Introduction to Multigrid Methods for Elliptic Boundary Value Problems

Arnold Reusken

Institut für Geometrie und Praktische Mathematik
RWTH Aachen, D-52056 Aachen, Germany
E-mail: reusken@igpm.rwth-aachen.de.

We treat multigrid methods for the efficient iterative solution of discretized elliptic boundary value problems. Two model problems are the Poisson equation and the Stokes problem. For the discretization we use standard finite element spaces. After discretization one obtains a large sparse linear system of equations. We explain multigrid methods for the solution of these linear systems. The basic concepts underlying multigrid solvers are discussed. Results of numerical experiments are presented which demonstrate the efficiency of these method. Theoretical convergence analyses are given that prove the typical grid independent convergence of multigrid methods.

1 Introduction

In these lecture notes we treat multigrid methods (MGM) for solving discrete elliptic boundary value problems. We assume that the reader is familiar with discretization methods for such partial differential equations. In our presentation we apply on finite element discretizations. We consider the following two model problems. Firstly, the Poisson equation

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega \subset \mathbb{R}^d, \\ u &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{1}$$

with f a (sufficiently smooth) source term and $d = 2$ or 3. The unknown is a scalar function u (for example, a temperature distribution) on Ω. We assume that the domain Ω is open, bounded and connected. The second problem consists of the Stokes equations

$$\begin{aligned} -\Delta \mathbf{u} + \nabla p &= \mathbf{f} \quad \text{in } \Omega \subset \mathbb{R}^d, \\ \operatorname{div} \mathbf{u} &= 0 \quad \text{in } \Omega, \\ \mathbf{u} &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{2}$$

The unknowns are the velocity vector function $\mathbf{u} = (u_1, \ldots, u_d)$ and the scalar pressure function p. To make this problem well-posed one needs an additional condition on p, for example, $\int_\Omega p\, dx = 0$. Both problems belong to the class of *elliptic boundary value problems*. Discretization of such partial differential equations using a finite difference, finite volume or finite element technique results in a *large sparse linear system of equations*. In the past three decades the development of *efficient iterative solvers* for such systems of equations has been an important research topic in numerical analysis and computational engineering. Nowadays it is recognized that multigrid iterative solvers are highly efficient for this type of problems and often have "optimal" complexity. There is an extensive literature on this subject. For a thorough treatment of multigrid methods we refer to the monograph

of Hackbusch[1]. For an introduction to multigrid methods requiring less knowledge of mathematics, we refer to Wesseling[2], Briggs[3], Trottenberg et al.[4]. A theoretical analysis of multigrid methods is presented in Bramble[5]. In these lecture notes we restrict ourselves to an introduction to the multigrid concept. We discuss several multigrid methods, heuristic concepts and theoretical analyses concerning convergence properties.

In the field of iterative solvers for discretized partial differential equations one can distinguish several classes of methods, namely *basic iterative methods* (eg., Jacobi, Gauss-Seidel), *Krylov subspace methods* (eg., CG, GMRES, BiCGSTAB) and *multigrid solvers*. For solving a linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ which results from the discretization of an elliptic boundary value problem the first two classes need as input (only) the matrix $\mathbf{A}$ and the righthand side $\mathbf{b}$. The fact that these data correspond to a certain underlying continuous boundary value problem is *not* used in the iterative method. However, the relation between the data ($\mathbf{A}$ and $\mathbf{b}$) and the underlying problem can be useful for the development of a fast iterative solver. Due to the fact that $\mathbf{A}$ results from a discretization procedure we know, for example, that there are other matrices which, in a certain natural sense, are similar to the matrix $\mathbf{A}$. These matrices result from the discretization of the underlying continuous boundary value problem on other grids than the grid corresponding to the given discrete problem $\mathbf{A}\mathbf{x} = \mathbf{b}$. *The use of discretizations of the given continuous problem on several grids with different mesh sizes plays an important role in the multigrid concept*. Due to the fact that in multigrid methods discrete problems on different grids are needed, the implementation of multigrid methods is in general (much) more involved than the implementation of, for example, Krylov subspace methods. We also note that for multigrid methods it is relatively hard to develop "black box" solvers which are applicable to a wide class of problems. In recent years so-called *algebraic multigrid methods* have become quite popular. In these methods one tries to reduce the amount of geometric information (eg., different grids) that is needed in the solver, thus making the multigrid method more algebraic. We will not discuss such algebraic MGM in these lecture notes.

We briefly outline the contents. In section 2 we explain the main ideas of the MGM using a simple one dimensional problem. In section 3 we introduce multigrid methods for discretizations of *scalar* elliptic boundary value problems like the Poisson equation (1). In section 4 we present results of a numerical experiment with a standard multigrid solver applied to a discrete Poisson equation in 3D. In section 5 we introduce the main ideas for a multigrid method applied to a (generalized) Stokes problem. In section 6 we present results of a numerical experiments with a Stokes equation. In the final part of these notes, the sections 7 and 8, we present convergence analyses of these multigrid methods for the two classes of elliptic boundary value problems.

2 Multigrid for a one-dimensional model problem

In this section we consider a simple model situation to show the basic principle behind the multigrid approach. We consider the two-point boundary value model problem

$$\begin{cases} -u''(x) = f(x), & x \in \Omega := (0,1). \\ u(0) = u(1) = 0\ . \end{cases} \tag{3}$$

We will use a finite element method for the discretization of this problem. This, however, is *not* essential: other discretization methods (finite differences, finite volumes) result in dis-

crete problems that are very similar. The corresponding multigrid methods have properties very similar to those in the case of a finite element discretization.

For the finite element discretization one needs a variational formulation of the boundary value problem in a suitable function space. We do not treat this issue here, but refer to the literature for information on this subject, eg. Hackbusch[6], Großmann[7]. For the two-point boundary value problem given above the appropriate function space is the Sobolov space $H_0^1(\Omega) := \{\, v \in L^2(\Omega) \mid v' \in L^2(\Omega),\ \ v(0) = v(1) = 0 \,\}$, where v' denotes a *weak* derivative of v. The variational formulation of the problem (3) is: find $u \in H_0^1(\Omega)$ such that

$$\int_0^1 u'v'\,dx = \int_0^1 fv\,dx \quad \text{for all } \ v \in H_0^1(\Omega).$$

For the discretization we introduce a sequence of nested uniform grids. For $\ell = 0, 1, 2, \ldots,$ we define

$$h_\ell = 2^{-\ell-1} \qquad \text{("mesh size")}\,, \tag{4}$$

$$n_\ell = h_\ell^{-1} - 1 \qquad \text{("number of interior grid points")}\,, \tag{5}$$

$$\xi_{\ell,i} = ih_\ell\,,\ \ i = 0, 1, ..., n_\ell + 1 \qquad \text{("grid points")}\,, \tag{6}$$

$$\Omega_\ell^{\rm int} = \{\xi_{\ell,i} \mid 1 \le i \le n_\ell\} \qquad \text{("interior grid")}\,, \tag{7}$$

$$\mathcal{T}_{h_\ell} = \cup\{\, [\xi_{\ell,i}, \xi_{\ell,i+1}] \mid 0 \le i \le n_\ell \,\} \qquad \text{("triangulation")}\,. \tag{8}$$

The space of *linear finite elements* corresponding to the triangulation $\mathcal{T}_{h_\ell}$ is given by

$$V_\ell := \{\, v \in C(\Omega) \mid v_{|[\xi_{\ell,i},\xi_{\ell,i+1}]} \in \mathcal{P}_1\,,\ i = 0, \ldots, n_\ell,\ \ v(0) = v(1) = 0 \,\}.$$

The standard nodal basis in this space is denoted by $(\phi_i)_{1\le i\le n_\ell}$. These functions satisfy $\phi_i(\xi_{\ell,i}) = 1,\ \phi_i(\xi_{\ell,j}) = 0$ for all $j \ne i$. This basis induces an isomorphism

$$P_\ell : \mathbb{R}^{n_\ell} \to V_\ell\,, \quad P_\ell \mathbf{x} = \sum_{i=1}^{n_\ell} x_i \phi_i. \tag{9}$$

The Galerkin discretization in the space V_ℓ is as follows: determine $u_\ell \in V_\ell$ such that

$$\int_0^1 u_\ell' v_\ell'\,dx = \int_0^1 fv_\ell\,dx \quad \text{for all } \ v_\ell \in V_\ell.$$

Using the representation $u_\ell = \sum_{j=1}^{n_\ell} x_j \phi_j$ this yields a linear system

$$\mathbf{A}_\ell \mathbf{x}_\ell = \mathbf{b}_\ell\,, \quad (A_\ell)_{ij} = \int_0^1 \phi_i'\phi_j'\,dx, \quad (b_\ell)_i = \int_0^1 f\phi_i\,dx. \tag{10}$$

The solution of this discrete problem is denoted by $\mathbf{x}_\ell^*$. The solution of the Galerkin discretization in the function space V_ℓ is given by $u_\ell = P_\ell \mathbf{x}_\ell^*$. A simple computation shows that

$$\mathbf{A}_\ell = h_\ell^{-1}\mathrm{tridiag}(-1, 2, -1) \in \mathbb{R}^{n_\ell \times n_\ell}.$$

Note that, apart from a scaling factor, the same matrix results from a standard discretization with finite differences of the problem (3).

