.g., [6], [9, Section 2.5] or the author's bachelor thesis [23, Chapter 2].

The smoothing is on each level $j = 1, \ldots, J$, described using a linear self-adjoint (w.r.t. the duality pairing (2.1)), coercive operator $\mathcal{B}_j : V_j \to V_j^{\#}$, $j = 1, \ldots, J$, called smoother, see, e.g., [26, Section 3.4]. Note that the properties of $\mathcal{B}_j$ yield that it defines an inner product on V_j

$$(\cdot,\cdot)_{\mathcal{B}_j} : V_j \times V_j \to \mathbb{R}, \quad (\cdot,\cdot)_{\mathcal{B}_j} := \langle \mathcal{B}_j \cdot, \cdot\rangle, \tag{2.8}$$

and the existence of its inverse $\mathcal{B}_j^{-1} : V_j^{\#} \to V_j$ is guaranteed (by the Lax-Milgram theorem).

The solver on the coarsest level is in the literature typically assumed to be exact, see, e.g., [26, Algorithm 3.7], [27, p. 294]. However, as discussed in the

[1] In agreement with the literature (see, e.g., [26, 27, 21]), we say that V_{j-1}, is a coarser space than V_j, and V_j is a finer space than V_{j-1}. Viewing (2.6) as a hierarchy of subspaces, the level $j-1$ is said to be coarser than the level j, respectively level j is finer than level $j-1$.

Introduction, in practical computations the iterative solvers are often used see; e.g., [16, Section 2.5.1, Chapters 5–7]. In this text, we consider a solver on the coarsest level, whose action can be expressed by a linear, self-adjoint (w.r.t. the duality pairing (2.1)), coercive operator $\mathcal{B}_0 : V_0 \to V_0^{\#}$, respectively its inverse $\mathcal{B}_0^{-1} : V_0^{\#} \to V_0$.

2.3 Multilevel schemes

A multilevel method, for solving the discretized problem (2.7) can be considered as a sequence of iterations

$$u^{(n)} \in V_J, \quad n = 0, 1, 2, \dots \quad ,$$

that starts with a chosen initial approximation[2] $u^{(0)}$. Having $u^{(n)} \in V_J$ the new approximation $u^{(n+1)}$ is computed according to a multilevel scheme.

We first describe the idea on a two-level scheme, i.e., $J = 1$, and later introduce its generalization, the so-called V-cycle scheme.

2.3.1 Two-level scheme

The two level scheme is stated in Algorithm 1; cf. [26, Algorithm 3.7]. The idea behind it can be described as follows. The smoothing should eliminate the oscillatory components of the error. Its smoother part should be then reduced by being approximated on the coarse level. In particular, the defect is computed and taken as a right-hand side of the problem on the coarse level

$$\text{find } v \in V_0 : \quad \mathcal{A}v = d_1 \quad \text{in } V_0^{\#}.$$

After computing the (approximate) solution of the coarse level problem, it is used to correct the approximation on the fine level.

Algorithm 1 Two-level scheme, operator formulation.

$v_1^{[1]} := u^{(n)} + \mathcal{B}_1^{-1}(b - \mathcal{A}u^{(n)})$ ▷*smoothing*

$d_1 := b - \mathcal{A}v_1^{[1]}$ ▷*computation of the defect*

$v_0^{[2]} := \mathcal{B}_0^{-1} d_1$ ▷*solution on the coarse level*

$v_1^{[2]} := v_1^{[1]} + v_0^{[2]}$ ▷*correction*

$u^{(n+1)} := v_1^{[2]}$

2.3.2 V-cycle scheme

The V-cycle scheme can be seen as a generalization of the two-level scheme for more levels. We consider two versions stated as Algorithms 2 and 3; cf. [26, Algorithms 3.6 and 3.7]. In Algorithm 2 smoothing is done before the solution on the coarsest level, whereas in Algorithm 3 smoothing is preformed after the solution on the coarsest level. See the graphic in Figure 2.1.

[2]If there are no information leading to a proper choice, we typically set $u^{(0)} = 0$.

Algorithm 2 V-cycle scheme with pre-smoothing, operator formulation.

$v_J^{[1]} := u^{(n)} + \mathcal{B}_J^{-1}(b - \mathcal{A}u^{(n)})$ ▷*smoothing on the finest level*

$d_J := b - \mathcal{A}v_J^{[1]}$

for $j = J-1, \dots, 1$ **do**

  $v_j^{[1]} := \mathcal{B}_j^{-1} d_{j+1}$ ▷*smoothing*

  $d_j := d_{j+1} - \mathcal{A}v_j^{[1]}$

end for

$v_0^{[2]} := \mathcal{B}_0^{-1} d_1$ ▷*solution on the coarsest level*

for $j = 1, \dots, J$ **do**

  $v_j^{[2]} := v_j^{[1]} + v_{j-1}^{[2]}$ ▷*correction*

end for

$u^{(n+1)} := v_J^{[2]}$

Algorithm 3 V-cycle scheme with post-smoothing, operator formulation.

$d_J := b - \mathcal{A}u^{(n)}$

$v_0 := \mathcal{B}_0^{-1} d_J$ ▷*solution on the coarsest level*

for $j = 1, \dots, J-1$ **do**

  $v_j := v_{j-1} + \mathcal{B}_j^{-1}(d_J - \mathcal{A}v_{j-1})$ ▷*smoothing*

end for

$v_J := u^{(n)} + v_{J-1}$ ▷*correction on the finest level*

$u^{(n+1)} := v_J + \mathcal{B}_J^{-1}(b - \mathcal{A}v_J)$ ▷*smoothing on the finest level*

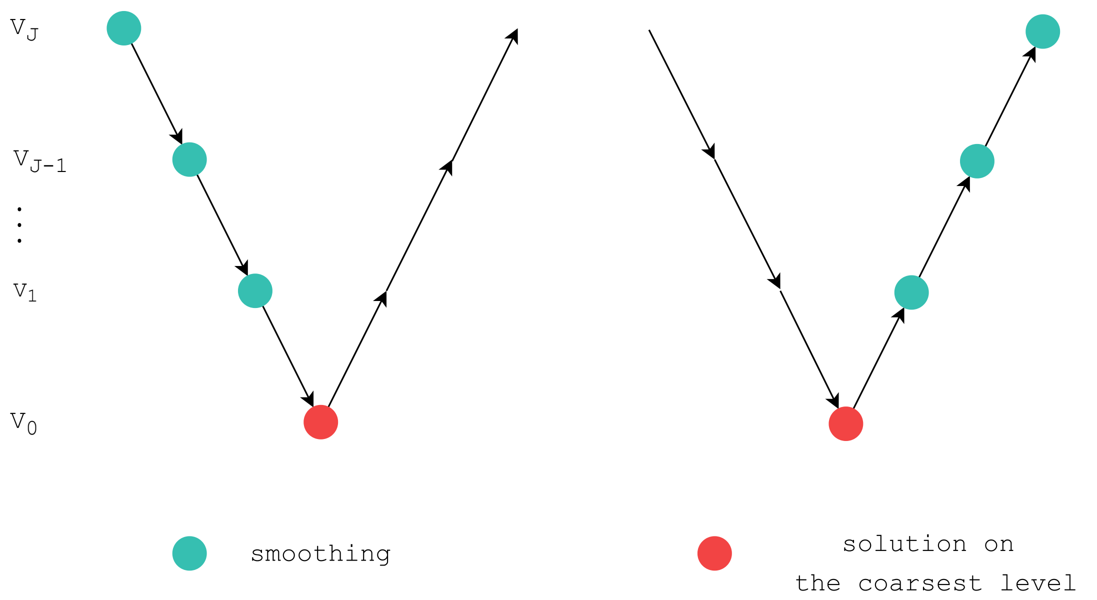

Figure 2.1: Ilustration of the V-cycle schemes - Algorithm 2 (left), and Algorithm 3 (right).

There exist also versions of the V-cycle scheme, where smoothing is performed before and after the solution on the coarsest level (see, e.g., [26, Algorithm 3.8]) and other multilevel schemes, e.g., the W-cycle scheme and the full multigrid algorithm (see, e.g., [14, Chapter 11]) which will be not considered in this text.

2.3.3 Error equations

Further we study the relation between the errors before and after one iteration of the V-cycle scheme.

Let $u^{(0)} \in V_J$ be an arbitrary initial approximation to the solution of $\mathcal{A}u_J = b$ in $V_J^{\#}$. Let $u^{(n)}$ and $u^{(n+1)}$ be the approximations computed after n and $n+1$ iterations of the V-cycle scheme with pre-smoothing (Algorithm 2). Writing the $(n+1)$st error $u_J - u^{(n+1)}$ using the individual steps of Algorithm 2 yields

$$\begin{aligned}
u_J - u^{(n+1)} &= u_J - (v_J^{[1]} + v_{J-1}^{[2]}) \\
&= u_J - (v_J^{[1]} + v_{J-1}^{[1]} + v_{J-2}^{[2]}) \\
&\dots \\
&= u_J - (v_J^{[1]} + v_{J-1}^{[1]} + \dots + v_1^{[1]} + v_0^{[2]}) \\
&= u_J - (v_J^{[1]} + v_{J-1}^{[1]} + \dots + v_1^{[1]} + \mathcal{B}_0^{-1} d_1) \\
&= u_J - (v_J^{[1]} + v_{J-1}^{[1]} + \dots + v_1^{[1]}) - \mathcal{B}_0^{-1} d_1.
\end{aligned}$$

Observing that $d_j = \mathcal{A}u_J - (\mathcal{A}v_J^{[1]} + \dots + \mathcal{A}v_j^{[1]})$ in $V_J^{\#}$, $j = 1, \dots, J$, and denoting I the identity operator on V_J leads to

$$\begin{aligned}
u_J - u^{(n+1)} &= (I - \mathcal{B}_0^{-1}\mathcal{A})(u_J - (v_J^{[1]} + v_{J-1}^{[1]} + \dots + v_1^{[1]})) \\
&= (I - \mathcal{B}_0^{-1}\mathcal{A})(u_J - (v_J^{[1]} + v_{J-1}^{[1]} + \dots + v_2^{[1]} + \mathcal{B}_1^{-1} d_2)) \\
&= (I - \mathcal{B}_0^{-1}\mathcal{A})(u_J - (v_J^{[1]} + v_{J-1}^{[1]} + \dots + v_2^{[1]}) - \mathcal{B}_1^{-1} d_2) \\
&= (I - \mathcal{B}_0^{-1}\mathcal{A})(I - \mathcal{B}_1^{-1}\mathcal{A})(u_J - (v_J^{[1]} + v_{J-1}^{[1]} + \dots + v_2^{[1]})) \\
&\dots \\
&= (I - \mathcal{B}_0^{-1}\mathcal{A})(I - \mathcal{B}_1^{-1}\mathcal{A}) \cdots (I - \mathcal{B}_J^{-1}\mathcal{A})(u_J - u^{(n)}).
\end{aligned}$$

Defining the operator E as

$$E := (I - \mathcal{B}_0^{-1}\mathcal{A})(I - \mathcal{B}_1^{-1}\mathcal{A}) \cdots (I - \mathcal{B}_J^{-1}\mathcal{A}) : V_J \to V_J$$

gives

$$u_J - u^{(n+1)} = E(u_J - u^{(n)}); \tag{2.9}$$

cf. [26, Section 3.4]. The operator E is in literature called the error propagation operator; see, e.g., [27, Section 5].