Clearly, in practice one should not solve the problem in (10) using an iterative method (a Cholesky factorization $\mathbf{A} = \mathbf{L}\mathbf{L}^T$ is stable and efficient). However, we do apply a basic

iterative method here, to illustrate a certain "smoothing" property which plays an important role in multigrid methods. We consider the damped Jacobi method

$$\mathbf{x}_\ell^{k+1} = \mathbf{x}_\ell^k - \frac{1}{2}\omega h_\ell(\mathbf{A}_\ell \mathbf{x}_\ell^k - \mathbf{b}_\ell) \quad \text{with} \quad \omega \in (0,1]\,. \tag{11}$$

The iteration matrix of this method, which describes the error propagation $\mathbf{e}_\ell^{k+1} = \mathbf{C}_\ell \mathbf{e}_\ell^k$, $\mathbf{e}_\ell^k := \mathbf{x}_\ell^* - \mathbf{x}_\ell^k$, is given by

$$\mathbf{C}_\ell = \mathbf{C}_\ell(\omega) = \mathbf{I} - \frac{1}{2}\omega h_\ell \mathbf{A}_\ell\,.$$

In this simple model problem an orthogonal eigenvector basis of $\mathbf{A}_\ell$, and thus of $\mathbf{C}_\ell$, too, is known. This basis is closely related to the "Fourier modes":

$$w^\nu(x) = \sin(\nu\pi x), \quad x \in [0,1], \quad \nu = 1, 2, \ldots\,.$$

Note that w^ν satisfies the boundary conditions in (3) and that $-(w^\nu)''(x) = (\nu\pi)^2 w^\nu(x)$ holds, and thus w^ν is an eigenfunction of the problem in (3). We introduce vectors $\mathbf{z}_\ell^\nu \in \mathbb{R}^{n_\ell}$, $1 \le \nu \le n_\ell$, which correspond to the Fourier modes w^ν restricted to the interior grid $\Omega_\ell^{\rm int}$:

$$\mathbf{z}_\ell^\nu := \left(w^\nu(\xi_{\ell,1}), w^\nu(\xi_{\ell,2}), \ldots, w^\nu(\xi_{\ell,n_\ell})\right)^T\,.$$

These vectors form an orthogonal basis of $\mathbb{R}^{n_\ell}$. For $\ell = 2$ we give an illustration in Fig. 1.

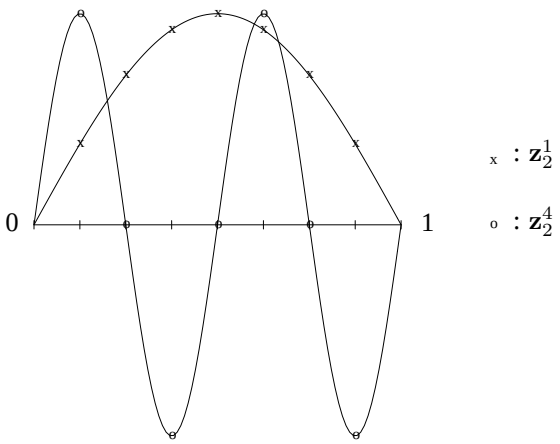


Figure 1. Two discrete Fourier modes.

To a vector $\mathbf{z}_\ell^\nu$ there corresponds a frequency ν. For $\nu < \frac{1}{2}n_\ell$ the vector $\mathbf{z}_\ell^\nu$, or the corresponding finite element function $P_\ell \mathbf{z}_\ell^\nu$, is called a "*low frequency mode*", and for $\nu \ge \frac{1}{2}n_\ell$ this vector [finite element function] is called a "*high frequency mode*". The vectors $\mathbf{z}_\ell^\nu$ are eigenvectors of the matrix $\mathbf{A}_\ell$:

$$\mathbf{A}_\ell \mathbf{z}_\ell^\nu = \frac{4}{h_\ell}\sin^2(\nu\frac{\pi}{2}h_\ell)\mathbf{z}_\ell^\nu,$$

and thus we have

$$\mathbf{C}_\ell \mathbf{z}_\ell^\nu = (1 - 2\omega \sin^2(\nu \frac{\pi}{2} h_\ell)) \mathbf{z}_\ell^\nu \ . \tag{12}$$

From this we obtain

$$\begin{aligned} \|\mathbf{C}_\ell\|_2 &= \max_{1 \le \nu \le n_\ell} \ |1 - 2\omega \sin^2(\nu \tfrac{\pi}{2} h_\ell)| \\ &= 1 - 2\omega \sin^2(\tfrac{\pi}{2} h_\ell) = 1 - \tfrac{1}{2}\omega \pi^2 h_\ell^2 + \mathcal{O}(h_\ell^4) \ . \end{aligned} \tag{13}$$

From this we see that the damped Jacobi method is convergent ($\|\mathbf{C}_\ell\|_2 < 1$), but that the rate of convergence will be very low for h_ℓ small.

Note that the eigenvalues and the eigenvectors of $\mathbf{C}_\ell$ are functions of $\nu h_\ell \in [0, 1]$:

$$\lambda_{\ell,\nu} := 1 - 2\omega \sin^2(\nu \frac{\pi}{2} h_\ell) =: g_\omega(\nu h_\ell) \ , \text{ with} \tag{14a}$$

$$g_\omega(y) \ = \ 1 - 2\omega \sin^2(\frac{\pi}{2} y), \qquad y \in [0, 1]. \tag{14b}$$

Hence, the size of the eigenvalues $\lambda_{\ell,\nu}$ can directly be obtained from the graph of the function g_ω. In Fig. 2 we show the graph of the function g_ω for a few values of ω.

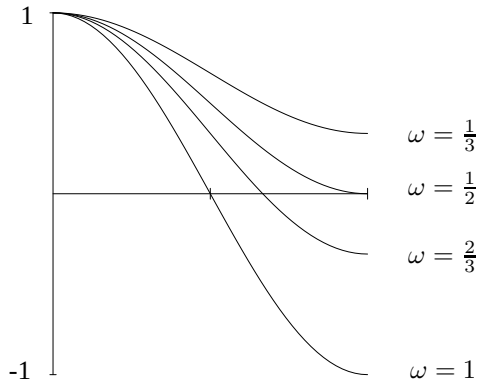


Figure 2. Graph of g_ω.

From the graphs in this figure we conclude that for a suitable choice of ω we have $|g_\omega(y)| \ll 1$ if $y \in [\frac{1}{2}, 1]$. We choose $\omega = \frac{2}{3}$ (then $|g_\omega(\frac{1}{2})| = |g_\omega(1)|$ holds). Then we have $|g_{\frac{2}{3}}(y)| \le \frac{1}{3}$ for $y \in [\frac{1}{2}, 1]$. Using this and the result in (14a) we obtain

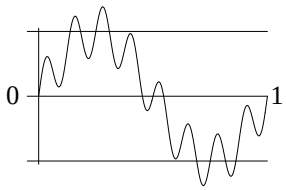
$$|\lambda_{\ell,\nu}| \le \frac{1}{3} \quad \text{for } \nu \ge \frac{1}{2} n_\ell \ .$$

Hence:

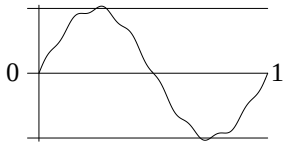
the high frequency modes are strongly damped by the iteration matrix $\mathbf{C}_\ell$.

From Fig. 2 it is also clear that the low rate of convergence of the damped Jacobi method is caused by the low frequency modes ($\nu h_\ell \ll 1$).

Summarizing, we draw the conclusion that in this example the damped Jacobi method will "smooth" the error. This elementary observation is of great importance for the two-grid method introduced below. In the setting of multigrid methods the damped Jacobi method is called a "smoother". The smoothing property of damped Jacobi is illustrated in Fig. 3. It is important to note that the discussion above concerning smoothing is related to



Graph of a starting error.



Graph of the error after one damped Jacobi iteration ($\omega = \frac{2}{3}$).

Figure 3. Smoothing property of damped Jacobi.

the iteration matrix $\mathbf{C}_\ell$, which means that the *error* will be made smoother by the damped Jacobi method, but not (necessarily) the new iterand $\mathbf{x}^{k+1}$.

In multigrid methods we have to transform information from one grid to another. For that purpose we introduce so-called *prolongations* and *restrictions*. In a setting with nested finite element spaces these operators can be defined in a very natural way. Due to the nestedness the identity operator

$$I_\ell : V_{\ell-1}, \rightarrow V_\ell, \quad I_\ell v = v,$$

is well-defined. This identity operator represents linear interpolation as is illustrated for $\ell = 2$ in Fig. 4. The matrix representation of this interpolation operator is given by

$$\mathbf{p}_\ell : \mathbb{R}^{n_{\ell-1}} \rightarrow \mathbb{R}^{n_\ell}, \quad \mathbf{p}_\ell := P_\ell^{-1} P_{\ell-1}. \tag{15}$$

A simple computation yields

$$\mathbf{p}_\ell = \begin{bmatrix} \frac{1}{2} & & \emptyset \\ 1 & & \\ \frac{1}{2} & \frac{1}{2} & \\ & 1 & \\ & \frac{1}{2} & \\ & \ddots & \\ & & \frac{1}{2} \\ & & 1 \\ \emptyset & & \frac{1}{2} \end{bmatrix}_{n_\ell \times n_{\ell-1}} . \tag{16}$$

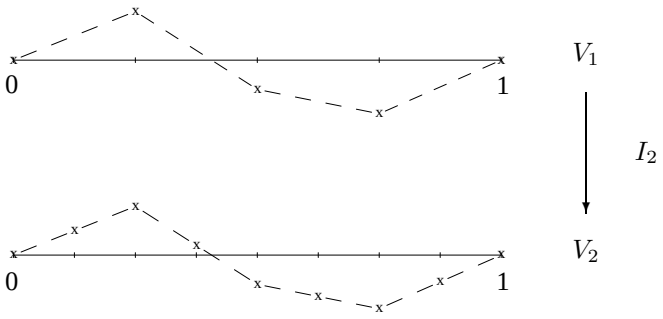


Figure 4. Canonical prolongation.

We can also restrict a given grid function v_ℓ on $\Omega_\ell^{\rm int}$ to a grid function on $\Omega_{\ell-1}^{\rm int}$. An obvious approach is to use a restriction r based on simple injection:

$$(r_{\rm inj} v_\ell)(\xi) = v_\ell(\xi) \quad \text{if} \ \ \xi \in \Omega_{\ell-1}^{\rm int} \ .$$

When used in a multigrid method then often this restriction based on injection is not satisfactory (cf. Hackbusch[1], section 3.5). A better method is obtained if a natural Galerkin property is satisfied. It can easily be verified (cf. also lemma 3.2) that with $\mathbf{A}_\ell, \ \mathbf{A}_{\ell-1}$ and $\mathbf{p}_\ell$ as defined in (10), (15) we have

$$\mathbf{r}_\ell \mathbf{A}_\ell \mathbf{p}_\ell = \mathbf{A}_{\ell-1} \quad \text{iff} \quad \mathbf{r}_\ell = \mathbf{p}_\ell^T . \tag{17}$$

Thus the natural Galerkin condition $\mathbf{r}_\ell \mathbf{A}_\ell \mathbf{p}_\ell = \mathbf{A}_{\ell-1}$ implies the choice

$$\mathbf{r}_\ell = \mathbf{p}_\ell^T \tag{18}$$

for the restriction operator.

The *two-grid* method is based on the idea that a smooth error, which results from the application of one or a few damped Jacobi iterations, can be approximated fairly well on a *coarser* grid. We now introduce this two-grid method.

Consider $\mathbf{A}_\ell \mathbf{x}_\ell^* = \mathbf{b}_\ell$ and let $\overline{\mathbf{x}}_\ell$ be the result of one or a few damped Jacobi iterations applied to a given starting vector $\mathbf{x}_\ell^0$. For the error $\mathbf{e}_\ell := \mathbf{x}_\ell^* - \overline{\mathbf{x}}_\ell$ we have

$$\mathbf{A}_\ell \mathbf{e}_\ell = \mathbf{b}_\ell - \mathbf{A}_\ell \overline{\mathbf{x}}_\ell =: \mathbf{d}_\ell \quad (\text{ “residual” or “defect”}). \tag{19}$$

Based on the assumption that $\mathbf{e}_\ell$ is smooth it seems reasonable to make the approximation $\mathbf{e}_\ell \approx \mathbf{p}_\ell \tilde{\mathbf{e}}_{\ell-1}$ with an appropriate vector (grid function) $\tilde{\mathbf{e}}_{\ell-1} \in \mathbb{R}^{n_{\ell-1}}$. To determine the vector $\tilde{\mathbf{e}}_{\ell-1}$ we use the equation (19) and the Galerkin property (17). This results in the equation

$$\mathbf{A}_{\ell-1} \tilde{\mathbf{e}}_{\ell-1} = \mathbf{r}_\ell \mathbf{d}_\ell$$

for the vector $\tilde{\mathbf{e}}_{\ell-1}$. Note that $\mathbf{x}^* = \overline{\mathbf{x}}_\ell + \mathbf{e}_\ell \approx \overline{\mathbf{x}}_\ell + \mathbf{p}_\ell \tilde{\mathbf{e}}_{\ell-1}$. Thus for the new iterand we take $\mathbf{x}_\ell := \overline{\mathbf{x}}_\ell + \mathbf{p}_\ell \tilde{\mathbf{e}}_{\ell-1}$. In a more compact formulation this two-grid method is as

follows:

$$
\left\{
\begin{array}{l}
\mathit{procedure}\ \mathrm{TGM}_\ell(\mathbf{x}_\ell, \mathbf{b}_\ell) \\
\quad \text{if } \ell = 0 \text{ then } \mathbf{x}_0 := \mathbf{A}_0^{-1}\mathbf{b}_0 \text{ else} \\
\quad \text{begin} \\
\qquad \mathbf{x}_\ell := J_\ell^\nu(\mathbf{x}_\ell, \mathbf{b}_\ell) \ \ (* \ \nu \text{ smoothing it., e.g. damped Jacobi } *) \\
\qquad \mathbf{d}_{\ell-1} := \mathbf{r}_\ell(\mathbf{b}_\ell - \mathbf{A}_\ell \mathbf{x}_\ell) \quad (* \text{ restriction of defect } *) \\
\qquad \tilde{\mathbf{e}}_{\ell-1} := \mathbf{A}_{\ell-1}^{-1}\mathbf{d}_{\ell-1} \quad (* \text{ solve coarse grid problem } *) \\
\qquad \mathbf{x}_\ell := \mathbf{x}_\ell + \mathbf{p}_\ell \tilde{\mathbf{e}}_{\ell-1} \quad (* \text{ add correction } *) \\
\qquad \mathrm{TGM}_\ell := \mathbf{x}_\ell \\
\quad \text{end;}
\end{array}
\right. \tag{20}
$$

Often, after the coarse grid correction $\mathbf{x}_\ell := \mathbf{x}_\ell + \mathbf{p}_\ell \tilde{\mathbf{e}}_{\ell-1}$, one or a few smoothing iterations are applied. Smoothing before/after the coarse grid correction is called pre/post-smoothing. Besides the smoothing property a second property which is of great importance for a multigrid method is the following:

The coarse grid system $\mathbf{A}_{\ell-1}\tilde{\mathbf{e}}_{\ell-1} = \mathbf{d}_{\ell-1}$ is of the same form as the system $\mathbf{A}_\ell \mathbf{x}_\ell = \mathbf{b}_\ell$.

Thus for solving the problem $\mathbf{A}_{\ell-1}\tilde{\mathbf{e}}_{\ell-1} = \mathbf{d}_{\ell-1}$ *approximately* we can apply the two-grid algorithm in (20) recursively. This results in the following *multigrid method* for solving $\mathbf{A}_\ell \mathbf{x}_\ell^* = \mathbf{b}_\ell$:

$$
\left\{
\begin{array}{l}
\mathit{procedure}\ \mathrm{MGM}_\ell(\mathbf{x}_\ell, \mathbf{b}_\ell) \\
\quad \text{if } \ \ell = 0 \text{ then } \mathbf{x}_0 := \mathbf{A}_0^{-1}\mathbf{b}_0 \text{ else} \\
\quad \text{begin} \\
\qquad \mathbf{x}_\ell := J_\ell^{\nu_1}(\mathbf{x}_\ell, \mathbf{b}_\ell) \quad (* \text{ presmoothing } *) \\
\qquad \mathbf{d}_{\ell-1} := \mathbf{r}_\ell(\mathbf{b}_\ell - \mathbf{A}_\ell \mathbf{x}_\ell) \\
\qquad \mathbf{e}_{\ell-1}^0 := \mathbf{0}; \ \text{ for } i = 1 \text{ to } \tau \text{ do } \ \mathbf{e}_{\ell-1}^i := \mathrm{MGM}_{\ell-1}(\mathbf{e}_{\ell-1}^{i-1}, \mathbf{d}_{\ell-1}); \\
\qquad \mathbf{x}_\ell := \mathbf{x}_\ell + \mathbf{p}_\ell \mathbf{e}_{\ell-1}^\tau \\
\qquad \mathbf{x}_\ell := J_\ell^{\nu_2}(\mathbf{x}_\ell, \mathbf{b}_\ell) \quad (* \text{ postsmoothing } *) \\
\qquad \mathrm{MGM}_\ell := \mathbf{x}_\ell \\
\quad \text{end;}
\end{array}
\right. \tag{21}
$$

If one wants to solve the system on a given finest grid, say with level number $\overline{\ell}$, i.e. $\mathbf{A}_{\overline{\ell}}\mathbf{x}_{\overline{\ell}}^* = \mathbf{b}_{\overline{\ell}}$, then we apply some iterations of $\mathrm{MGM}_{\overline{\ell}}(\mathbf{x}_{\overline{\ell}}, \mathbf{b}_{\overline{\ell}})$.

Based on efficiency considerations (cf. section 3) we usually take $\tau = 1$ ("V-cycle") or $\tau = 2$ ("W-cycle") in the recursive call in (21). For the case $\ell = 3$ the structure of one multigrid iteration with $\tau \in \{1, 2\}$ is illustrated in Fig. 5.

3 Multigrid for scalar elliptic problems

In this section we introduce multigrid methods which can be used for solving discretized scalar elliptic boundary value problems. A model example from this problem class is the Poisson equation in (1). Opposite to the CG method, the applicability of multigrid methods

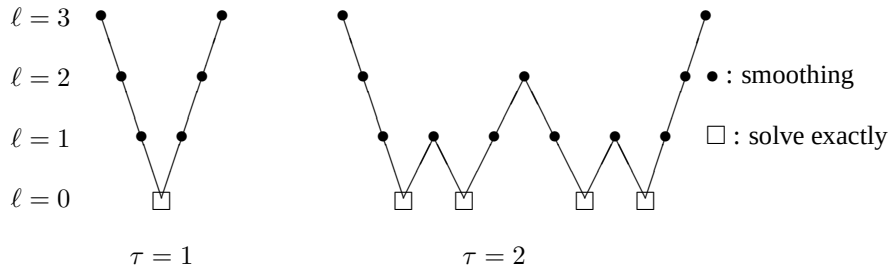


Figure 5. Structure of one multigrid iteration

is not restricted to symmetric problems. Multigrid methods can also be used for solving problems which are nonsymmetric (i.e., convection terms are present in the equation). If the problem is convection-dominated (the corresponding stiffness matrix then is strongly nonsymmetic) one usually has to modify the standard multigrid approach in the sense that special smoothers and/or special prolongations and restrictions should be used. We do not discuss this issue here.