For the approximations $u^{(n)}$, $u^{(n+1)}$ computed after n and $n+1$ iterations of the V-cycle scheme with post-smoothing (Algorithm 3) holds

$$\begin{aligned}
u_J - u^{(n+1)} &= u_J - (v_J + \mathcal{B}_J^{-1}(b - \mathcal{A}v_J)) \\
&= u_J - (u^{(n)} + v_{J-1} + \mathcal{B}_J^{-1}(\mathcal{A}u_J - \mathcal{A}(u^{(n)} + v_{J-1}))) \\
&= (I - \mathcal{B}_J^{-1}\mathcal{A})(u_J - (u^{(n)} + v_{J-1})) \\
&= (I - \mathcal{B}_J^{-1}\mathcal{A})(u_J - (u^{(n)} + v_{J-2} + \mathcal{B}_{J-1}^{-1}(d_J - \mathcal{A}v_{J-2}))) \\
&= (I - \mathcal{B}_J^{-1}\mathcal{A})(u_J - (u^{(n)} + v_{J-2} + \mathcal{B}_{J-1}^{-1}(\mathcal{A}u_J - \mathcal{A}u^{(n)} - \mathcal{A}v_{J-2}))) \\
&= (I - \mathcal{B}_J^{-1}\mathcal{A})(I - \mathcal{B}_{J-1}^{-1}\mathcal{A})(u_J - (u^n + \ v_{J-2})) \\
&\dots \\
&= (I - \mathcal{B}_J^{-1}\mathcal{A}) \cdots (I - \mathcal{B}_1^{-1}\mathcal{A})(u_J - (u^{(n)} - v_0)) \\
&= (I - \mathcal{B}_J^{-1}\mathcal{A}) \cdots (I - \mathcal{B}_1^{-1}\mathcal{A})(u_J - (u^{(n)} - \mathcal{B}_0^{-1}(b - \mathcal{A}u^{(n)}))) \\
&= (I - \mathcal{B}_J^{-1}\mathcal{A}) \cdots (I - \mathcal{B}_1^{-1}\mathcal{A})(I - \mathcal{B}_0^{-1}\mathcal{A})(u_J - u^{(n)}).
\end{aligned}$$

Thus the errors satisfy the relation (cf. [26, Section 3.4])

$$u_J - u^{(n+1)} = E^*(u_J - u^{(n)}), \tag{2.10}$$

where E^* is the error propagation operator

$$E^* := (I - \mathcal{B}_J^{-1}\mathcal{A})(I - \mathcal{B}_{J-1}^{-1}\mathcal{A}) \cdots (I - \mathcal{B}_0^{-1}\mathcal{A}) : V_J \to V_J. \tag{2.11}$$

The self-adjointness of $\mathcal{A}$ and $\mathcal{B}_j^{-1}$ imply that the operators $I - \mathcal{B}_j^{-1}\mathcal{A}$ are adjoint w.r.t. the $\mathcal{A}$-inner product. Indeed,

$$\begin{aligned}
\langle \mathcal{A}(I - \mathcal{B}_j^{-1}\mathcal{A})u, v\rangle &= \langle \mathcal{A}u, v\rangle - \langle \mathcal{A}\mathcal{B}_j^{-1}\mathcal{A}u, v\rangle \\
&= \langle \mathcal{A}v, u\rangle - \langle \mathcal{A}v, \mathcal{B}_j^{-1}\mathcal{A}u\rangle \\
&= \langle \mathcal{A}v, u\rangle - \langle \mathcal{A}u, \mathcal{B}_j^{-1}\mathcal{A}v\rangle \\
&= \langle \mathcal{A}v, u\rangle - \langle \mathcal{A}\mathcal{B}_j^{-1}\mathcal{A}v, u\rangle, \\
&= \langle \mathcal{A}(I - \mathcal{B}_j^{-1}\mathcal{A})v, u\rangle.
\end{aligned}$$

Consequently, E^* is the adjoint of E w.r.t. the $\mathcal{A}$-inner product; cf. [26, Remark 3.1].

2.4 Matrix-vector representation

This section gives a matrix-vector representation of the discrete problem (2.7) and the multilevel schemes presented above. We use analogous approach as in [18, Chapter 6] and adopt it for the multilevel methods.

2.4.1 Discrete problem

Consider a basis of the N_J-dimensional space V_J

$$\Phi_J = (\phi_1^{(J)}, \dots, \phi_{N_J}^{(J)})$$

and the associated canonical dual basis of $V_J^{\#}$

$$\Phi_J^{\#} = (\phi_1^{(J)\#}, \dots, \phi_{N_J}^{(J)\#}),$$

i.e.,

$$\langle \phi_i^{(J)\#}, \phi_j^{(J)} \rangle = \delta_{ij}, \quad i,j = 1,\dots,N_J. \tag{2.12}$$

Then each $u \in V_J$ can be represented in $\mathbb{R}^{N_J}$ using the vector $\mathbf{u} \in \mathbb{R}^{N_J}$, whose components are the coordinates of u in the basis Φ_J, i.e.,

$$\mathbf{u} = (\langle \phi_1^{(J)\#}, u \rangle, \dots, \langle \phi_{N_J}^{(J)\#}, u \rangle)^T,$$

symbolically written as

$$u = \Phi_J \mathbf{u}.$$

Analogously each $f \in V_J^\#$ can be represented in $\mathbb{R}^{N_J}$ by its coordinates $\mathbf{f} \in \mathbb{R}^{N_J}$ in the basis $\Phi_J^\#$

$$f = \Phi_J^\# \mathbf{f}, \quad \mathbf{f} = (\langle f, \phi_1^{(J)} \rangle, \dots, \langle f, \phi_{N_J}^{(J)} \rangle)^T.$$

The restriction of the operator $\mathcal{A}$ to the finite-dimensional space V_J can be represented by a matrix $\mathbf{A_J} \in \mathbb{R}^{N_J \times N_J}$ as follows. Since $\mathcal{A}\phi_j^{(J)}$ belongs to $V^\# \subset V_J^\#$ it can be expressed in the basis $\Phi_J^\#$ as

$$\mathcal{A}\phi_j^{(J)} = \sum_{i=1}^{N_J} \nu_i \phi_i^{(J)\#}, \quad \nu_i = \langle \mathcal{A}\phi_j^{(J)}, \phi_i^{(J)} \rangle.$$

Denoting

$$\mathbf{A_J} \in \mathbb{R}^{N_J \times N_J}, \quad (\mathbf{A_J})_{ij} := \langle \mathcal{A}\phi_j^{(J)}, \phi_i^{(J)} \rangle \tag{2.13}$$

and defining

$$\mathcal{A}\Phi_J : \mathbb{R}^{N_J} \to V_J^\#, \quad \mathcal{A}\Phi_J := (\mathcal{A}\phi_1^{(J)}, \dots, \mathcal{A}\phi_{N_J}^{(J)})$$

leads to

$$\mathcal{A}\Phi_J = \Phi_J^\# \mathbf{A_J}, \tag{2.14}$$

i.e., the jth column of the matrix $\mathbf{A_J}$ contains the coordinates of the image $\mathcal{A}\phi_j^{(J)}$ in the basis $\Phi_J^\#$.

The self-adjointness and coercivity of $\mathcal{A}$ yields the symmetry and positive definiteness of the matrix $\mathbf{A_J}$; let $\mathbf{v} \in \mathbb{R}^{N_J}$, $\mathbf{v} \neq 0$ and consider the function $v = \Phi_J \mathbf{v}$, then

$$\mathbf{v}^* \mathbf{A_J} \mathbf{v} = \langle \Phi_J^\# \mathbf{A_J} \mathbf{v}, \Phi_J \mathbf{v} \rangle = \langle \mathcal{A}v, v \rangle \geq c_{\mathcal{A}} \|v\|_V > 0.$$

Rewriting the formulas in (2.7) using these matrix and vector representations

$$\mathcal{A}u_J = \mathcal{A}\Phi_J \mathbf{u_J} = \Phi_J^\# \mathbf{A_J} \mathbf{u_J} = b = \Phi_J^\# \mathbf{b},$$

yields the reformulation of (2.7) as a system of linear algebraic equations for the coordinates of u_J in the basis Φ_J

$$\text{find } \mathbf{u_J} \in \mathbb{R}^{N_J} : \quad \mathbf{A_J} \mathbf{u_J} = \mathbf{b}. \tag{2.15}$$

2.4.2 Prolongation and restriction matrices

For each $j = 0, \dots, J$, consider a basis of the N_j-dimensional space V_j

$$\Phi_j = (\phi_1^{(j)}, \dots, \phi_{N_j}^{(j)})$$

and the associated canonical dual basis of $V_j^{\#}$

$$\Phi_j^{\#} = (\phi_1^{(j)\#}, \dots, \phi_{N_j}^{(j)\#}).$$

Since V_{j-1} and V_j are nested subspaces, a function from V_{j-1} can be expressed in the basis Φ_{j-1} as well as in the basis Φ_j. Specially a basis function $\phi_\ell^{(j-1)}$ can be written as

$$\phi_\ell^{(j-1)} = \sum_{k=1}^{N_j} \langle \phi_k^{(j)\#}, \phi_\ell^{(j-1)} \rangle \phi_k^{(j)}. \tag{2.16}$$

Introducing the *prolongation matrix*

$$\mathbf{I_{j-1}^{j}} \in \mathbb{R}^{N_j \times N_{j-1}}, \quad (\mathbf{I_{j-1}^{j}})_{k\ell} := \langle \phi_k^{(j)\#}, \phi_\ell^{(j-1)} \rangle \quad k = 1, \dots, N_j; \quad \ell = 1, \dots, N_{j-1},$$

gives the relation between the basis functions of V_{j-1} and V_j

$$\Phi_{j-1} = \Phi_j \mathbf{I_{j-1}^{j}}. \tag{2.17}$$

For each function $u \in V_{j-1}$ then holds

$$u = \Phi_{j-1}\mathbf{u} = \Phi_j \mathbf{I_{j-1}^{j}} \mathbf{u},$$

i.e., the prolongation matrix $\mathbf{I_{j-1}^{j}}$ maps (*prolongates*) the coefficients of u in the basis Φ_{j-1} to the coefficients of u in the basis Φ_j.

Analogously, since $V_j^{\#} \subset V_{j-1}^{\#}$, a functional from $V_j^{\#}$ can be expressed in both dual bases $\Phi_j^{\#}$ and $\Phi_{j-1}^{\#}$. Specially, a basis functional $\phi_m^{(j)\#}$ can be written as

$$\phi_m^{(j)\#} = \sum_{\ell=1}^{N_{j-1}} \langle \phi_m^{(j)\#}, \phi_\ell^{(j-1)} \rangle \phi_\ell^{(j-1)\#}. \tag{2.18}$$

Using (2.16) in (2.18) yields

$$\begin{aligned} \phi_m^{(j)\#} &= \sum_{\ell=1}^{N_{j-1}} \langle \phi_m^{(j)\#}, \sum_{k=1}^{N_j} \langle \phi_k^{(j)\#}, \phi_\ell^{(j-1)} \rangle \phi_k^{(j)} \rangle \phi_\ell^{(j-1)\#} \\ &= \sum_{\ell=1}^{N_{j-1}} \sum_{k=1}^{N_j} \langle \phi_k^{(j)\#}, \phi_\ell^{(j-1)} \rangle \langle \phi_m^{(j)\#}, \phi_k^{(j)} \rangle \phi_\ell^{(j-1)\#} \\ &= \sum_{\ell=1}^{N_{j-1}} \langle \phi_m^{(j)\#}, \phi_\ell^{(j-1)} \rangle \phi_\ell^{(j-1)\#}, \end{aligned}$$

where the property of the canonical dual basis $\Phi_j^{\#}$ was used. Summarizing, there holds

$$\Phi_j^{\#} = \Phi_{j-1}^{\#} \left(\mathbf{I_{j-1}^{j}}\right)^*, \tag{2.19}$$

where $\left(\mathbf{I^j_{j\text{-}1}}\right)^*$ denotes the transpose of the matrix $\mathbf{I^j_{j\text{-}1}}$ and each functional $f \in V_j^\#$ can be written as

$$f = \Phi_j^\# \mathbf{f} = \Phi_{j-1}^\# \left(\mathbf{I^j_{j\text{-}1}}\right)^* \mathbf{f},$$

i.e., the coefficients of f in the basis $\Phi_j^\#$ are mapped (*restricted*) using the *restriction matrix* $\left(\mathbf{I^j_{j\text{-}1}}\right)^*$ to the coefficients of f in the basis $\Phi_{j-1}^\#$.

Moreover, denoting $\mathbf{I^J_j} \in \mathbb{R}^{N_J \times N_j}$ the matrix

$$\mathbf{I^J_j} := \mathbf{I^J_{J\text{-}1}} \mathbf{I^{J\text{-}1}_{J\text{-}2}} \cdots \mathbf{I^{j+1}_j}, \quad j = 0, \ldots, J-1,$$

and using (2.17) and (2.19) recursively yields

$$\Phi_j = \Phi_J \mathbf{I^J_j} \quad \text{and} \quad \Phi_J^\# = \Phi_j^\# \left(\mathbf{I^J_j}\right)^*, \quad j = 0, \ldots, J-1. \tag{2.20}$$

2.4.3 Representation of operators

For each $j = 0, 1, \ldots, J$, the restrictions of the operator $\mathcal{A}$ to V_j can be represented using a symmetric positive definite matrices

$$\mathbf{A_j} \in \mathbb{R}^{N_j \times N_j}, \quad \left(\mathbf{A_j}\right)_{k\ell} := \langle \mathcal{A}\phi_\ell^{(j)}, \phi_k^{(j)} \rangle, \quad \mathcal{A}\Phi_j = \Phi_j^\# \mathbf{A_j}; \tag{2.21}$$

see the derivation of (2.14).

Having the matrix representation of the restriction of the operator $\mathcal{A}$ to V_J, the matrix representation of the restriction of $\mathcal{A}$ to V_j, $j = 0, 1, \ldots, J-1$, can be also obtained using the prolongation and restriction relations (2.20)

$$\mathcal{A}\Phi_j = \mathcal{A}\Phi_J \mathbf{I}_j^J = \Phi_J^\# \mathbf{A_J} \mathbf{I^J_j} = \Phi_j^\# \left(\mathbf{I^J_j}\right)^* \mathbf{A_J} \mathbf{I^J_j}. \tag{2.22}$$

Comparing (2.21) and (2.22) yields

$$\mathbf{A_j} = \left(\mathbf{I^J_j}\right)^* \mathbf{A_J} \mathbf{I^J_j}. \tag{2.23}$$

Analogously, for each $j = 0, 1, \ldots, J$, the operator $\mathcal{B}_j$ can be represented by a symmetric positive definite matrix

$$\mathbf{B_j} \in \mathbb{R}^{N_j \times N_j}, \quad \left(\mathbf{B_j}\right)_{k\ell} := \langle \mathcal{B}_j\phi_\ell^{(j)}, \phi_k^{(j)} \rangle, \quad \mathcal{B}_j\Phi_j = \Phi_j^\# \mathbf{B_j}. \tag{2.24}$$

The matrix representation of the inverse operator $\mathcal{B}_j^{-1}$ is then given by the inverse of the matrix $\mathbf{B_j}$, i.e., there holds

$$\mathcal{B}_j^{-1}\Phi_j^\# = \Phi_j \mathbf{B_j}^{-1}. \tag{2.25}$$

To show this, let $\mathbf{v}, \mathbf{f} \in \mathbb{R}^{N_j}$ and consider the function $v = \Phi_j \mathbf{v}$ and the functional $f = \Phi_j^\# \mathbf{f}$. Assuming that $\mathcal{B}_j^{-1}\Phi_j^\# = \Phi_j \mathbf{X}$ for some matrix $\mathbf{X} \in \mathbb{R}^{N_j \times N_j}$

$$\mathbf{v}^*\mathbf{f} = \langle f, v \rangle = \langle \mathcal{B}_j\mathcal{B}_j^{-1} f, v \rangle = \langle \mathcal{B}_j\mathcal{B}_j^{-1}\Phi_j^\# \mathbf{f}, v \rangle = \langle \mathcal{B}_j \Phi_j \mathbf{X}\mathbf{f}, \Phi_j \mathbf{v} \rangle = \mathbf{v}^* \mathbf{B_j} \mathbf{X} \mathbf{f},$$

where the matrix representation (2.24) of the operator $\mathcal{B}_j$ was used. Since $\mathbf{v}$ and $\mathbf{f}$ can be chosen arbitrary, comparison of the left and the right term yields $\mathbf{X} = \mathbf{B_j}^{-1}$.

2.4.4 Two-level scheme

Using the above matrix and vector representations the formulas in the two-level scheme (Algorithm 1) can be rewritten as follows: the smoothing process

$$\begin{aligned} v_1^{[1]} = \Phi_1 \mathbf{v_1^{[1]}} &= u^{(n)} + \mathcal{B}_1^{-1}(b - \mathcal{A}u^{(n)}) \\ &= \Phi_1 \mathbf{u^{(n)}} + \mathcal{B}_1^{-1}(\Phi_1^{\#}\mathbf{b} - \mathcal{A}\Phi_1 \mathbf{u^{(n)}}) \\ &= \Phi_1 \mathbf{u^{(n)}} + \mathcal{B}_1^{-1}\Phi_1^{\#}(\mathbf{b} - \mathbf{A_1 u^{(n)}}) \\ &= \Phi_1(\mathbf{u^{(n)}} + \mathbf{B_1^{-1}}(\mathbf{b} - \mathbf{A_1 u^{(n)}})), \end{aligned}$$

the computation of the defect

$$d_1 = b - \mathcal{A}v_1^{[1]} = \Phi_1^{\#}(\mathbf{b} - \mathbf{A_1 v_1^{[1]}}) = \Phi_0^{\#}\left(\mathbf{I_0^1}\right)^*(\mathbf{b} - \mathbf{A_1 v_1^{[1]}}),$$

the solution on the coarse level

$$\begin{aligned} v_0^{[2]} = \Phi_0 \mathbf{v_0^{[2]}} &= \mathcal{B}_0^{-1} d_1 \\ &= \mathcal{B}_0^{-1}\Phi_0^{\#}\left(\mathbf{I_0^1}\right)^*(\mathbf{b} - \mathbf{A_1 v_1^{[1]}}) \\ &= \Phi_0 \mathbf{B_0^{-1}}\left(\mathbf{I_0^1}\right)^*(\mathbf{b} - \mathbf{A_1 v_1^{[1]}}), \end{aligned}$$

and finally the correction

$$\begin{aligned} v_1^{[2]} = \Phi_1 \mathbf{v_1^{[2]}} &= v_1^{[1]} + v_0^{[2]} \\ &= \Phi_1 \mathbf{v_1^{[1]}} + \Phi_0 \mathbf{v_0^{[2]}} \\ &= \Phi_1(\mathbf{v_1^{[1]}} + \left(\mathbf{I_0^1}\right)\mathbf{v_0^{[2]}}). \end{aligned}$$

The two-level method defined by Algorithm 1 can be hence reformulated purely in terms of vectors and matrices as an iterative process

$$\mathbf{u^{(n)}} \in \mathbb{R}^{N_1}, \quad n = 0, 1, 2, \ldots \quad ,$$

that computes an approximation of the solution $\mathbf{u_1} \in \mathbb{R}^{N_1}$ of (2.15). Given $\mathbf{u^{(n)}}$, the new approximation $\mathbf{u^{(n+1)}}$ is obtained by Algorithm 4.

Algorithm 4 Two-level scheme with pre-smoothing, matrix-vector formulation.

$\mathbf{v_1^{[1]}} := \mathbf{u^{(n)}} + \mathbf{B_1^{-1}}(\mathbf{b} - \mathbf{A_J u^{(n)}})$ ▷*smoothing*

$\mathbf{d_1} := \mathbf{b} - \mathbf{A_1 v^{[1]}}$ ▷*computation of the defect*

$\mathbf{f_0} := \left(\mathbf{I_0^1}\right)^* \mathbf{d_1}$ ▷*restriction*

$\mathbf{v_0^{[2]}} := \mathbf{B_0^{-1} f_0}$ ▷*solution on the coarse level*

$\mathbf{v_1^{[2]}} := \mathbf{v_1^{[1]}} + \left(\mathbf{I_0^1}\right)\mathbf{v_0^{[2]}}$ ▷*prolongation and correction*

$\mathbf{u^{(n+1)}} = \mathbf{v_1^{[2]}}$

Algorithm 5 V-cycle scheme with pre-smoothing, matrix-vector formulation.

$\mathbf{v_J^{[1]} := u^{(n)} + B_j^{-1}(b - A_J u^{(n)})}$ ▷*smoothing on the finest level*

$\mathbf{d_J := b - A_J v^{[1]}}$ ▷*computation of the defect*

$\mathbf{f_{J\text{-}1} := \left(I_{J\text{-}1}^{J}\right)^* d_J}$ ▷*restriction*

for j=J-1,...,1 do

  $\mathbf{v_j^{[1]} := B_j^{-1} f_j}$ ▷*smoothing*

  $\mathbf{d_j := \left(I_{j\text{-}1}^{j}\right)^* (f_j - A_j v_j^{[1]})}$ ▷*computation of the defect*

  $\mathbf{f_{j\text{-}1} := \left(I_{j\text{-}1}^{j}\right)^* d_j}$ ▷*restriction*

end for

$\mathbf{v_0^{[2]} := B_0^{-1} f_0}$ ▷*solution on the coarsest level*

for j=1,...,J do

  $\mathbf{v_j^{[2]} := v_j^{[1]} + \left(I_{j-1}^{j}\right) v_{j\text{-}1}^{[2]}}$ ▷*prolongation and correction*

end for

$\mathbf{u^{(n+1)} = v_J^{[2]}}$

Algorithm 6 V-cycle scheme with post-smoothing, operator formulation.

$\mathbf{d_J := b - A_J u^{(n)}}$ ▷*computation of the defect*

$\mathbf{f_{J\text{-}1} := \left(I_{J\text{-}1}^{J}\right)^* d_J}$ ▷*restriction*

for j = J-1,...,1 do

  $\mathbf{f_{j\text{-}1} := \left(I_{j\text{-}1}^{j}\right)^* f_j}$ ▷*restriction*

end for

$\mathbf{v_0 := B_0^{-1} f_0}$ ▷*solution on the coarsest level*

for j=1,...,J-1 do

  $\mathbf{v_j := \left(I_{j-1}^{j}\right) v_{j\text{-}1}}$ ▷*prolongation*

  $\mathbf{v_j := v_j + B_j^{-1}(f_j - A_J v_j)}$ ▷*smoothing*

end for

$\mathbf{v_J := u^{(n+1)} + \left(I_{J-1}^{J}\right) v_{J\text{-}1}}$ ▷*prolongation and correction*

$\mathbf{u^{(n+1)} := v_J + B_j^{\text{-}1}(b - A_J v_J)}$ ▷*smoothing on the finest level*

2.4.5 V-cycle scheme

Analogously to the two-level scheme, the V-cycle schemes (Algorithms 2 and 3) can be reformulated using the above matrix and vector representations in purely algebraic terms, see Algorithms 5 and 6.

We note that Algorithms 5 and 6 falls into the more general class of classical multigrid algorithms presented in [14, Section 11.4], where the considered restriction and prolongation matrices does not need to be transpose of each other.

2.5 Link to multiplicative subspace correction methods

The subspace correction methods, also known as Schwarz methods, see, e.g., [12], [26, Section 3], [21, Chapters 1-2], are based on a decomposition of the search space into subspaces. The approximation is then computed by performing corrections on the individual subspaces.

In this section we describe a multiplicative subspace correction method for finding an approximation to the solution u_J of the discretized problem (2.7) that uses the sequence of subspaces $V_0, \ldots, V_J$, the smoothers $\mathcal{B}_j$, $j = 1, \ldots, J$, and the operator $\mathcal{B}_0$ describing the solver on the coarsest level.

Let $v \in V_J$ be an approximation to u_J, the subspace correction w.r.t. the subspace V_j reads as

$$v^{\{j\}} := v + \mathcal{B}_j^{-1}(b - \mathcal{A}v).$$

Combining the subspace corrections successively w.r.t. all subspaces V_j yields one iteration of a multiplicative subspace correction method. We consider two versions depending on the ordering of the subspace corrections. In Algorithm 7 the corrections are successively carried out on the subspaces $V_J, V_{J-1}, \ldots, V_0$, whereas the reverse ordering is used in Algorithm 8; cf. [26, Algorithm 3.3]. See the graphic in Figure 2.2.

Algorithm 7 Multiplicative subspace correction method - version 1, operator formulation.

$v^{\{J\}} := u^{(n)} + \mathcal{B}_J^{-1}(b - \mathcal{A}u^{(n)})$
for $j = J, \ldots, 1$ **do**
  $v^{\{j-1\}} := v^{\{j\}} + \mathcal{B}_{j-1}^{-1}(b - \mathcal{A}v^{\{j\}})$
end for
$u^{(n+1)} := v^{\{0\}}$

Algorithm 8 Multiplicative subspace correction method - version 2, operator formulation.