We will introduce the two-grid and multigrid method by generalizing the approach of section 2 to the higher (i.e., two and three) dimensional case. We consider a scalar elliptic boundary value problems of the form

$$\begin{aligned} -\nabla \cdot (a\nabla u) + \mathbf{b} \cdot \nabla u + cu &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned}$$

This problem is considered in a domain $\Omega \subset \mathbb{R}^d$, $d = 2$ or 3. We assume that the functions a, c and the vector function $\mathbf{b}$ are sufficiently smooth on Ω and

$$a(x) \geq a_0 > 0, \;\; c(x) - \frac{1}{2}\mathrm{div}\, \mathbf{b}(x) \geq 0 \quad \text{for all } x \in \bar{\Omega}. \tag{22}$$

These assumptions guarantee that the problem is elliptic and well-posed. In view of the finite element discretization we introduce the variational formulation of this problem. For this we need the Sobolov space $H_0^1(\Omega) := \{ v \in L^2(\Omega) \mid \frac{\partial v}{\partial x_i} \in L^2(\Omega),\ i = 1, \ldots, d, \;\; v_{|\partial\Omega} = 0 \}$. The partial derivative $\frac{\partial v}{\partial x_i}$ has to be interpreted in a suitable weak sense. The variational formulation is as follows:

$$\begin{cases} \text{find } u \in H_0^1(\Omega) \text{ such that} \\ k(u, v) = f(v) \quad \text{for all } v \in H_0^1(\Omega), \end{cases} \tag{23}$$

with a bilinear form and righthand side

$$k(u, v) = \int_\Omega a\nabla u^T \nabla v + \mathbf{b} \cdot \nabla u v + cuv \, dx. \,, \quad f(v) = \int_\Omega fv \, dx.$$

If (22) holds then this bilinear form is *continuous and elliptic* on $H_0^1(\Omega)$, i.e. there exist constants $\gamma > 0$ and c such that

$$k(u, u) \geq \gamma |u|_1^2, \quad k(u, v) \leq c|u|_1 |v|_1 \quad \text{for all } u, v \in H_0^1(\Omega).$$

Here we use $|u|_1 := \left(\int_\Omega \nabla u^T \nabla u \, dx \right)^{\frac{1}{2}}$, which is a norm on $H_0^1(\Omega)$. For the discretization of this problem we use simplicial finite elements. Let $\{\mathcal{T}_h\}$ be a regular family of triangulations of Ω consisting of d-simplices and V_h a corresponding finite element space. For simplicity we only consider *linear* finite elements:

$$V_h = \{ \, v \in C(\Omega) \mid v_{|T} \in \mathcal{P}_1 \ \text{ for all } \ T \in \mathcal{T}_h \, \}.$$

The presentation and implementation of the multigrid method is greatly simplified if we assume a given sequence of *nested* finite element spaces.

Assumption 3.1 *In the remainder we always assume that we have a sequence V_ℓ, $\ell = 0, 1, \ldots$, of simplicial finite element spaces which are nested:*

$$V_\ell \subset V_{\ell+1} \quad \textit{for all } \ \ell. \tag{24}$$

We note that this assumption is not necessary for a succesful application of multigrid methods. For a treatment of multigrid methods in case of non-nestedness we refer to Trottenberg et al.[4]. The construction of a hierarchy of triangulations such that the corresponding finite element spaces are nested is discussed in Bey[8].

In V_ℓ we use the standard nodal basis $(\phi_i)_{1 \le i \le n_\ell}$. This basis induces an isomorphism

$$P_\ell : \mathbb{R}^{n_\ell} \to V_\ell \ , \quad P_\ell \mathbf{x} = \sum_{i=1}^{n_\ell} x_i \phi_i.$$

The Galerkin discretization: Find $u_\ell \in V_\ell$ such that

$$k(u_\ell, v_\ell) = f(v_\ell) \quad \text{for all } \ v_\ell \in V_\ell \tag{25}$$

can be represented as a linear system

$$\mathbf{A}_\ell \mathbf{x}_\ell = \mathbf{b}_\ell \, , \ \text{with } \ (A_\ell)_{ij} = k(\phi_j, \phi_i), \ (b_\ell)_i = f(\phi_i), \ \ 1 \le i, j \le n_\ell. \tag{26}$$

The solution $\mathbf{x}_\ell^*$ of this linear system yields the Galerkin finite element solution $u_\ell = P_\ell \mathbf{x}_\ell^*$. Along the same lines as in the one-dimensional case we introduce a multigrid method for solving this system of equations on an arbitrary level $\ell \ge 0$.

For the *smoother* we use a basic iterative method such as, for example, a *Richardson method*

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \omega_\ell (\mathbf{A}_\ell \mathbf{x}^k - \mathbf{b}),$$

a *damped Jacobi method*

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \omega \mathbf{D}_\ell^{-1} (\mathbf{A}_\ell \mathbf{x}^k - \mathbf{b}), \tag{27}$$

or a *Gauss-Seidel method*

$$\mathbf{x}^{k+1} = \mathbf{x}^k - (\mathbf{D}_\ell - \mathbf{L}_\ell)^{-1} (\mathbf{A}_\ell \mathbf{x}^k - \mathbf{b}), \tag{28}$$

where $\mathbf{D}_\ell - \mathbf{L}_\ell$ is the lower triangular part of the matrix $\mathbf{A}_\ell$. For such a method we use the general notation

$$\mathbf{x}^{k+1} = \mathcal{S}_\ell(\mathbf{x}^k, \mathbf{b}_\ell) = \mathbf{x}^k - \mathbf{M}_\ell^{-1} (\mathbf{A}_\ell \mathbf{x}^k - \mathbf{b}) \ , \quad k = 0, 1, \ldots$$

The corresponding iteration matrix is denoted by

$$\mathbf{S}_\ell = \mathbf{I} - \mathbf{M}_\ell^{-1} \mathbf{A}_\ell.$$

For the *prolongation* we use the matrix representation of the identity $I_\ell : V_{\ell-1} \to V_\ell$, i.e.,

$$\mathbf{p}_\ell := P_\ell^{-1} P_{\ell-1}. \tag{29}$$

The choice of the restriction is based on the following elementary lemma:

Lemma 3.2 *Let $\mathbf{A}_\ell$, $\ell \geq 0$, be the stiffness matrix defined in* (26) *and $\mathbf{p}_\ell$ as in* (29)*. Then for $\mathbf{r}_\ell : \mathbb{R}^{n_\ell} \to \mathbb{R}^{n_{\ell-1}}$ we have:*

$$\mathbf{r}_\ell \mathbf{A}_\ell \mathbf{p}_\ell = \mathbf{A}_{\ell-1} \quad \textit{if and only if} \quad \mathbf{r}_\ell = \mathbf{p}_\ell^T.$$

Proof: For the stiffness matrix matrix the identity

$$\langle \mathbf{A}_\ell \mathbf{x}, \mathbf{y} \rangle = k(P_\ell \mathbf{x}, P_\ell \mathbf{y}) \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^{n_\ell}$$

holds. From this we get

$$\begin{aligned} & \mathbf{r}_\ell \mathbf{A}_\ell \mathbf{p}_\ell = \mathbf{A}_{\ell-1} \\ \Leftrightarrow \quad & \langle \mathbf{A}_\ell \mathbf{p}_\ell \mathbf{x}, \mathbf{r}_\ell^T \mathbf{y} \rangle = \langle \mathbf{A}_{\ell-1} \mathbf{x}, \mathbf{y} \rangle \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^{n_{\ell-1}} \\ \Leftrightarrow \quad & k(P_{\ell-1}\mathbf{x}, P_\ell \mathbf{r}_\ell^T \mathbf{y}) = k(P_{\ell-1}\mathbf{x}, P_{\ell-1}\mathbf{y}) \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^{n_{\ell-1}}. \end{aligned}$$

Using the ellipticity of $k(\cdot,\cdot)$ it now follows that

$$\begin{aligned} & \mathbf{r}_\ell \mathbf{A}_\ell \mathbf{p}_\ell = \mathbf{A}_{\ell-1} \\ \Leftrightarrow \quad & P_\ell \mathbf{r}_\ell^T \mathbf{y} = P_{\ell-1}\mathbf{y} \quad \text{for all } \mathbf{y} \in \mathbb{R}^{n_{\ell-1}} \\ \Leftrightarrow \quad & \mathbf{r}_\ell^T \mathbf{y} = P_\ell^{-1} P_{\ell-1}\mathbf{y} = \mathbf{p}_\ell \mathbf{y} \quad \text{for all } \mathbf{y} \in \mathbb{R}^{n_{\ell-1}} \\ \Leftrightarrow \quad & \mathbf{r}_\ell^T = \mathbf{p}_\ell. \end{aligned}$$

Thus the claim is proved. ∎

This motivates that for the *restriction* we take:

$$\mathbf{r}_\ell := \mathbf{p}_\ell^T. \tag{30}$$

Using these components we can define a multigrid method with exactly the same structure as in (21):

```
procedure MGM_ℓ(x_ℓ, b_ℓ)
  if ℓ = 0 then x_0 := A_0^{-1} b_0 else
  begin
    x_ℓ := S_ℓ^{ν1}(x_ℓ, b_ℓ)   (* presmoothing *)
    d_{ℓ-1} := r_ℓ(b_ℓ − A_ℓ x_ℓ)
    e^0_{ℓ-1} := 0;  for i = 1 to τ do  e^i_{ℓ-1} := MGM_{ℓ-1}(e^{i-1}_{ℓ-1}, d_{ℓ-1});
    x_ℓ := x_ℓ + p_ℓ e^τ_{ℓ-1}
    x_ℓ := S_ℓ^{ν2}(x_ℓ, b_ℓ)   (* postsmoothing *)
    MGM_ℓ := x_ℓ
  end;
```

(31)

We briefly comment on some important issues related to this multigrid method.

Smoothers
For many problems basic iterative methods provide good smoothers. In particular the Gauss-Seidel method is often a very effective smoother. Other smoothers used in practice are the damped Jacobi method and the ILU method.

Prolongation and restriction
If instead of a discretization with nested finite element spaces one uses a finite difference or a finite volume method then one can not use the approach in (29) to define a prolongation. However, for these cases other canonical constructions for the prolongation operator exist. We refer to Hackbusch[1], Trottenberg et al.[4] or Wesseling[2] for a treatment of this topic. A general technique for the construction of a prolongation operator in case of nonnested finite element spaces is given in Braess[9].