$v^{\{0\}} := u^{(n)} + \mathcal{B}_0^{-1}(b - \mathcal{A}u^{(n)})$
for $j = 1, \ldots, J$ **do**
  $v^{\{j\}} := v^{\{j-1\}} + \mathcal{B}_j^{-1}(b - \mathcal{A}v^{\{j-1\}})$
end for
$u^{(n+1)} := v^{\{J\}}$

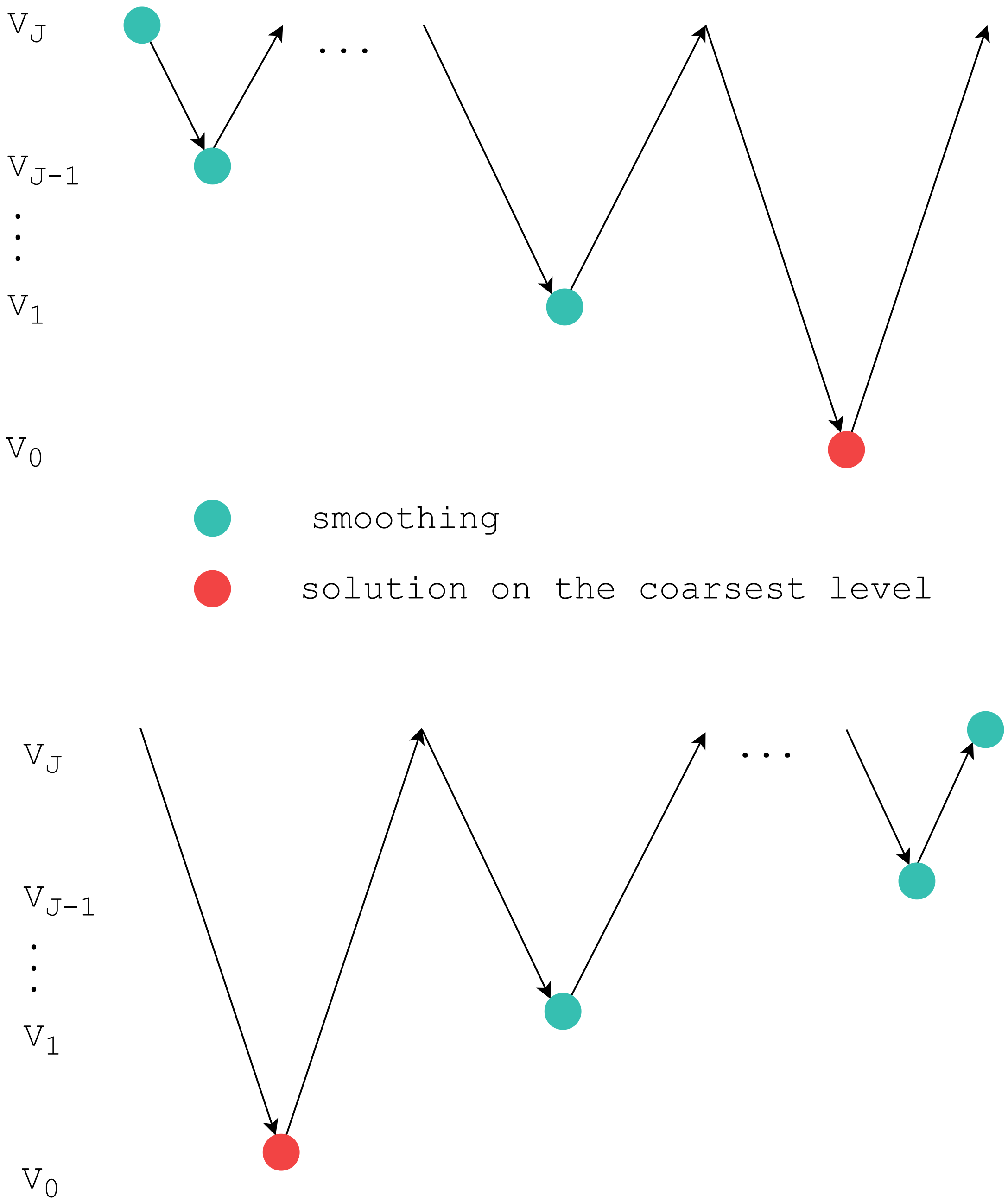

Figure 2.2: Illustration of the multiplicative subspace correction methods - Algorithm 7 (top) and Algorithm 8 (bottom).

Error equations

Let $u^{(0)} \in V_J$ be an arbitrary initial approximation to the solution of $\mathcal{A}u_J = b$ in $V_J^{\#}$. Let $u^{(n)}$ and $u^{(n+1)}$ be the approximations given utilizing Algorithm 7 after n and $n+1$ iterations. Expanding the $(n+1)$st error using the individual steps of Algorithm 7 yields

$$\begin{aligned}
u_J - u^{(n+1)} &= u_J - v^{\{0\}} \\
&= u_J - v^{\{1\}} - \mathcal{B}_J^{-1}(b - \mathcal{A}v^{\{1\}}) \\
&= u_J - v^{\{1\}} - \mathcal{B}_J^{-1}(\mathcal{A}u_J - \mathcal{A}v^{\{1\}}) \\
&= (I - \mathcal{B}_J^{-1}\mathcal{A})(u_J - v^{\{1\}}) \\
&\dots \\
&= (I - \mathcal{B}_J^{-1}\mathcal{A})(I - \mathcal{B}_{J-1}^{-1}\mathcal{A}) \cdots (I - \mathcal{B}_0^{-1}\mathcal{A})(u_J - u^{(n)}),
\end{aligned}$$

i.e.,

$$u_J - u^{(n+1)} = E(u_J - u^{(n)}).$$

Analogously for the approximations $u^{(n)}$, $u^{(n+1)}$ computed by Algorithm 8

$$u_J - u^{(n+1)} = E^*(u_J - u^{(n)}).$$

We see that the multilevel V-cycle scheme Algorithm 2 (respectively Algorithm 3) and the multiplicative subspace correction method Algorithm 7 (respectively Algorithm 8) give the same approximations $u^{(n)}$; when started with the same initial approximation $u^{(0)}$.

This connection is the reason why the multilevel methods are in literature often analysed from a viewpoint of the subspace correction methods; see, e.g., the approaches in [26, 27].

3. Convergence of multilevel methods: Abstract description

In this chapter we first introduce the concept of *uniform convergence* of the multilevel methods. After specifying and commenting on the assumptions, the main result on the uniform convergence is formulated. Even though the multilevel methods are described in the finite-dimensional framework, it is important to note that in order to prove the main result and also to verify its assumptions, the properties of the infinite-dimensional problem will be used.

In contrast with the exposition in the previous chapter we first discuss the properties of the V-cycle scheme with post-smoothing and later comment on the variant with pre-smoothing.

Let $u^{(0)} \in V_J$ be an arbitrary initial approximation of the solution of $\mathcal{A}u_J = b$ in $V_J^{\#}$. Let $u^{(n)}$ and $u^{(n+1)}$ be the approximations computed after n and $n+1$ iterations of the V-cycle scheme with post-smoothing (Algorithm 3). As we have shown in the previous chapter (see equation (2.10)) the errors satisfy

$$u_J - u^{(n+1)} = E^*(u_J - u^{(n)}), \quad n = 0, 1, \dots \quad , \tag{3.1}$$

where E^* is the operator

$$E^* = (I - \mathcal{B}_J^{-1}\mathcal{A})(I - \mathcal{B}_{J-1}^{-1}\mathcal{A}) \cdots (I - \mathcal{B}_0^{-1}\mathcal{A}) : V_J \to V_J.$$

Motivated by [27, 26] we will show that the operator E^* is a contraction with respect to the $\mathcal{A}$-norm with a contraction factor *independent of the number of levels in the V-cycle scheme*, i.e., show that there exists a contraction factor $\rho \in (0, 1)$ *independent of* J such that

$$\|E^*\|_{\mathcal{A}} := \sup_{v \in V_J; \|v\|_{\mathcal{A}}=1} \|E^* v\|_{\mathcal{A}} \le \rho. \tag{3.2}$$

Provided that (3.2) holds,

$$\begin{aligned} \|u_J - u^{(n+1)}\|_{\mathcal{A}} &= \|E^*(u_J - u^{(n)})\|_{\mathcal{A}} \\ &\le \|E^*\|_{\mathcal{A}} \|u_J - u^{(n)}\|_{\mathcal{A}} \\ &\le \rho \|u_J - u^{(n)}\|_{\mathcal{A}}, \quad n = 0, 1, \dots \quad , \end{aligned}$$

i.e., the $\mathcal{A}$-norm of the error is after each iteration of the multilevel method reduced at least by the factor ρ that does not depend on the number of levels in the multilevel method. This property is in literature known as *the uniform convergence* of multilevel methods, see, e.g., [27, 26].

For approximations $u^{(n)}$ and $u^{(n+1)}$ computed using the V-cycle scheme with pre-smoothing (Algorithm 2), holds (see equation (2.9))

$$u_J - u^{(n+1)} = E(u_J - u^{(n)}), \tag{3.3}$$

where

$$E = (I - \mathcal{B}_0^{-1}\mathcal{A})(I - \mathcal{B}_1^{-1}\mathcal{A}) \cdots (I - \mathcal{B}_J^{-1}\mathcal{A}) : V_J \to V_J. \tag{3.4}$$

Since E and E^* are adjoint w.r.t. the $\mathcal{A}$-inner product (see Section 2.3), their $\mathcal{A}$-norms coincide (see, e.g., [8, Proposition 2.7]), i.e.,

$$\|E\|_{\mathcal{A}} = \|E^*\|_{\mathcal{A}}. \tag{3.5}$$

The results mentioned above for the V-cycle scheme with post-smoothing are therefore valid also for the version with pre-smoothing.

In order to show that (3.2) holds, we consider the following five (rather technical) assumptions inspired by [27, Section 5]. The first assumption regards the (geometric) relationship between the restrictions of the operator $\mathcal{A}$ to the subspaces V_j and the smoothers $\mathcal{B}_j$.

A1: There exists a constant $\omega \in (0, 2)$ *independent of* J such that for all $j = 1, \dots, J$ (cf. [27, p. 305])

$$\langle \mathcal{A}v, v\rangle \leq \omega\langle \mathcal{B}_j v, v\rangle, \quad \text{for all } v \in V_j. \tag{3.6}$$

The convergence analysis of multilevel methods typically assumes that the problem on the coarsest level is solved exactly; see, e.g., [27, 26]. Here we weaken this assumption and allow inexact solver satisfying the following assumption. This assumption is an analogy of **A1**.

A2: There exists a constant $\omega_0 \in (0, 2)$ such that

$$\langle \mathcal{A}v, v\rangle \leq \omega_0\langle \mathcal{B}_0 v, v\rangle, \quad \text{for all } v \in V_0. \tag{3.7}$$

In order to formulate the next assumptions, we introduce the concept of splitting into subspaces, see, e.g., [27, Section 5], [26], [21, Chapter 1], [12] and the references therein. This concept is here used only as a tool to show the uniform convergence of the multilevel methods and it does not enter the practical computation. The concept of splitting into subspaces is also used in the context of operator preconditioning, see, e.g., [15] and references therein.

Motivated by [27, Section 5], we consider splitting of the space V_J into subspaces,

$$W_0 := V_0 \quad \text{and} \quad W_j \subset V_j, \quad j = 0, 1, \dots, J, \tag{3.8}$$

such that each function v from V_J can be *uniquely* represented as

$$v = \sum_{j=0}^{J} w_j, \quad w_j \in W_j.$$

In other words the space V_J is the direct sum of subspaces W_j

$$V_J = \bigoplus_{j=0,\dots,J} W_j.$$

Each subspace W_j, $j = 0, 1, \dots, J$, is considered with its own inner product and the associated norm

$$(\cdot,\cdot)_j : W_j \times W_j \to \mathbb{R} \quad \text{and} \quad \|\cdot\|_j := \sqrt{(\cdot,\cdot)_j}. \tag{3.9}$$

Since each $v \in V_J$ can be uniquely written as $v = \sum_{j=0}^{J} w_j$, $w_j \in W_j$, the individual norms $\|\cdot\|_j$ induce a norm in the space V_J

$$\|v\|_S^2 := \sum_{j=0}^{J} \|w_j\|_j^2. \tag{3.10}$$

This norm is in literature called the additive Schwarz norm; see, e.g., [21, Definition 2.1.1].

The following assumptions describe the required relations between the introduced splitting and the multilevel framework; cf. [27, Section 5].

A3: There exists a constant $C_S > 0$ *independent of* J such that for all $v \in V_J$ holds

$$\|v\|_S^2 \le C_S \|v\|_V^2, \tag{3.11}$$

where $\|v\|_S$ is defined in (3.10).

A4: There exists a constant $C_{\mathcal{B}} > 0$ *independent of* J such that for all $j = 1, \ldots, J$,

$$\langle \mathcal{B}_j w, w \rangle \le C_{\mathcal{B}} \|w\|_j^2, \quad \text{for all } w \in W_j. \tag{3.12}$$

A5: There exist constants γ_{jk}, $j = 0, 1, \ldots, J$, $k = 0, 1, \ldots, j$, such that for all $v \in V_k$ and all $w \in W_j$ holds

$$\langle \mathcal{A} v, w \rangle \le \gamma_{jk} \|v\|_V \|w\|_j. \tag{3.13}$$

More importantly, forming the symmetric matrix

$$\mathbf{M} \in \mathbb{R}^{(J+1)\times(J+1)}, \quad \mathbf{M}_{jk} := \gamma_{jk} \quad \text{for } j \ge k,$$

there exist a constant $\Gamma > 0$ *independent of* J that bounds the spectral radius of $\mathbf{M}$ from above[1].

Let us further denote $C_{\mathcal{B}_0} > 0$ the constant such that

$$\langle \mathcal{B}_0 w, w \rangle \le C_{\mathcal{B}_0} \|w\|_0^2 \quad \text{for all } w \in W_0. \tag{3.14}$$

Since $\langle \mathcal{B}_0 \cdot, \cdot \rangle$ induces a norm in the finite-dimensional space W_0, where all norms are topologically equivalent, the constant $C_{\mathcal{B}_0}$ exists; see Section 2.2.

Providing the assumptions **A1-A5** are satisfied, the contraction property (3.2) holds with the contraction factor

$$\rho = \sqrt{1 - \frac{2 - \max\{\omega, \omega_0\}}{\frac{C_S}{c_A}\left(\max\left\{\sqrt{C_{\mathcal{B}}}, \sqrt{C_{\mathcal{B}_0}}\right\} + \frac{\Gamma}{c_{\mathcal{A}}} \max\{\sqrt{\omega}, \sqrt{\omega_0}\}\right)^2}} < 1, \tag{3.15}$$

where $c_{\mathcal{A}}$ is the coercivity constant of the *infinite-dimensional* operator $\mathcal{A}$ defined in (2.3). Proof of this statement, which is inspired by the proof of Theorem 5.1 in [27, Section 5], is included in Appendix.

Let us summarise the results of this section in the following theorem.

[1]The assumption **A5** is in the literature known as the Cauchy-Schwarz type inequality, [27, Section 5].

Theorem 1. *Let* $u^{(0)} \in V_J$ *be an arbitrary initial approximation to the solution of* $\mathcal{A}u_J = b$ *in* $V_J^{\#}$*. Let*

$$u^{(n)}, \quad n = 1, 2, \ldots,$$

*be the approximations computed by the V-cycle scheme - Algorithm 2 or Algorithm 3. Providing that Assumptions **A1-A5** are satisfied,* $u^{(n)}$ *converges to the solution* u_J *and there holds*

$$\|u_J - u^{(n+1)}\|_{\mathcal{A}} \leq \rho \|u_J - u^{(n)}\|_{\mathcal{A}}, \quad n = 0, 1, 2, \ldots, \quad , \tag{3.16}$$

where ρ *is the contraction factor* (3.15) *independent of the number of levels in the V-cycle scheme.*

4. Convergence of multilevel methods for the model problem

In this chapter we consider the application of the presented multilevel methods to the model problem (1.5) and its finite element discretization. After defining the hierarchy of the finite element spaces we briefly comment on the choice of the smoothers and the solvers on the coarsest level. Further we discuss the verification of the assumptions **A1-A5**. The chapter ends with commenting on the dependence of the convergence behavior on the mesh size of the initial triangulation.

Definition of spaces

Consider the setting and model problem (1.5) presented in Section 1.2. Let V be the Sobolev space $H_0^1(\Omega)$ equipped with the inner product and the associated norm

$$(u, v)_V := \int_\Omega \nabla u \cdot \nabla v, \quad \text{for all } u, v \in V,$$
$$\|u\|_V := |u|_{H^1}, \quad \text{for all } u \in V.$$

The subspace V_j from the definition of the multilevel framework (Section 2.2) will be defined as the finite element subspaces corresponding to the uniformly refined triangulations. Let T_0 be an initial triangulation of Ω and let $T_1, \dots, T_J$ be the triangulations obtained by successive uniform refinements of T_0, i.e., the triangles in T_{j+1} are generated by dividing the triangles in T_j into four congruent subtriangles. It follows from the construction that the corresponding mesh sizes $h_0, \dots, h_J$ satisfy $h_j = h_0/2^j$, $j = 0, \dots, J$ and that the shape parameters (see (1.11))

$$C_{T_j} := \max_{K \in T_j} \frac{h_K}{\rho_K} \tag{4.1}$$

coincide for all $j = 0, \dots, J$, i.e., $C_{T_j} = C_{T_0}$. Moreover, let $K^{(j-1)} \in T_{j-1}$ be the element in the coarser triangulation containing $K^{(j)}$ and let $K^{(0)} \in T_0$ be the element in the initial triangulation containing both $K^{(j)}$ and $K^{(j-1)}$ then

$$|K^{(j)}| = \frac{|K^{(j-1)}|}{4} = \frac{|K^{(0)}|}{4^j}, \tag{4.2}$$

where $|K^{(j)}|$ denote its Lebesgue measure.

Using this construction, the space V_j is, for each $j = 0, \dots, J$, defined as the finite element space of continuous piecewise linear functions on T_j that vanish on the boundary $\partial\Omega$.

Smoothing

Smoothing is in the multilevel methods typically done using few iterations of a classic stationary iterative method such as the Richardson method, the (damped)

Jacobi method, or the (damped) Gauss-Seidel method, see, e.g., [6, Chapter 2], [14, Chapter 11].

As an example, we consider smoothing performed by one iteration of the Richardson method. For each $j = 1, \dots, J$, let $\mathbf{I_j}$ denote the identity matrix in $\mathbb{R}^{N_j}$ and let $\sigma(\mathbf{A_j})$ be the spectral radius of the matrix $\mathbf{A_j}$. The corresponding operators $\mathcal{B}_j = \Phi_j \mathbf{B_j}$, $j = 1, \dots, J$, are defined as

$$\mathbf{B_j} := \sigma(\mathbf{A_j})\mathbf{I_j}.$$

Since $\mathbf{A_j}$ is a symmetric matrix, for all $j = 1, \dots, J$, holds

$$\mathbf{v}^* \mathbf{A_j} \mathbf{v} \leq \sigma(\mathbf{A_j}) \mathbf{v}^* \mathbf{v} = \mathbf{v}^* \mathbf{B_j} \mathbf{v}, \quad \text{for all } \mathbf{v} \in \mathbb{R}^{N_j},$$

or equivalently in the operator form (see Section 2.4)

$$\langle \mathcal{A} v, v \rangle \leq \langle \mathcal{B}_j v, v \rangle, \quad \text{for all } v \in V_j,$$

The assumption **A1** is thus satisfied with $\omega = 1$.

We remark that it is also possible to define the operators $\mathcal{B}_j$, $j = 1, \dots, J$, such that they correspond to two or more iterations of a stationary iterative method, see, e.g., [27, p. 293].

Solver on the coarsest level

The solver on the coarsest level is the literature typically assumed to be exact; see, e.g., [26, Algorithm 3.7], [27, p. 294]. This is represented by defining the operator $\mathcal{B}_0$ as the restriction of $\mathcal{A}$ to V_0, i.e.,

$$\langle \mathcal{B}_0 u, v \rangle := \langle \mathcal{A} u, v \rangle, \quad \text{for all } u, v \in V_0. \tag{4.3}$$

The assumption **A2** is then trivially satisfied with $\omega_0 = 1$.

The framework described in Chapters 2 and 3 allow us to consider also an inexact solvers. As an example we consider the solver composed of one iteration of the Richardson method, i.e.,

$$\mathcal{B}_0 = \Phi_0 \mathbf{B_0}, \quad \mathbf{B_0} := \sigma(\mathbf{A_0})\mathbf{I_0}, \tag{4.4}$$

where $\mathbf{I_0}$ denotes the identity matrix in $\mathbb{R}^{N_0}$ and $\sigma(\mathbf{A_0})$ is the spectral radius of $\mathbf{A_0}$. The assumption **A2** is for this choice again satisfied with $\omega_0 = 1$.

Splitting into subspaces

In order to verify the assumptions **A3-A5**, the splitting (3.8) - (3.9) of V_J has to be chosen. We consider the L^2-like orthogonal splitting presented in [27, Section 7] that enable us to verify the assumption **A3-A5** for the problem (1.5) with the weak solution $u \in H^1(\Omega)$. In the cases where the problem (1.5) has a regular weak solution $u \in H^2(\Omega)$ one can also use the $\mathcal{A}$ orthogonal splitting; see the discussion in [27, Sections 6-7].

Let $(\cdot,\cdot)_{L^2,T_0} : V \times V \to \mathbb{R}$ denote the *scaled* L^2 inner product on V

$$(u, v)_{L^2,T_0} := \sum_{K \in T_0} \frac{1}{|K|} \int_K uv, \quad \text{for all } u, v \in V, \tag{4.5}$$

and let for each $j = 0, 1, \ldots, J$, $Q_j : V \to V_j$ be the projection onto V_j which is orthogonal with respect to this inner product, i.e.,

$$(Q_j u, v)_{L^2,T_0} = (u, v)_{L^2,T_0}, \quad \text{for all } u \in V, v \in V_j.$$

The projections Q_j are used to decompose the space V_J into the space $W_0 = V_0$ and the orthogonal complements (cf. [27, Section 7])

$$W_j := \{Q_j v - Q_{j-1} v \; ; \; v \in V\}.$$

The inner product on W_0 is chosen as the restriction of the inner product on V to W_0, i.e.,

$$(\cdot,\cdot)_0 = W_0 \times W_0 \to \mathbb{R}, \quad (u, v)_0 := (u, v)_V = (u, v)_{H^1}, \quad \text{for all } u, v \in W_0.$$

The spaces W_j, $j = 1, \ldots, J$, are considered with the *scaled* L^2 inner products

$$(\cdot,\cdot)_j = W_j \times W_j \to \mathbb{R}, \quad (u, v)_j := \sum_{K\in T_j} \frac{1}{|K|} \int_K uv, \quad \text{for all } u, v \in W_j. \tag{4.6}$$

The definition of the scaled inner products (4.6) takes into the account the sizes of the triangles in the corresponding triangulations and use their sizes as weights. This choice will enable us to verify the assumptions **A3**-**A5**. Note that by (4.2) there holds for all $u, v \in W_j$, $j = 1, \ldots, J$,

$$(u, v)_j = \sum_{K\in T_j} \frac{1}{|K|} \int_K uv = 4^j \sum_{K\in T_0} \frac{1}{|K|} \int_K uv = 4^j (u, v)_{L^2,T_0}.$$

Verification of the assumptions A3 and A5

Let us now comment on the verification of the assumptions **A3**, **A5**. In order to do that we will need the following Lemma 2 and Lemma 3, which are proven in [3, Theorem 7.6][1] and [26, Lemma 6.1], [27, Lemma 6.1], respectively. The proofs of these lemmas use the properties of the infinite-dimensional problem.

Lemma 2. *There exists a constant $C_S > 0$ depending only on the domain Ω and the shape parameter C_{T_0} of the triangulations such that for all $v \in V_J$ holds*

$$\|v\|_S^2 = |Q_0 v|_{H^1}^2 + \sum_{j=1}^{J} 4^j \|(Q_j - Q_{j-1})v\|_{L^2,T_0}^2 \le C_S |v|_{H^1}^2. \tag{4.7}$$

Lemma 3. *There exist a constant $C_\alpha > 0$ depending only on the domain Ω, the shape parameter C_{T_0} of the triangulations, boundedness of $\mathcal{A}$ and boundedness of the derivative of $\tilde{a}$ such that for all $j = 0, 1, \ldots, J$ and $k = 0, 1, \ldots, j$ holds*

$$\langle \mathcal{A} v, w \rangle \le C_\alpha \left(\frac{\sqrt{2}}{2} \right)^{j-k} |v|_{H^1} 2^j \|w\|_{L^2,T_0}, \quad \textit{for all } v \in V_k, w \in V_j. \tag{4.8}$$

[1]See also [27, Section 7], [21, Sections 2.2-2.4] and [20].