Arithmetic costs per iteration
We discuss the arithmetic costs of one MGM_ℓ iteration as defined in (31). For this we introduce a unit of arithmetic work on level ℓ:

$$WU_\ell := \#\text{ flops needed for } \mathbf{A}_\ell \mathbf{x}_\ell - \mathbf{b}_\ell \text{ computation.} \tag{32}$$

We assume:

$$WU_{\ell-1} \lesssim g\, WU_\ell \quad \text{with } g < 1 \text{ independent of } \ell. \tag{33}$$

Note that if $\mathcal{T}_\ell$ is constructed through a uniform global grid refinement of $\mathcal{T}_{\ell-1}$ (for $n = 2$: subdivision of each triangle $T \in \mathcal{T}_{\ell-1}$ into four smaller triangles by connecting the midpoints of the edges) then (33) holds with $g = (\frac{1}{2})^d$. Furthermore we make the following assumptions concerning the arithmetic costs of each of the substeps in the procedure MGM_ℓ:

$$\left.\begin{aligned} \mathbf{x}_\ell &:= \mathcal{S}_\ell(\mathbf{x}_\ell, \mathbf{b}_\ell) \;: \quad \text{costs} \lesssim WU_\ell \\ \mathbf{d}_{\ell-1} &:= \mathbf{r}_\ell(\mathbf{b}_\ell - \mathbf{A}_\ell \mathbf{x}_\ell) \\ \mathbf{x}_\ell &:= \mathbf{x}_\ell + \mathbf{p}_\ell \mathbf{e}_{\ell-1}^\tau \end{aligned}\right\} \text{total costs} \lesssim 2\, WU_\ell$$

For the amount of work in one multigrid V-cycle ($\tau = 1$) on level ℓ, which is denoted by VMG_ℓ, we get using $\nu := \nu_1 + \nu_2$:

$$\begin{aligned} VMG_\ell &\lesssim \nu WU_\ell + 2WU_\ell + VMG_{\ell-1} = (\nu+2)WU_\ell + VMG_{\ell-1} \\ &\lesssim (\nu+2)\big(WU_\ell + WU_{\ell-1} + \ldots + WU_1\big) + VMG_0 \\ &\lesssim (\nu+2)\big(1 + g + \ldots + g^{\ell-1}\big)WU_\ell + VMG_0 \\ &\lesssim \frac{\nu+2}{1-g}\, WU_\ell. \end{aligned} \tag{34}$$

In the last inequality we assumed that the costs for computing $\mathbf{x}_0 = \mathbf{A}_0^{-1}\mathbf{b}_0$ (i.e., VMG_0) are negligible compared to WU_ℓ. The result in (34) shows that the arithmetic costs for one V-cycle are proportional (if $\ell \to \infty$) to the costs of a residual computation. For example, for $g = \frac{1}{8}$ (uniform refinement in 3D) the arithmetic costs of a V-cycle with $\nu_1 = \nu_2 = 1$ on level ℓ are comparable to $4\frac{1}{2}$ times the costs of a residual computation on level ℓ.

For the W-cycle ($\tau = 2$) the arithmetic costs on level ℓ are denoted by WMG_ℓ. We have:

$$\begin{aligned} WMG_\ell &\lesssim \nu WU_\ell + 2WU_\ell + 2WMG_{\ell-1} = (\nu+2)WU_\ell + 2WMG_{\ell-1} \\ &\lesssim (\nu+2)\big(WU_\ell + 2WU_{\ell-1} + 2^2 WU_{\ell-2} + \ldots + 2^{\ell-1}WU_1\big) + WMG_0 \\ &\lesssim (\nu+2)\big(1 + 2g + (2g)^2 + \ldots + (2g)^{\ell-1}\big)WU_\ell + WMG_0. \end{aligned}$$

From this we see that to obtain a bound proportional to WU_ℓ we have to assume

$$g < \frac{1}{2}.$$

Under this assumption we get for the W-cycle

$$WMG_\ell \lesssim \frac{\nu+2}{1-2g}\, WU_\ell$$

(again we neglected WMG_0). Similar bounds can be obtained for $\tau \geq 3$, provided $\tau g < 1$ holds.

3.1 Nested Iteration

We consider a sequence of discretizations of a given boundary value problem, as for example in (26):

$$\mathbf{A}_\ell \mathbf{x}_\ell = \mathbf{b}_\ell, \qquad \ell = 0, 1, 2, \ldots.$$

We assume that for a certain $\ell = \overline{\ell}$ we want to compute the solution $\mathbf{x}^*_{\overline{\ell}}$ of the problem $\mathbf{A}_{\overline{\ell}}\mathbf{x}_{\overline{\ell}} = \mathbf{b}_{\overline{\ell}}$ using an iterative method (not necessarily a multigrid method). In the nested iteration method we use the systems on coarse grids to obtain a *good starting vector* $\mathbf{x}^0_{\overline{\ell}}$ for this iterative method with relatively low computational costs. The nested iteration method for the computation of this starting vector $\mathbf{x}^0_{\overline{\ell}}$ is as follows

$$\left\{\begin{array}{l} \text{compute the solution } \mathbf{x}^*_0 \text{ of } \mathbf{A}_0\mathbf{x}_0 = \mathbf{b}_0 \\ \mathbf{x}^0_1 := \tilde{\mathbf{p}}_1 \mathbf{x}^*_0 \quad \text{(prolongation of } \mathbf{x}^*_0\text{)} \\ \mathbf{x}^k_1 := \text{result of } k \text{ iterations of an iterative method} \\ \qquad \text{applied to } \mathbf{A}_1\mathbf{x}_1 = \mathbf{b}_1 \text{ with starting vector } \mathbf{x}^0_1 \\ \mathbf{x}^0_2 := \tilde{\mathbf{p}}_2 \mathbf{x}^k_1 \quad \text{(prolongation of } \mathbf{x}^k_1\text{)} \\ \mathbf{x}^k_2 := \text{ result of } k \text{ iterations of an iterative method} \\ \qquad \text{applied to } \mathbf{A}_2\mathbf{x}_2 = \mathbf{b}_2 \text{ with starting vector } \mathbf{x}^0_2 \\ \vdots \\ \text{etc.} \\ \vdots \\ \mathbf{x}^0_{\overline{\ell}} := \tilde{\mathbf{p}}_{\overline{\ell}} \mathbf{x}^k_{\overline{\ell}-1}\,. \end{array}\right. \tag{35}$$

In this nested iteration method we use a prolongation $\tilde{\mathbf{p}}_\ell : \mathbb{R}^{n_{\ell-1}} \to \mathbb{R}^{n_\ell}$. The nested iteration principle is based on the idea that $\tilde{\mathbf{p}}_\ell \mathbf{x}^*_{\ell-1}$ is expected to be a reasonable approximation of $\mathbf{x}^*_\ell$, because $\mathbf{A}_{\ell-1}\mathbf{x}^*_{\ell-1} = \mathbf{b}_{\ell-1}$ and $\mathbf{A}_\ell \mathbf{x}^*_\ell = \mathbf{b}_\ell$ are discretizations of the same

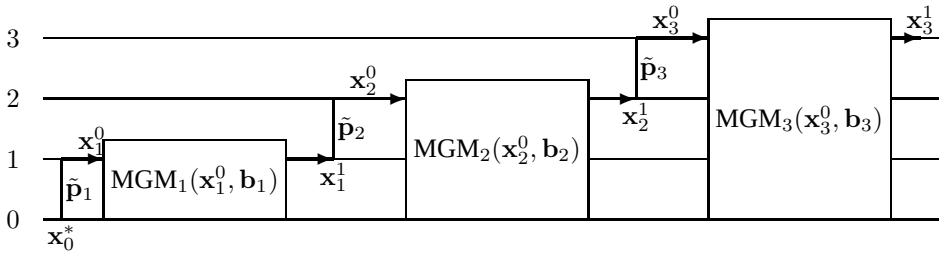


Figure 6. Multigrid and nested iteration.

continuous problem. With respect to the computational costs of this approach we note the following (cf. Hackbusch[1], section 5.3). For the nested iteration to be a feasible approach, the number of iterations applied on the coarse grids (i.e. k in (35)) should not be "too large" and the number of grid points in the union of all coarse grids (i.e. level $0, 1, 2, ..., \overline{\ell} - 1$) should be at most of the same order of magnitude as the number of grid points in the level $\overline{\ell}$ grid. Often, if one uses a multigrid solver these two conditions are satisfied. Usually in multigrid we use coarse grids such that the number of grid points decreases in a geometric fashion, and for k in (35) we can often take $k = 1$ or $k = 2$ due to the fact that on the coarse grids we use the multigrid method, which has a high rate of convergence.

If one uses the algorithm $\mathrm{MGM}_{\overline{\ell}}$ from (31) as the solver on level $\overline{\ell}$ then the implementation of the nested iteration method can be realized with only little additional effort because the coarse grid data structure and coarse grid operators (e.g. $\mathbf{A}_\ell,\ \ell < \overline{\ell}$) needed in the nested iteration method are already available.

If in the nested iteration method we use a multigrid iterative solver on all levels we obtain the following algorithmic structure:

$$\left\{\begin{array}{l} \mathbf{x}_0^* := \mathbf{A}_0^{-1}\mathbf{b}_0;\ \mathbf{x}_0^k := \mathbf{x}_0^* \\ \text{for}\ \ \ell = 1 \text{ to } \overline{\ell} \text{ do} \\ \quad \text{begin} \\ \quad\quad \mathbf{x}_\ell^0 := \tilde{\mathbf{p}}_\ell \mathbf{x}_{\ell-1}^k \\ \quad\quad \text{for}\ \ i = 1 \text{ to } k \text{ do}\ \ \mathbf{x}_\ell^i := \mathrm{MGM}_\ell(\mathbf{x}_\ell^{i-1}, \mathbf{b}_\ell) \\ \quad \text{end;} \end{array}\right. \tag{36}$$

For the case $\overline{\ell} = 3$ and $k = 1$ this method is illustrated in Fig. 6.

Remark 3.3 The prolongation $\tilde{\mathbf{p}}_\ell$ used in the nested iteration may be the same as the prolongation $\mathbf{p}_\ell$ used in the multigrid method. However, from the point of view of efficiency it is sometimes better to use in the nested iteration a prolongation $\tilde{\mathbf{p}}_\ell$ that has a higher order of accuracy than the prolongation used in the multigrid method. □

4 Numerical experiment: Multigrid applied to a Poisson equation

In this section we present results of a standard multigrid solver applied to the model problem of the Poisson equation:

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega := (0,1)^3, \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned}$$

We take $f(x_1, x_2, x_3) = x_1^2 + e^{x_2} x_1 + x_3^2 x_2$. For the discretization we start with a uniform subdivision of Ω into cubes with edges of length $h_0 := \frac{1}{4}$. Each cube is subdivided into six tetrahedra. This yields the starting triangulation $\mathcal{T}_0$ of Ω. The triangulation $\mathcal{T}_1$ with mesh size $h_1 = \frac{1}{8}$ is constructed by regular subdivision of each tetrahedron in $\mathcal{T}_0$ into 8 child tetrahedra. This uniform refinement strategy is repeated, resulting in a family of triangulations $(\mathcal{T}_\ell)_{\ell \geq 0}$ with corresponding mesh size $h_\ell = 2^{-\ell-2}$. For discretization of this problem we use the space of linear finite elements on these triangulations. The resulting linear system is denoted by $\mathbf{A}_\ell \mathbf{x}_\ell = \mathbf{b}_\ell$. We consider the problem of solving this linear system on a fixed finest level $\ell = \bar{\ell}$. Below we consider $\bar{\ell} = 1, \ldots, 5$. For $\bar{\ell} = 5$ the triangulation contains 14.380.416 tetrahedra and in the linear system we have 2.048.383 unknowns.