The assumption **A3** follows directly from Lemma 2. To verify the assumption **A5** we use Lemma 3 which gives (3.13) with $\gamma_{jk} = \left(\frac{\sqrt{2}}{2}\right)^{j-k}$. The matrix $\mathbf{M}$ is then of the form

$$\mathbf{M} = C_\alpha \begin{bmatrix} 1 & \frac{\sqrt{2}}{2} & \left(\frac{\sqrt{2}}{2}\right)^2 & \cdots & \left(\frac{\sqrt{2}}{2}\right)^J \\ \frac{\sqrt{2}}{2} & 1 & \frac{\sqrt{2}}{2} & \cdots & \left(\frac{\sqrt{2}}{2}\right)^{J-1} \\ \left(\frac{\sqrt{2}}{2}\right)^2 & \frac{\sqrt{2}}{2} & 1 & \cdots & \left(\frac{\sqrt{2}}{2}\right)^{J-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \left(\frac{\sqrt{2}}{2}\right)^J & \left(\frac{\sqrt{2}}{2}\right)^{J-1} & \left(\frac{\sqrt{2}}{2}\right)^{J-2} & \cdots & 1 \end{bmatrix}.$$

The following derivation is inspired by the proof of [26, Lemma 4.6]. The spectral radius $\sigma(\mathbf{M})$ of matrix $\mathbf{M}$ can be estimated using the Gershgorin theorem (see e.g., [22, Theorem 2.1], [24, Corollary 1.12]) as

$$\sigma(\mathbf{M}) \leq \max_{k=0,1,\ldots,J} \sum_{j=0}^{J} \mathbf{M}_{jk}.$$

Then

$$\max_{k=0,1,\ldots,J} \sum_{j=0}^{J} \mathbf{M}_{jk} < 2C_\alpha \sum_{j=0}^{J} \left(\frac{\sqrt{2}}{2}\right)^j < 2C_\alpha \sum_{j=0}^{+\infty} \left(\frac{\sqrt{2}}{2}\right)^j = \frac{2C_\alpha}{1-\frac{\sqrt{2}}{2}} = 2C_\alpha(2+\sqrt{2}).$$

Finally, taking the constant Γ as $2C_\alpha(2+\sqrt{2})$ yields that the assumption **A5** is fulfilled.

Verification of the assumption A4

We comment on the verification of the assumption **A4**. The algebraic representation of the inequality (3.12) considered on the whole V_j reads as

$$\sigma(\mathbf{A_j})\mathbf{v}^*\mathbf{v} \leq C_\mathcal{B}\mathbf{v}^*\mathbf{M_j}\mathbf{v}, \quad \text{for all } \mathbf{v} \in \mathbb{R}^{N_j}, \quad j = 1,\ldots,J, \tag{4.9}$$

where $\mathbf{M_j}$ denotes the mass matrix corresponding to the inner product $(\cdot,\cdot)_j$, i.e.,

$$\left(\mathbf{M_j}\right)_{mn} := (\phi_n^{(j)}, \phi_m^{(j)})_j = \sum_{K\in T_j} \frac{1}{|K|}\int_K \phi_n^{(j)}\phi_m^{(j)}, \quad m,n = 1,\ldots,N_j, \tag{4.10}$$

where $\phi_n^{(j)}$, $n = 1,\ldots,N_j$ are the basis finite element functions of the space V_j.

Inequality (4.9) can be equivalently reformulated as

$$C_\mathcal{B} \geq \sigma(\mathbf{A_j})\frac{1}{\frac{\mathbf{v}^*\mathbf{M_j}\mathbf{v}}{\mathbf{v}^*\mathbf{v}}}, \quad \text{for all } \mathbf{v} \in \mathbb{R}^{N_0}, \mathbf{v} \neq \mathbf{0}, \quad j = 1,\ldots,J. \tag{4.11}$$

Thus, in order to find a constant $C_\mathcal{B} > 0$ satisfying (4.11), it is sufficient to give an upper bound on $\sigma(\mathbf{A_j})$ and lower bound on $\frac{\mathbf{v}^*\mathbf{M_j}\mathbf{v}}{\mathbf{v}^*\mathbf{v}}$ for all nonzero $\mathbf{v} \in \mathbb{R}^{N_j}$ valid for all $j = 1,\ldots,J$. If those bounds are independent of J, the constant $C_\mathcal{B}$ satisfies the same.

Let us now focus on the spectral properties of the scaled mass matrix $\mathbf{M_j}$. The result in [9, Equation (1.116)] gives for the standard mass matrix

$$\left(\mathbf{P_j}\right)_{mn} := \int_\Omega \phi_n^{(j)} \phi_m^{(j)} = \sum_{K \in T_j} \int_K \phi_n^{(j)} \phi_m^{(j)}, \quad m, n = 1, \dots, N_j, \tag{4.12}$$

that there exist constants $c_\mathbf{P} > 0$ and $C_\mathbf{P} > 0$ depending only on the shape parameter C_{T_0} such that

$$c_\mathbf{P} h_j^2 \le \frac{\mathbf{v}^* \mathbf{P_j} \mathbf{v}}{\mathbf{v}^* \mathbf{v}} \le C_\mathbf{P} h_j^2, \quad \text{for all } \mathbf{v} \in \mathbb{R}^{N_j}. \tag{4.13}$$

Although we do not give the complete proof here, we believe that using the same technique as in the proof of (4.13), it can be shown that there exist constants $c_\mathbf{M} > 0$ and $C_\mathbf{M} > 0$ depending only on the shape parameter C_{T_0} such that

$$c_\mathbf{M} \le \frac{\mathbf{v}^* \mathbf{M_j} \mathbf{v}}{\mathbf{v}^* \mathbf{v}} \le C_\mathbf{M}, \quad \text{for all } \mathbf{v} \in \mathbb{R}^{N_j}. \tag{4.14}$$

The reason is that the elements in the matrices $\mathbf{M_j}$ and $\mathbf{P_j}$ differs only in the weights which are of the order h_j^{-2}. Thus $c_\mathbf{M}$ gives the desired lower bound.

It remains to show the existence of the upper bound on $\sigma(\mathbf{A_j})$. The algebraic representation of the inequality (2.4) considered on the finite-dimensional space V_j reads as

$$c_\mathcal{A} \mathbf{v}^* \mathbf{L_j} \mathbf{v} \le \mathbf{v}^* \mathbf{A_j} \mathbf{v} \le C_\mathcal{A} \mathbf{v}^* \mathbf{L_j} \mathbf{v}, \quad \text{for all } \mathbf{v} \in \mathbb{R}^{N_j}, \tag{4.15}$$

where $\mathbf{L_j}$ denotes the matrix corresponding to the inner product $(\cdot, \cdot)_j$, i.e.,

$$\left(\mathbf{L_j}\right)_{mn} = \int_\Omega \nabla \phi_n^{(j)} \cdot \nabla \phi_m^{(j)}, \quad m, n = 1, \dots, N_j,$$

where $\phi_n^{(j)}$, $n = 1, \dots, N_j$ are the basis finite element function of the space V_j. The result [9, Theorem 1.33] gives the existence of the constant $C_\mathbf{L} > 0$ depending only on the shape parameter C_{T_0} such that the spectral radius $\sigma(\mathbf{L_j}) \le C_\mathbf{L}$.

Summarizing, defining

$$C_\mathcal{B} = c_\mathbf{M} \cdot C_\mathcal{A} \mathbf{v} \cdot C_\mathbf{L}$$

ensures that the assumption **A4** is satisfied.

Dependence of the convergence behavior on the mesh size of the initial triangulation

At this point we have verified or comment on the verification of all of the assumptions **A1-A5** of Theorem 1. We already know that the contraction factor ρ (3.15) is independent of the number of levels of the multilevel scheme. However as it will be shown below it can depend on the mesh size of the initial triangulation, via the constant $C_{\mathcal{B}_0}$. Let us discuss the dependence of the constant $C_{\mathcal{B}_0}$ for the above mentioned choices of the solver on the coarsest level.

For the multilevel method with exact solver on the coarsest level, the inequality (3.14) reads as

$$\langle \mathcal{A} w, w \rangle \le C_{\mathcal{B}_0} \|w\|_{H^1}, \quad \text{for all } w \in W_0. \tag{4.16}$$

Using (2.4) and (1.9), the constant $C_{\mathcal{B}_0}$ can be taken as

$$C_{\mathcal{B}_0} = C_{\mathcal{A}} = \max_{x \in \overline{\Omega}} \tilde{a}(x),$$

and thus depends only on the properties of the infinite-dimensional problem.

The situation is different for the method with the inexact solver on the coarsest level defined in (4.4). We will use the algebraic formulation of (3.14)

$$\sigma(\mathbf{A_0})\mathbf{w}^*\mathbf{w} \leq C_{\mathcal{B}_0}\mathbf{w}^*\mathbf{L_0}\mathbf{w}, \quad \text{for all } \mathbf{w} \in \mathbb{R}^{N_0}, \tag{4.17}$$

where $\mathbf{L_0}$ denotes the matrix corresponding to the inner product $(\cdot,\cdot)_0$, i.e.,

$$(\mathbf{L_0})_{mn} = \int_\Omega \nabla\phi_n^{(0)} \cdot \nabla\phi_m^{(0)}, \quad m, n = 1, \ldots, N_0,$$

where $\phi_n^{(0)}$, $n = 1, \ldots, N_0$ are the basis finite element function of the space V_0. Inequality (4.17) can be equivalently reformulated as

$$\frac{1}{C_{\mathcal{B}_0}} \leq \frac{1}{\sigma(\mathbf{A_0})} \frac{\mathbf{w}^*\mathbf{L_0}\mathbf{w}}{\mathbf{w}^*\mathbf{w}}, \quad \text{for all } \mathbf{w} \in \mathbb{R}^{N_0}, \mathbf{w} \neq \mathbf{0}.$$

Taking the minimum over all vectors $\mathbf{w} \in \mathbb{R}^{N_j}$ we have

$$C_{\mathcal{B}_0} \geq \sigma(\mathbf{A_0})\lambda_{\min}^{-1}(\mathbf{L_0}). \tag{4.18}$$

Note that the term $\sigma(\mathbf{A_0})$ comes from the choice of the Richardson method, whereas the term $\lambda_{\min}^{-1}(\mathbf{L_0})$ comes from the norm on the space W_0. From (4.18) we see that the constant $C_{\mathcal{B}_0}$ is bounded from below by the spectral radius of $\mathbf{A_0}$ multiplied by the inverse of the smallest eigenvalue of the matrix $\mathbf{L_0}$.

In concrete examples with additional assumptions on the mesh geometry we are able to use the Fourier analysis and show that

$$\lambda_{\min}(\mathbf{L_0}) = \mathcal{O}(h_0^2) \quad \text{and} \quad \lambda_{\max}(\mathbf{L_0}) = \mathcal{O}(1), \tag{4.19}$$

see, e.g., [9, pp. 58-59]. Using (2.4) restricted on the finite-dimensional space W_0, which in the algebraic representation reads as

$$c_{\mathcal{A}}\mathbf{w}^*\mathbf{L_0}\mathbf{w} \leq \mathbf{w}^*\mathbf{A_0}\mathbf{w} \leq C_{\mathcal{A}}\mathbf{w}^*\mathbf{L_0}\mathbf{w}, \quad \text{for all } \mathbf{w} \in \mathbb{R}^{N_0}, \tag{4.20}$$

gives together with (4.19) that $\sigma(\mathbf{A_0}) = \mathcal{O}(1)$. Thus

$$C_{\mathcal{B}_0} = \mathcal{O}(h_0^{-2}). \tag{4.21}$$

This indicates that the constant $C_{\mathcal{B}_0}$ and thus also the contraction factor ρ can in general depend on the mesh size of the initial triangulation.

Conclusion

The aim of this thesis was to study the convergence behavior of the multilevel methods with inexact solver on the coarsest level. Such methods are often used in practice. The analysis presented in literature is typically carried out under the assumption that the problem on the coarsest level is solved exactly. Our exposition is built upon the articles [26, 27]. We present a coherent abstract description of the multilevel methods in the Hilbert spaces and discuss in detail their operator and matrix-vector formulations. In compliance with our aim, we allow inexact solve on the coarsest level, and modify existing convergence proof. The rate of the convergence is independent of the number of levels.

Further we consider a boundary value problem formulated using a second order elliptic PDE and apply the described multilevel methods to its finite element discretization. For the choice of the exact solver on the coarsest level the bound on the rate of convergence is independent on the mesh size of the initial triangulation. On the other hand, the presented discussion indicates that for the choice of the inexact solver the convergence bound *can depend* on the mesh size of the initial triangulation.

Many questions remain open. In the future we would like to generalize the presented analysis to an arbitrary inexact solver on the coarsest level, e.g., the conjugate gradient method or the inaccurate direct solvers based, e.g., on incomplete LU factorization.

This thesis focus on the a priori convergence analysis. In practice it is important to have a reliable a posteriori error estimates. The results in this field are typically derived also under the assumption on the exact solve on the coarsest level; see e.g., [2]. We would like to weaken this assumption to allow inexact solvers. Another interesting question is to take into the account the effects of finite precision arithmetic on the convergence properties of the multilevel methods.