We briefly discuss the components used in the multigrid method for solving this linear system. For the prolongation and restriction we use the canonical ones as in (29), (30). For the smoother we use two different methods, namely a damped Jacobi method and a symmetric Gauss-Seidel method (SGS). The damped Jacobi method is as in (27) with $\omega := 0.7$. The symmetric Gauss-Seidel method consists of two substeps. In the first step we use a Gauss-Seidel iteration as in (28). In the second step we apply this method with a reversed ordering of the equations and the unknowns. The arithmetic costs per iteration for such a symmetric Gauss-Seidel smoother are roughly twice as high as for a damped Jacobi method. In the experiment we use the same number of pre- and post-smoothing iterations, i.e. $\nu_1 = \nu_2$. The total number of smoothing iterations per multigrid iteration is $\nu := \nu_1 + \nu_2$. We use a multigrid V-cycle. i.e., $\tau = 1$ in the recursive call in (31). The coarsest grid used in the multigrid method is $\mathcal{T}_0$, i.e. with a mesh size $h_0 = \frac{1}{4}$. In all experiments we use a starting vector $\mathbf{x}^0 := 0$. The rate of convergence is measured by looking at relative residuals:

$$r_k := \frac{\|\mathbf{A}_{\bar{\ell}} \mathbf{x}^k - \mathbf{b}_{\bar{\ell}}\|_2}{\|\mathbf{b}_{\bar{\ell}}\|_2}.$$

In Fig. 7 (left) we show results for SGS with $\nu = 4$. For $\bar{\ell} = 1, \ldots, 5$ we plotted the relative residuals r_k for $k = 1, \ldots, 8$. In Fig. 7 (right) we show results for the SGS method with varying number of smoothing iterations, namely $\nu = 2, 4, 6$. For $\bar{\ell} = 1, \ldots, 5$ we give the average residual reduction per iteration $r := (r_8)^{\frac{1}{8}}$.

These results show the very fast and essentially level independent rate of convergence of this multigrid method. For a larger number of smoothing iterations the convergence is faster. On the other hand, also the costs per iteration then increase, cf. (34) (with $g = \frac{1}{8}$). Usually, in practice the number of smoothings per iteration is not taken very large. Typical values are $\nu = 2$ or $\nu = 4$. In the Fig. 8 we show similar results but now for the damped Jacobi smoother (damping with $\omega = 0.7$) instead of the SGS method.

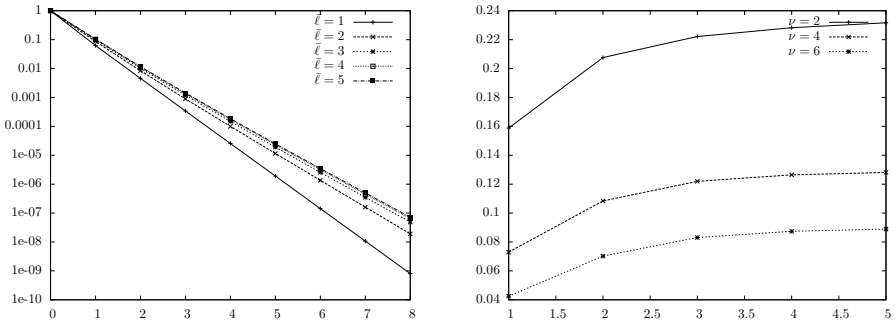


Figure 7. Convergence of multigrid V-cycle with SGS smoother. Left: r_k, for $k = 0, \ldots, 8$ and $\bar{\ell} = 1, \ldots, 5$. Right: $(r_8)^{\frac{1}{8}}$ for $\bar{\ell} = 1, \ldots, 5$ and $\nu = 2, 4, 6$.

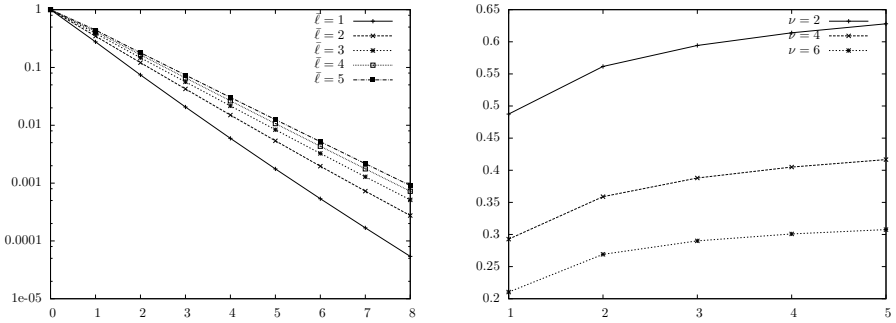


Figure 8. Convergence of multigrid V-cycle with damped Jacobi smoother. Left: r_k, for $k = 0, \ldots, 8$ and $\bar{\ell} = 1, \ldots, 5$. Right: $(r_8)^{\frac{1}{8}}$ for $\bar{\ell} = 1, \ldots, 5$ and $\nu = 2, 4, 6$.

For the method with damped Jacobi smoothing we also clearly observe an essentially level independent rate of convergence. Furthermore there is an increase in the rate of convergence when the number ν of smoothing step gets larger. Comparing the results of the multigrid method with Jacobi smoothing to those with SGS smoothing we see that the latter method has a significantly faster convergence. Note, however, that the arithmetic costs per iteration for the latter method are higher (the ratio lies between 1.5 and 2).

5 Multigrid methods for generalized Stokes equations

Let $\Omega \subset \mathbb{R}^d$, $d = 2$ or 3 be a bounded connected domain. We consider the following generalized Stokes problem: Given $\vec{f}$, find a velocity $\vec{u}$ and a pressure p such that

$$\begin{aligned} \xi\vec{u} - \nu\Delta\vec{u} + \nabla p &= \vec{f} \quad \text{in } \Omega, \\ \nabla \cdot \vec{u} &= 0 \quad \text{in } \Omega, \\ \vec{u} &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{37}$$

The parameters $\nu > 0$ (viscosity) and $\xi \geq 0$ are given. Often the latter is proportional to the inverse of the time step in an implicit time integration method applied to a nonstationary Stokes problem. Note that this general setting includes the classical (stationary) Stokes problem ($\xi = 0$). The weak formulation of (37) is as follows: Given $\vec{f} \in L^2(\Omega)^d$, we seek $\vec{u} \in H_0^1(\Omega)^d$ and $p \in L_0^2(\Omega) := \{\, q \in L^2(\Omega) \mid \int_\Omega q\, dx = 0 \,\}$ such that

$$\begin{aligned} \xi(\vec{u},\vec{v}) + \nu(\nabla\vec{u},\nabla\vec{v}) - (\text{div}\,\vec{v}, p) &= (\vec{f},\vec{v}) \quad \text{for all } \vec{v} \in H_0^1(\Omega)^d, \\ (\text{div}\,\vec{u}, q) &= 0 \quad \text{for all } q \in L_0^2(\Omega). \end{aligned} \tag{38}$$

Here $(\cdot,\cdot)$ denotes the L^2 scalar product.

For discretization of (38) we use a standard finite element approach. Based on a regular family of *nested* tetrahedral grids $\mathcal{T}_\ell = \mathcal{T}_{h_\ell}$ with $\mathcal{T}_0 \subset \mathcal{T}_1 \subset \ldots$ we use a sequence of nested finite element spaces

$$(\mathbf{V}_{\ell-1}, Q_{\ell-1}) \subset (\mathbf{V}_\ell, Q_\ell), \quad \ell = 1, 2, \ldots.$$

The pair of spaces $(\mathbf{V}_\ell, Q_\ell)$, $\ell \geq 0$, is assumed to be stable. By h_ℓ we denote the mesh size parameter corresponding to $\mathcal{T}_\ell$. In our numerical experiments we use the Hood-Taylor $\mathcal{P}_2 - \mathcal{P}_1$ pair:

$$\begin{aligned} &\mathbf{V}_\ell = V_\ell^d, \quad V_\ell := \{\, v \in C(\Omega) \mid v_{|T} \in \mathcal{P}_2 \quad \text{for all } T \in \mathcal{T}_\ell \,\}, \\ &Q_\ell = \{\, v \in C(\Omega) \mid v_{|T} \in \mathcal{P}_1 \quad \text{for all } T \in \mathcal{T}_\ell \,\}. \end{aligned} \tag{39}$$

The discrete problem is given by the Galerkin discretization of (38) with the pair $(\mathbf{V}_\ell, Q_\ell)$. We are interested in the solution of this discrete problem on a given finest discretization level $\ell = \bar{\ell}$. The resulting discrete problem can be represented using the standard nodal bases in these finite element spaces. The representation of the discrete problem on level ℓ in these bases results in a *linear saddle point problem* of the form

$$\mathcal{A}_\ell \mathbf{x}_\ell = \mathbf{b}_\ell \quad \text{with} \quad \mathcal{A}_\ell = \begin{pmatrix} A_\ell & B_\ell^T \\ B_\ell & 0 \end{pmatrix}, \quad \mathbf{x}_\ell = \begin{pmatrix} \mathbf{u}_\ell \\ \mathbf{p}_\ell \end{pmatrix}. \tag{40}$$

The dimensions of the spaces $\mathbf{V}_\ell$ and Q_ℓ are denoted by n_ℓ and m_ℓ, respectively. The matrix $A_\ell \in \mathbb{R}^{n_\ell \times n_\ell}$ is the discrete representation of the differential operator $\xi I - \nu\Delta$ and is symmetric positive definite. Note that A_ℓ depends on the parameters ξ and ν. The matrix $\mathcal{A}_\ell$ depends on these parameters, too, and is *symmetric and strongly indefinite*.