Appendix

In the Appendix we give the proof of the contraction property of the error propagation operator presented in Chapter 3. More precisely, let E^* be the operator defined in (2.11), i.e.,

$$E^* = (I - \mathcal{B}_J^{-1}\mathcal{A})(I - \mathcal{B}_{J-1}^{-1}\mathcal{A}) \cdots (I - \mathcal{B}_0^{-1}\mathcal{A}) : V_J \to V_J.$$

We will prove that if the assumptions **A1-A5** formulated in Chapter 3 are satisfied, there holds

$$\|E^*\|_{\mathcal{A}} = \sup_{v \in V_J; \|v\|_{\mathcal{A}}=1} \|E^* v\|_{\mathcal{A}} \le \rho, \tag{i}$$

where

$$\rho = \sqrt{1 - \frac{2 - \max\{\omega, \omega_0\}}{\frac{C_S}{c_{\mathcal{A}}}\left(\max\left\{\sqrt{C_{\mathcal{B}}}, \sqrt{C_{\mathcal{B}_0}}\right\} + \frac{\Gamma}{c_{\mathcal{A}}}\max\{\sqrt{\omega}, \sqrt{\omega_0}\}\right)^2}} < 1.$$

The constants $c_{\mathcal{A}}$ and $C_{\mathcal{B}_0}$ are defined in (2.3) and (3.14) and the constants ω, ω_0, C_S, $C_{\mathcal{B}}$ and Γ are specified in the assumptions **A1-A5**. The following proof, which is rather technical, is inspired by the proof of Theorem 5.1 in [27, Section 5].

Let us first establish a useful notation. For each $j = 1, \ldots, J-1$, let E_j be the operator defined as

$$E_j^* := (I - \mathcal{B}_j^{-1}\mathcal{A}) \cdots (I - \mathcal{B}_1^{-1}\mathcal{A})(I - \mathcal{B}_0^{-1}\mathcal{A}) : V_J \to V_J.$$

The operator E_j^* can be seen as the error propagation operator corresponding to the V-cycle scheme where the post-smoothing is done only on levels $1, \ldots, j$. Denoting further $E_J^* := E^*$ and $E_{-1}^* := I$, the following relation hold

$$E_j^* = (I - \mathcal{B}_j^{-1}\mathcal{A})E_{j-1}^*, \qquad j = 0, \ldots, J, \tag{ii}$$

$$E_{j-1}^* - E_j^* = \mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^*, \qquad j = 0, \ldots, J,$$

$$I - E_{j-1}^* = \sum_{k=0}^{j-1} E_{k-1}^* - E_k^* = \sum_{k=0}^{j-1} \mathcal{B}_k^{-1}\mathcal{A}E_{k-1}^*, \qquad j = 1, \ldots, J. \tag{iii}$$

In order to show (i), let v be an arbitrary function from V_J, such that $\|v\|_{\mathcal{A}} = 1$. The term $\|E^* v\|_{\mathcal{A}}^2$ can be written as

$$\begin{aligned} \|E^* v\|_{\mathcal{A}}^2 &= \|v\|_{\mathcal{A}}^2 - \|v\|_{\mathcal{A}}^2 + \|E_0^* v\|^2 - \|E_0^* v\|^2 \\ &\qquad + \cdots + \|E_{J-1}^* v\|^2 - \|E_{J-1}^* v\|^2 + \|E^* v\|_{\mathcal{A}}^2 \\ &= \|v\|_{\mathcal{A}}^2 - \sum_{j=0}^{J} \left(\|E_{j-1}^* v\|_{\mathcal{A}}^2 - \|E_j^* v\|_{\mathcal{A}}^2 \right). \end{aligned} \tag{iv}$$

The differences $\|E_{j-1}^* v\|_{\mathcal{A}}^2 - \|E_j^* v\|_{\mathcal{A}}^2$, $j = 0, 1, \ldots, J$, can be rewritten using (ii) and the self-adjointness of $\mathcal{A}$ as

$$\begin{aligned}
&\|E_{j-1}^* v\|_{\mathcal{A}}^2 - \|E_j^* v\|_{\mathcal{A}}^2 \\
&\qquad = \langle \mathcal{A}E_{j-1}^* v, E_{j-1}^* v\rangle - \langle \mathcal{A}E_j^* v, E_j^* v\rangle \\
&\qquad = \langle \mathcal{A}E_{j-1}^* v, E_{j-1}^* v\rangle - \langle \mathcal{A}(I - \mathcal{B}_j^{-1}\mathcal{A})E_{j-1}^* v, (I - \mathcal{B}_j^{-1}\mathcal{A})E_{j-1}^* v\rangle \\
&\qquad = \langle \mathcal{A}E_{j-1}^* v, E_{j-1}^* v\rangle - \Big(\langle \mathcal{A}E_{j-1}^* v, E_{j-1}^* v\rangle - \langle \mathcal{A}E_{j-1}^* v, \mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^* v\rangle \\
&\qquad\qquad - \langle \mathcal{A}\mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^* v, E_{j-1}^* v\rangle + \langle \mathcal{A}\mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^* v, \mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^* v\rangle\Big) \\
&\qquad = 2\langle \mathcal{A}\mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^* v, E_{j-1}^* v\rangle - \langle \mathcal{A}\mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^* v, \mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^* v\rangle. \qquad \text{(v)}
\end{aligned}$$

Bounding

$$\langle \mathcal{A}\mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^* v, \mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^* v\rangle, \quad j = 0, 1, \ldots, J,$$

in (v) using the assumptions **A1** and **A2**, we have

$$\begin{aligned}
&\|E_{j-1}^* v\|_{\mathcal{A}}^2 - \|E_j^* v\|_{\mathcal{A}}^2 \\
&\qquad \le 2\langle \mathcal{A}\mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^* v, E_{j-1}^* v\rangle - \max\{\omega, \omega_0\}\,\langle \mathcal{B}_j\mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^* v, \mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^* v\rangle \\
&\qquad = (2 - \max\{\omega, \omega_0\})\,\langle \mathcal{A}E_{j-1}^* v, \mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^* v\rangle. \qquad \text{(vi)}
\end{aligned}$$

Combining (iv) and (vi) yields

$$\|E^* v\|_{\mathcal{A}}^2 \le \|v\|_{\mathcal{A}}^2 - (2 - \max\{\omega, \omega_0\}) \sum_{j=0}^{J} \langle \mathcal{A}E_{j-1}^* v, \mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^* v\rangle, \qquad \text{(vii)}$$

The sum on the right-hand side of (vii), will be bounded using the following lemma. Its prove is for the ease of the exposition postpone for later.

Lemma 4. *Provided that the assumptions **A3-A5** are satisfied, there holds*

$$\|v\|_{\mathcal{A}}^2 \le K \sum_{j=0}^{J} \langle \mathcal{A}E_{j-1}^* v, \mathcal{B}_j^{-1}E_{j-1}^* v\rangle, \quad \textit{for all } v \in V_J, \qquad \text{(viii)}$$

where

$$K = \frac{C_S}{c_A}\left(\max\left\{\sqrt{C_{\mathcal{B}}}, \sqrt{C_{\mathcal{B}_0}}\right\} + \frac{\Gamma}{c_{\mathcal{A}}}\max\left\{\sqrt{\omega}, \sqrt{\omega_0}\right\}\right)^2.$$

Bounding the sum on the right-hand side of (vii) using Lemma 4 gives

$$\|E^* v\|_{\mathcal{A}}^2 \le \|v\|_{\mathcal{A}}^2 - \frac{(2 - \max\{\omega, \omega_0\})}{K}\|v\|_{\mathcal{A}}^2,$$

and consequently

$$\|E^* v\|_{\mathcal{A}} \le \sqrt{1 - \frac{(2 - \max\{\omega, \omega_0\})}{K}}\,\|v\|_{\mathcal{A}}.$$

Taking the supremum over all $v \in V_J$, $\|v\|_{\mathcal{A}} = 1$, yields (i).

The proof of Lemma 4 is given in the end of the Appendix. To be able to complete the proof of Lemma 4 we need to formulate and prove the following auxiliary Lemmas 5 and 6.

Lemma 5. *Let $u_0, u_1, \dots, u_J$ be arbitrary functions from V_J and for each $j = 0, 1, \dots, J$, let w_j be an arbitrary function from W_j. Provided that the assumption **A4** is satisfied, there holds*

$$\sum_{j=0}^{J}\langle \mathcal{A}u_j, w_j\rangle \le \max\left\{\sqrt{C_{\mathcal{B}}}, \sqrt{C_{\mathcal{B}_0}}\right\}\left(\sum_{j=0}^{J}\|w_j\|_j^2\right)^{\frac{1}{2}}\left(\sum_{j=0}^{J}\langle \mathcal{A}u_j, \mathcal{B}_j^{-1}\mathcal{A}u_j\rangle\right)^{\frac{1}{2}}. \quad \text{(ix)}$$

The following proof of Lemma 5 is inspired by the the proof of Lemma 5.2 in [27].

Proof of Lemma 5. Rewriting the left-hand side of (ix) as

$$\sum_{j=0}^{J}\langle \mathcal{A}u_j, w_j\rangle = \sum_{j=0}^{J}\langle \mathcal{B}_j\mathcal{B}_j^{-1}\mathcal{A}u_j, w_j\rangle = \sum_{j=0}^{J}(\mathcal{B}_j^{-1}\mathcal{A}u_j, w_j)_{\mathcal{B}_j}$$

and using the Cauchy–Schwarz inequality for each $\mathcal{B}_j$-inner product give

$$\begin{aligned}\sum_{j=0}^{J}\langle \mathcal{A}u_j, w_j\rangle = \sum_{j=0}^{J}(\mathcal{B}_j^{-1}\mathcal{A}u_j, w_j)_{\mathcal{B}_j} &\le \sum_{j=0}^{J}(\mathcal{B}_j^{-1}\mathcal{A}u_j, \mathcal{B}_j^{-1}\mathcal{A}u_j)_{\mathcal{B}_j}^{\frac{1}{2}}(w_j, w_j)_{\mathcal{B}_j}^{\frac{1}{2}}\\ &= \sum_{j=0}^{J}\langle \mathcal{B}_j\mathcal{B}_j^{-1}\mathcal{A}u_j, \mathcal{B}_j^{-1}\mathcal{A}u_j\rangle^{\frac{1}{2}}\langle \mathcal{B}_jw_j, w_j\rangle^{\frac{1}{2}}\\ &= \sum_{j=0}^{J}\langle \mathcal{A}u_j, \mathcal{B}_j^{-1}\mathcal{A}u_j\rangle^{\frac{1}{2}}\langle \mathcal{B}_jw_j, w_j\rangle^{\frac{1}{2}}.\end{aligned}$$

Using the Cauchy–Schwarz inequality

$$\sum_{j=0}^{J}a_jb_j \le \left(\sum_{j=0}^{J}a_j^2\right)^{\frac{1}{2}}\left(\sum_{j=0}^{J}b_j^2\right)^{\frac{1}{2}}, \quad a_j, b_j \in \mathbb{R},$$

for

$$a_j := \langle \mathcal{A}u_j, \mathcal{B}_j^{-1}\mathcal{A}u_j\rangle^{\frac{1}{2}} \quad \text{and} \quad b_j := \langle \mathcal{B}_jw_j, w_j\rangle^{\frac{1}{2}}$$

yields

$$\sum_{j=0}^{J}\langle \mathcal{A}u_j, w_j\rangle \le \left(\sum_{j=0}^{J}\langle \mathcal{A}u_j, \mathcal{B}_j^{-1}\mathcal{A}u_j\rangle\right)^{\frac{1}{2}}\left(\sum_{j=0}^{J}\langle \mathcal{B}_jw_j, w_j\rangle\right)^{\frac{1}{2}}.$$

Bounding $\langle \mathcal{B}_0w_0, w_0\rangle$ using the inequality (3.14) and $\langle \mathcal{B}_jw_j, w_j\rangle$, $j = 1, \dots, J$ using Assumption **A4**, we have

$$\sum_{j=0}^{J}\langle \mathcal{A}u_j, w_j\rangle \le \left(\sum_{j=0}^{J}\langle \mathcal{A}u_j, \mathcal{B}_j^{-1}\mathcal{A}u_j\rangle\right)^{\frac{1}{2}}\left(C_{\mathcal{B}_0}\|w_0\|_0^2 + C_{\mathcal{B}}\sum_{j=1}^{J}\|w_j\|_j^2\right)^{\frac{1}{2}},$$

which yields (ix). □

Lemma 6. *Providing that the assumption **A5** is satisfied, there holds*

$$\sum_{j=0}^{J}\sum_{k=0}^{J}\boldsymbol{M}_{jk}x_jy_k \le \Gamma\left(\sum_{j=0}^{J}x_j^2\right)^{\frac{1}{2}}\left(\sum_{j=0}^{J}y_j^2\right)^{\frac{1}{2}}, \quad \textit{for all } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{J+1}. \quad \text{(x)}$$

The following proof of Lemma 6 is inspired by the proof of Lemma 4.6 in [26].