We describe a multigrid method that can be used for the iterative solution of the system (40). This method has the same algorithmic structure as in (31). We need intergrid transfer operators (prolongation and restriction) and a smoother. These components are described below.

Intergrid transfer operators. For the prolongation and restriction of vectors (or corresponding finite element functions) between different level we use the canonical operators. The prolongation between level $\ell - 1$ and ℓ is given by

$$P_\ell = \begin{pmatrix} P_{\mathbf{V}} & 0 \\ 0 & P_Q \end{pmatrix}, \tag{41}$$

where the matrices $P_{\mathbf{V}} : \mathbb{R}^{n_{\ell-1}} \to \mathbb{R}^{n_\ell}$ and $P_Q : \mathbb{R}^{m_{\ell-1}} \to \mathbb{R}^{m_\ell}$ are matrix representations of the embeddings $\mathbf{V}_{\ell-1} \subset \mathbf{V}_\ell$ (quadratic interpolation for $\mathcal{P}_2$) and $Q_{\ell-1} \subset Q_\ell$

(linear interpolation for $\mathcal{P}_1$), respectively. For the restriction operator R_ℓ between the levels ℓ and $\ell-1$ we take the adjoint of P_ℓ (w.r.t. a scaled Euclidean scalar product). Then the Galerkin property $\mathcal{A}_{\ell-1} = R_\ell \mathcal{A}_\ell P_\ell$ holds.

Braess-Sarazin smoother. This smoother is introduced in Braess[10]. With $D_\ell = \mathrm{diag}(A_\ell)$ and a given $\alpha > 0$ the smoothing iteration has the form

$$\begin{pmatrix} \mathbf{u}_\ell^{k+1} \\ \mathbf{p}_\ell^{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{u}_\ell^{k} \\ \mathbf{p}_\ell^{k} \end{pmatrix} - \begin{pmatrix} \alpha D_\ell & B_\ell^T \\ B_\ell & 0 \end{pmatrix}^{-1} \left\{ \begin{pmatrix} A_\ell & B_\ell^T \\ B_\ell & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u}_\ell^{k} \\ \mathbf{p}_\ell^{k} \end{pmatrix} - \begin{pmatrix} \mathbf{f}_\ell \\ 0 \end{pmatrix} \right\}. \tag{42}$$

Each iteration (42) requires the solution of the auxiliary problem

$$\begin{pmatrix} \alpha D_\ell & B_\ell^T \\ B_\ell & 0 \end{pmatrix} \begin{pmatrix} \hat{\mathbf{u}}_\ell \\ \hat{\mathbf{p}}_\ell \end{pmatrix} = \begin{pmatrix} \mathbf{r}_\ell^k \\ B_\ell \mathbf{u}_\ell^k \end{pmatrix} \tag{43}$$

with $\mathbf{r}_\ell^k = A_\ell \mathbf{u}_\ell^k + B_\ell^T \mathbf{p}_\ell^k - \mathbf{f}_\ell$. From (43) one obtains

$$B_\ell \hat{\mathbf{u}}_\ell = B_\ell \mathbf{u}_\ell^k,$$

and hence,

$$B_\ell \mathbf{u}_\ell^{k+1} = B_\ell(\mathbf{u}_\ell^k - \hat{\mathbf{u}}_\ell) = 0 \quad \text{for all} \quad j \geq 0. \tag{44}$$

Therefore, the Braess-Sarazin method can be considered as a smoother on the subspace of vectors that satisfy the constraint equation $B_\ell \mathbf{u}_\ell = 0$.

The problem (43) can be reduced to a problem for the auxiliary pressure unknown $\hat{\mathbf{p}}_\ell$:

$$Z_\ell \hat{\mathbf{p}}_\ell = B_\ell D_\ell^{-1} \mathbf{r}_\ell^k - \alpha B_\ell \mathbf{u}_\ell^k, \tag{45}$$

where $Z_\ell = B_\ell D_\ell^{-1} B_\ell^T$.

Remark 5.1 The matrix Z_ℓ is similar to a discrete Laplace operator on the pressure space. In practice the system (45) is solved approximately using an efficient iterative solver, cf. Braess[10], Zulehner[11]. □

Once $\hat{\mathbf{p}}_\ell$ is known (approximately), an approximation for $\hat{\mathbf{u}}_\ell$ can easily be determined from $\alpha D_\ell \hat{\mathbf{u}}_\ell = \mathbf{r}_\ell^k - B_\ell^T \hat{\mathbf{p}}_\ell$.

Vanka smoother. The Vanka-type smoothers, originally proposed by Vanka[12] for finite difference schemes, are block Gauß-Seidel type of methods. If one uses such a method in a finite element setting then a block of unknowns consists of all degrees of freedom that correspond with one element. Numerical tests given in John[13] show that the use of this element-wise Vanka smoother can be problematic for continuous pressure approximations. In John[13] the pressure-oriented Vanka smoother for continuous pressure approximations has been suggested as a good alternative. In this method a local problem corresponds to the block of unknowns consisting of one pressure unknown and all velocity degrees of freedom that are connected with this pressure unknown. We only consider this type of Vanka smoother. We first give a more precise description of this method.

We take a fixed level ℓ in the discretization hierarchy. To simplify the presentation we drop the level index ℓ from the notation, i.e. we write, for example, $\begin{pmatrix} \mathbf{u} \\ \mathbf{p} \end{pmatrix} \in \mathbb{R}^{n+m}$ instead of $\begin{pmatrix} \mathbf{u}_\ell \\ \mathbf{p}_\ell \end{pmatrix} \in \mathbb{R}^{n_\ell + m_\ell}$. Let $r_P^{(j)} : \mathbb{R}^m \to \mathbb{R}$ be the pressure projection (injection)

$$r_P^{(j)} \mathbf{p} = p_j, \qquad j = 1, \ldots, m.$$

For each j ($1 \leq j \leq m$) let the set of velocity indices that are "connected" to j be given by

$$\mathcal{V}_j = \{1 \leq i \leq n \mid (r_P^{(j)} B)_i \neq 0\}.$$

Define $d_j := |\mathcal{V}_j|$ and write $\mathcal{V}_j = \{i_1 < i_2 < \ldots < i_{d_j}\}$. A corresponding velocity projection operator $r_V^{(j)} : \mathbb{R}^n \to \mathbb{R}^{d_j}$ is given by

$$r_V^{(j)} \mathbf{u} = (u_{i_1}, u_{i_2}, \ldots, u_{i_{d_j}})^T.$$

The combined pressure and velocity projection is given by

$$r^{(j)} = \begin{pmatrix} r_V^{(j)} & 0 \\ 0 & r_P^{(j)} \end{pmatrix} \in \mathbb{R}^{(d_j+1)\times(n+m)}.$$

Furthermore, define $p^{(j)} = \left(r^{(j)}\right)^T$. Using these operators we can formulate a standard multiplicative Schwarz method. Define

$$\mathcal{A}^{(j)} := r^{(j)} \mathcal{A} p^{(j)} =: \begin{pmatrix} A^{(j)} & {B^{(j)}}^T \\ B^{(j)} & 0 \end{pmatrix} \in \mathbb{R}^{(d_j+1)\times(d_j+1)}.$$

Note that $B^{(j)}$ is a row vector of length d_j. In addition, we define

$$\mathcal{D}^{(j)} = \begin{pmatrix} \operatorname{diag}(A^{(j)}) & {B^{(j)}}^T \\ B^{(j)} & 0 \end{pmatrix} = \begin{pmatrix} \ddots & 0 & \vdots \\ 0 & \ddots & \vdots \\ \ldots & \ldots & 0 \end{pmatrix} \in \mathbb{R}^{(d_j+1)\times(d_j+1)}.$$

The *full* Vanka smoother is a multiplicative Schwarz method (or block Gauss-Seidel method) with iteration matrix

$$\mathcal{S}_{\text{full}} = \prod_{j=1}^{m} \left(I - p^{(j)} (\mathcal{A}^{(j)})^{-1} r^{(j)} \mathcal{A}\right). \tag{46}$$

The *diagonal* Vanka smoother is similar, but with $\mathcal{D}^{(j)}$ instead of $\mathcal{A}^{(j)}$:

$$\mathcal{S}_{\text{diag}} = \prod_{j=1}^{m} \left(I - p^{(j)} (\mathcal{D}^{(j)})^{-1} r^{(j)} \mathcal{A}\right). \tag{47}$$

Thus, a smoothing step with a Vanka-type smoother consists of a loop over all pressure degrees of freedom ($j = 1, \ldots, m$), where for each j a linear system of equations with the matrix $\mathcal{A}^{(j)}$ (or $\mathcal{D}^{(j)}$) has to be solved. The degrees of freedom are updated in a Gauss-Seidel manner. These two methods are well-defined if all matrices $\mathcal{A}^{(j)}$ and $\mathcal{D}^{(j)}$ are nonsingular.

The linear systems with the diagonal Vanka smoother can be solved very efficiently using the special structure of the matrix $\mathcal{D}^{(j)}$ whereas for the systems with the full Vanka smoother a direct solver for the systems with the matrices $\mathcal{A}^{(j)}$ is required. The computational costs for solving a local (i.e. for each block) linear system of equations is $\sim d_j$ for the diagonal Vanka smoother and $\sim d_j^3$ for the full Vanka smoother. Typical values for d_j are given in Table 2.

Using the prolongation, restriction and smoothers as explained above a multigrid algorithm for solving the discretized Stokes problem (40) is defined as in (31).

	$h_0 = 2^{-1}$	$h_1 = 2^{-2}$	$h_2 = 2^{-3}$	$h_3 = 2^{-4}$	$h_4 = 2^{-5}$
n_ℓ	81	1029	10125	89373	750141
m_ℓ	27	125	729	4913	35937

Table 1. Dimensions: n_ℓ = number of velocity unknowns, m_ℓ = number of pressure unknowns.