Proof of Lemma 6. Let $\mathbf{x}, \mathbf{y}$ be arbitrary vectors from $\mathbb{R}^{J+1}$. Writting the left-hand side of (x) in matrix-vector form we have

$$\sum_{j=0}^{J}\sum_{J=0}^{J} \mathbf{M}_{jk} x_j y_k = \mathbf{y}^* \mathbf{M} \mathbf{x}. \tag{xi}$$

Let $\|\cdot\|$ denote the Euclidean norm in $\mathbb{R}^{J+1}$. The Cauchy-Schwarz inequality gives

$$\mathbf{y}^* \mathbf{M} \mathbf{x} \le \|\mathbf{M}\mathbf{x}\| \|\mathbf{y}\|. \tag{xii}$$

Let $\sigma(\mathbf{M})$ denotes the spectral radius of matrix $\mathbf{M}$. Since $\mathbf{M}$ is a symmetric matrix there holds

$$\|\mathbf{M}\mathbf{x}\| \le \sigma(\mathbf{M}) \|\mathbf{x}\|, \tag{xiii}$$

see, e.g., [24, Corollary 1.8]. Combining (xi), (xii), (xiii) and bounding $\sigma(\mathbf{M})$ using Γ from above finishes the proof. □

Now we are ready to give the proof of the Lemma 4. It is inspired by the proofs of Theorem 5.1 in [27], and the proofs of Lemma 4.3 and Theorem 4.4 in [26].

Proof of Lemma 4. Let v be arbitrary function from V_J and consider its decomposition into the subspaces W_j, $j = 0, 1, \ldots, J$, i.e., $v = \sum_{j=0}^{J} w_j$, $w_j \in W_j$. Then

$$\begin{aligned}
\|v\|_{\mathcal{A}}^2 &= \sum_{j=0}^{J} \langle \mathcal{A} v, w_j \rangle \\
&= \sum_{j=0}^{J} \langle \mathcal{A} \left(E_{j-1}^* + I - E_{j-1}^* \right) v, w_j \rangle \\
&= \sum_{j=0}^{J} \langle \mathcal{A} E_{j-1}^* v, w_j \rangle + \sum_{j=0}^{J} \langle \mathcal{A} \left(I - E_{j-1}^* \right) v, w_j \rangle \\
&= \sum_{j=0}^{J} \langle \mathcal{A} E_{j-1}^* v, w_j \rangle + \sum_{j=1}^{J} \langle \mathcal{A} \left(I - E_{j-1}^* \right) v, w_j \rangle, \qquad \text{(since } E_{-1}^* = I\text{)}.
\end{aligned} \tag{xiv}$$

The first sum in (xiv) is bounded by Lemma 5 as

$$\begin{aligned}
&\sum_{j=0}^{J} \langle \mathcal{A} E_{j-1}^* v, w_j \rangle \\
&\quad \le \max\left\{ \sqrt{C_{\mathcal{B}}}, \sqrt{C_{\mathcal{B}_0}} \right\} \left(\sum_{j=0}^{J} \|w_j\|_j^2 \right)^{\frac{1}{2}} \left(\sum_{j=0}^{J} \langle \mathcal{A} E_{j-1}^* v, \mathcal{B}_j^{-1} \mathcal{A} E_{j-1}^* v \rangle \right)^{\frac{1}{2}} \\
&\quad = \max\left\{ \sqrt{C_{\mathcal{B}}}, \sqrt{C_{\mathcal{B}_0}} \right\} \left(\sum_{j=0}^{J} \langle \mathcal{A} E_{j-1}^* v, \mathcal{B}_j^{-1} \mathcal{A} E_{j-1}^* v \rangle \right)^{\frac{1}{2}} \|v\|_S,
\end{aligned} \tag{xv}$$

we have also used the definition of the additive Schwarz norm $\|\cdot\|_S$; see (3.10). Rewriting the second sum in (xiv) utilizing the relation (iii) gives

$$\sum_{j=1}^{J} \langle \mathcal{A} \left(I - E_{j-1}^* \right) v, w_j \rangle = \sum_{j=1}^{J} \sum_{k=0}^{j-1} \langle \mathcal{A} \mathcal{B}_k^{-1} \mathcal{A} E_{k-1}^* v, w_j \rangle. \tag{xvi}$$

Using the assumption **A5** and the coersivity of $\mathcal{A}$ leads to

$$\sum_{j=1}^{J}\sum_{k=0}^{j-1}\langle \mathcal{A}\mathcal{B}_k^{-1}\mathcal{A}E_{k-1}^{*}v, w_j\rangle \leq \frac{1}{c_{\mathcal{A}}}\sum_{j=1}^{J}\sum_{k=0}^{j-1}\gamma_{j,k}\|\mathcal{B}_k^{-1}\mathcal{A}E_{k-1}^{*}v\|_{\mathcal{A}}\|w_j\|_j$$
$$\leq \frac{1}{c_{\mathcal{A}}}\sum_{j=0}^{J}\sum_{k=0}^{J}\gamma_{j,k}\|\mathcal{B}_k^{-1}\mathcal{A}E_{k-1}^{*}v\|_{\mathcal{A}}\|w_j\|_j. \qquad \text{(xvii)}$$

Applying Lemma 6 for

$$x_j = \|w_j\|_j, \quad y_k = \|\mathcal{B}_k^{-1}\mathcal{A}E_{k-1}^{*}v\|_{\mathcal{A}},$$

gives

$$\sum_{j=0}^{J}\sum_{k=0}^{J}\gamma_{j,k}\|\mathcal{B}_k^{-1}\mathcal{A}E_{k-1}^{*}v\|_{\mathcal{A}}\|w_j\|_j \leq \Gamma\left(\sum_{j=0}^{J}\|\mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^{*}v\|_{\mathcal{A}}^2\right)^{\frac{1}{2}}\left(\sum_{j=0}^{J}\|w_j\|_j^2\right)^{\frac{1}{2}}$$
$$= \Gamma\left(\sum_{j=0}^{J}\langle \mathcal{A}\mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^{*}v, \mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^{*}v\rangle\right)^{\frac{1}{2}}\|v\|_S. \qquad \text{(xviii)}$$

Bounding $\langle \mathcal{A}\mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^{*}v, \mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^{*}v\rangle$, $j = 0, 1, \ldots, J$, using the assumptions **A1** and **A2** we have

$$\sum_{j=0}^{J}\langle \mathcal{A}\mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^{*}v, \mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^{*}v\rangle \leq \max\{\omega,\omega_0\}\sum_{j=0}^{J}\langle \mathcal{B}_j\mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^{*}v, \mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^{*}v\rangle$$
$$= \max\{\omega,\omega_0\}\sum_{j=0}^{J}\langle \mathcal{A}E_{j-1}^{*}v, \mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^{*}v\rangle. \qquad \text{(xix)}$$

Combining (xvi), (xvii), (xviii) and (xix) gives

$$\sum_{j=1}^{J}\langle \mathcal{A}\left(I - E_{j-1}^{*}\right)v, w_j\rangle$$
$$\leq \frac{\Gamma}{c_{\mathcal{A}}}\max\left\{\sqrt{\omega},\sqrt{\omega_0}\right\}\left(\sum_{j=0}^{J}\langle \mathcal{A}E_{j-1}^{*}v, \mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^{*}v\rangle\right)^{\frac{1}{2}}\|v\|_S. \qquad \text{(xx)}$$

Bounding the sums on the right-hand side of (xiv) using (xv) and (xx) yields

$$\|v\|_{\mathcal{A}}^2 \leq \tilde{K}\left(\sum_{j=0}^{J}\langle \mathcal{A}E_{j-1}^{*}v, \mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^{*}v\rangle\right)^{\frac{1}{2}}\|v\|_S$$

where

$$\tilde{K} := \max\left\{\sqrt{C_{\mathcal{B}}},\sqrt{C_{\mathcal{B}_0}}\right\} + \frac{\Gamma}{c_{\mathcal{A}}}\max\left\{\sqrt{\omega},\sqrt{\omega_0}\right\}.$$

Finally, using the assumption **A3** and subsequently the coercivity of $\mathcal{A}$ to bound the norm $\|v\|_S$ we have

$$\|v\|_{\mathcal{A}}^2 \leq \frac{\sqrt{C_S}}{\sqrt{c_{\mathcal{A}}}}\tilde{K}\left(\sum_{j=0}^{J}\langle \mathcal{A}E_{j-1}^{*}v, \mathcal{B}_j^{-1}\mathcal{A}E_{j-1}^{*}v\rangle\right)^{\frac{1}{2}}\|v\|_{\mathcal{A}}$$

and consequently

$$\|v\|_{\mathcal{A}}^2 \leq \frac{C_S}{c_{\mathcal{A}}} \tilde{K}^2 \sum_{j=0}^{J} \langle \mathcal{A} E_{j-1}^* v, \mathcal{B}_j^{-1} \mathcal{A} E_{j-1}^* v \rangle,$$

which finishes the proof. □

Bibliography

[1] R. A. Adams and J. J. Fournier. *Sobolev spaces*, volume 140. Elsevier, 2003.

[2] R. Becker, C. Johnson, and R. Rannacher. Adaptive error control for multigrid finite element. *Computing*, 55:271–288, 1995.

[3] F. Bornemann and H. Yserentant. A basic norm equivalence for the theory of multilevel methods. *Numerische Mathematik*, 64(1):455–476, 1993.

[4] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*, volume 15 of *Texts in Applied Mathematics*. Springer-Verlag, New York, third edition, 2007.

[5] H. Brezis. *Functional analysis, Sobolev spaces and Partial Differential Equations*. Universitext, Springer, New York, 2011.

[6] W. L. Briggs, V. E. Henson, and S. F. McCormick. *A Multigrid Tutorial*. SIAM, Philadelphia, PA, second edition, 2000.

[7] P. G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, New York, Oxford, 1978.

[8] J. B. Conway. *A course in functional analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1990.

[9] H. C. Elman, D. J. Silvester, and A. J. Wathen. *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*. Numerical mathematics and scientific computation. Oxford University Press, Oxford, 2005.

[10] L. C. Evans. *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society, 1998.

[11] M. S. Gockenbach. *Understanding and implementing the finite element method*. SIAM, Philadelphia, PA, 2006.

[12] M. Griebel and P. Oswald. On the abstract theory of additive and multiplicative Schwarz algorithms. *Numerische Mathematik*, 70(2):163–180, 1995.

[13] P. Grisvard. *Elliptic problems in nonsmooth domains*, volume 69. SIAM, Philadelphia, PA, 2011.

[14] W. Hackbusch. *Iterative solution of large sparse systems of equations*, volume 95 of *Applied Mathematical Sciences*. Springer, second edition, 2016.

[15] J. Hrnčíř, I. Pultarová, and Z. Strakoš. Decomposition into subspaces preconditioning: abstract framework. *Numerical Algorithms*, pages 1–42, 2019.

[16] M. Huber. Massively parallel and fault-tolerant multigrid solvers on petascale systems. Dissertation Technische Universität München, 2018.

[17] J. Liesen and Z. Strakoš. *Krylov Subspace Methods: Principles and Analysis*. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford, 2013.

[18] J. Málek and Z. Strakoš. *Preconditioning and the conjugate gradient method in the context of solving PDEs*, volume 1 of *SIAM Spotlights*. SIAM, Philadelphia, PA, 2015.

[19] J. Nečas. *Direct methods in the theory of elliptic equations*. Springer monographs in mathematics. Springer, Heidelberg, 2012.

[20] P. Oswald. Stable subspace splittings for sobolev spaces and domain decomposition algorithms. *Contemporary Mathematics*, 180:87–87, 1994.

[21] U. Rüde. *Mathematical and computational techniques for multilevel adaptive methods*. SIAM, Philadelphia, PA, 1993.

[22] G. W. Stewart and J. Sun. *Matrix Perturbation Theory*. Computer Science and Scientific Computing, Academic Press, 1990.

[23] P. Vacek. Multilevel methods and adaptivity. Bachelor thesis, Charles University, 2016.

[24] R. S. Varga. *Matrix iterative analysis*, volume 27 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, expanded edition, 2000.

[25] R. Verfürth. *A Posteriori Error Estimation Techniques for Finite Element Methods*. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford, 2013.

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

[27] H. Yserentant. Old and new convergence proofs for multigrid methods. *Acta Numerica*, 2:285–326, 1993.