6 Numerical experiment: Multigrid applied to a generalized Stokes equation

We consider the generalized Stokes equation as in (37) on the unit cube $\Omega = (0,1)^3$. The right-hand side $\vec{f}$ is taken such that the continuous solution is

$$\vec{u}(x,y,z) = \frac{1}{3}\begin{pmatrix} \sin(\pi x)\sin(\pi y)\sin(\pi z) \\ -\cos(\pi x)\cos(\pi y)\sin(\pi z) \\ 2\cdot\cos(\pi x)\sin(\pi y)\cos(\pi z)\end{pmatrix},$$

$$p(x,y,z) = \cos(\pi x)\sin(\pi y)\sin(\pi z) + C$$

with a constant C such that $\int_\Omega p\,dx = 0$. For the discretization we start with a uniform tetrahedral grid with $h_0 = \frac{1}{2}$ and we apply regular refinements to this starting discretization. For the finite element discretization we use the Hood-Taylor $\mathcal{P}_2$-$\mathcal{P}_1$ pair, cf. (39). In Table 1 the dimension of the system to be solved on each level and the corresponding mesh size are given.

In all tests below the iterations were repeated until the condition

$$\frac{\|\mathbf{r}^{(k)}\|}{\|\mathbf{r}^{(0)}\|} < 10^{-10},$$

with $\mathbf{r}^{(k)} = \mathbf{b} - \mathcal{A}\mathbf{x}^{(k)}$, was satisfied.

We first consider an experiment to show that for this problem class the multigrid method with *full* Vanka smoother is very time consuming. In Table 2 we show the maximal and mean values of d_j on the level ℓ. These numbers indicate the dimensions of the local systems that have to be solved in the Vanka smoother.

	$h_0 = 2^{-1}$	$h_1 = 2^{-2}$	$h_2 = 2^{-3}$	$h_3 = 2^{-4}$	$h_4 = 2^{-5}$
$\frac{\text{mean}(d_j)}{\max_j d_j}$	21.8 / 82	51.7 / 157	88.8 / 157	119.1 / 165	138.1 / 166

Table 2. The maximal and mean values of d_j on different grids.

We use a multigrid W-cycle with 2 pre- and 2 post-smoothing iterations. In Table 3 we show the computing time (in seconds) and the number of iterations needed both for the full Vanka $\mathcal{S}_{\text{full}}$ and the diagonal Vanka $\mathcal{S}_{\text{diag}}$ smoother.

As can be seen from these results, the rather high dimensions of the local systems lead to high computing times for the multigrid method with the full Vanka smoother compared to the method with the diagonal Vanka smoother. Therefore we prefer the method with the diagonal Vanka smoother. In numerical experiments we observed that the multigrid

$\xi = 0$	$\mathcal{S}_{\rm full}$, $h_3 = 2^{-4}$	$\mathcal{S}_{\rm diag}$,$h_3 = 2^{-4}$	$\mathcal{S}_{\rm full}$, $h_4 = 2^{-5}$	$\mathcal{S}_{\rm diag}$,$h_4 = 2^{-5}$
$\nu = 1$	287 (4)	19 (10)	3504 (5)	224 (13)
$\nu = 10^{-1}$	283 (4)	19 (10)	3449 (5)	238 (13)
$\nu = 10^{-2}$	284 (4)	19 (10)	3463 (5)	238 (13)
$\nu = 10^{-3}$	356 (5)	20 (11)	3502 (5)	238 (13)

Table 3. CPU time and number of iterations for multigrid with the full and the diagonal Vanka smoother.

W-cycle with only *one* pre- and post-smoothing iteration with the diagonal Vanka method sometimes diverges. Further tests indicate that often for the method with diagonal Vanka smoothing the choice $\nu_1 = \nu_2 = 4$ is (slightly) better (w.r.t. CPU time) than $\nu_1 = \nu_2 = 2$.

Results for two variants of the multigrid W-cycle method, one with diagonal Vanka smoothing (V-MGM) and one with Braess-Sarazin smoothing (BS-MGM) are given in the tables 4 and 5. In the V-MGM we use $\nu_1 = \nu_2 = 4$. Based on numerical experiments, in the method with the Braess-Sarazin smoother we use $\nu_1 = \nu_2 = 2$ and $\alpha = 1.25$. For other values $\alpha \in [1.1, 1.5]$ the efficiency is very similar. The linear system in (45) is solved approximately using a conjugate gradient method with a fixed relative tolerance $\varepsilon_{CG} = 10^{-2}$. To investigate the robustness of these method we give results for several values of ℓ, ν and ξ.

$\xi = 0$	$h_3 = 2^{-4}$	
ν	V-MGM	BS-MGM
$\nu = 1$	19 (5)	20 (11)
$\nu = 10^{-1}$	19 (5)	20 (11)
$\nu = 10^{-3}$	19 (5)	17 (8)
$\xi = 10$	$h_3 = 2^{-4}$	
ν	V-MGM	BS-MGM
$\nu = 1$	19 (5)	20 (11)
$\nu = 10^{-1}$	17 (4)	20 (11)
$\nu = 10^{-3}$	15 (3)	21 (7)
$\xi = 100$	$h_3 = 2^{-4}$	
ν	V-MGM	BS-MGM
$\nu = 1$	17 (4)	20 (11)
$\nu = 10^{-1}$	15 (3)	19 (7)
$\nu = 10^{-3}$	15 (3)	19 (6)

Table 4. CPU time and the number of iterations for BS- and V-MGM methods.

The results show that the rate of convergence is essentially independent of the parameters ν and ξ, i.e., these methods have a robustness property. Furthermore we observe that if for fixed ν, ξ we compare the results for $\ell = 3$ ($h_3 = 2^{-4}$) with those for $\ell = 4$ ($h_4 = 2^{-5}$) then for the V-MGM there is (almost) no increase in the number of iterations. This illustrates the mesh independent rate of convergence of the method. For the BS-MGM there is a (small) growth in the number of iterations. For both methods the CPU time needed per iteration grows with a factor of roughly 10 when going from $\ell = 3$ to

$\xi = 0$	$h_4 = 2^{-5}$	
ν	V-MGM	BS-MGM
$\nu = 1$	198 (5)	274 (14)
$\nu = 10^{-1}$	199 (5)	276 (14)
$\nu = 10^{-3}$	198 (5)	241 (11)
$\xi = 10$	$h_3 = 2^{-5}$	
ν	V-MGM	BS-MGM
$\nu = 1$	190 (5)	244 (13)
$\nu = 10^{-1}$	189 (5)	224 (10)
$\nu = 10^{-3}$	145 (3)	238 (7)
$\xi = 100$	$h_3 = 2^{-5}$	
ν	V-MGM	BS-MGM
$\nu = 1$	190 (5)	241 (13)
$\nu = 10^{-1}$	167 (4)	243 (13)
$\nu = 10^{-3}$	122 (2)	282 (9)

Table 5. CPU time and the number of iterations for BS- and V-MGM methods.

$\ell = 4$. The number of unknowns then grows with about a factor 8.3, cf. Table 1. This indicates that the arithmetic work per iteration is almost linear in the number of unknowns.

7 Convergence analysis for scalar elliptic problems

In this section we present a convergence analysis for the multigrid method introduced in section 3. Our approach is based on the so-called approximation- and smoothing property, introduced by Hackbusch[1, 14]. For a discussion of other analyses we refer to remark 7.23.

7.1 Introduction

One easily verifies that the two-grid method is a linear iterative method. The iteration matrix of this method with ν_1 presmoothing and ν_2 postsmoothing iterations on level ℓ is given by

$$\mathbf{C}_{TG,\ell} = \mathbf{C}_{TG,\ell}(\nu_2, \nu_1) = \mathbf{S}_\ell^{\nu_2}(\mathbf{I} - \mathbf{p}_\ell \mathbf{A}_{\ell-1}^{-1} \mathbf{r}_\ell \mathbf{A}_\ell)\mathbf{S}_\ell^{\nu_1} \tag{48}$$

with $\mathbf{S}_\ell = \mathbf{I} - \mathbf{M}_\ell^{-1}\mathbf{A}_\ell$ the iteration matrix of the smoother.

Theorem 7.1 *The multigrid method* (31) *is a linear iterative method with iteration matrix* $\mathbf{C}_{MG,\ell}$ *given by*

$$\mathbf{C}_{MG,0} = 0 \tag{49a}$$

$$\mathbf{C}_{MG,\ell} = \mathbf{S}_\ell^{\nu_2}\big(\mathbf{I} - \mathbf{p}_\ell(\mathbf{I} - \mathbf{C}_{MG,\ell-1}^{\tau})\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\big)\mathbf{S}_\ell^{\nu_1} \tag{49b}$$

$$= \mathbf{C}_{TG,\ell} + \mathbf{S}_\ell^{\nu_2}\mathbf{p}_\ell\mathbf{C}_{MG,\ell-1}^{\tau}\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\mathbf{S}_\ell^{\nu_1}, \quad \ell = 1, 2, \ldots \tag{49c}$$

Proof: The result in (49a) is trivial. The result in (49c) follows from (49b) and the definition of $\mathbf{C}_{TG,\ell}$. We now prove the result in (49b) by induction. For $\ell = 1$ it follows from (49a) and (48). Assume that the result is correct for $\ell - 1$. Then $\mathrm{MGM}_{\ell-1}(\mathbf{y}_{\ell-1}, \mathbf{z}_{\ell-1})$ defines a linear iterative method and for arbitrary $\mathbf{y}_{\ell-1},\ \mathbf{z}_{\ell-1} \in \mathbb{R}^{n_{\ell-1}}$ we have

$$\mathrm{MGM}_{\ell-1}(\mathbf{y}_{\ell-1}, \mathbf{z}_{\ell-1}) - \mathbf{A}_{\ell-1}^{-1}\mathbf{z}_{\ell-1} = \mathbf{C}_{MG,\ell-1}(\mathbf{y}_{\ell-1} - \mathbf{A}_{\ell-1}^{-1}\mathbf{z}_{\ell-1}) \tag{50}$$

We rewrite the algorithm (31) as follows:

$$\begin{aligned}
\mathbf{x}^1 &:= \mathcal{S}_\ell^{\nu_1}(\mathbf{x}_\ell^{\mathrm{old}}, \mathbf{b}_\ell) \\
\mathbf{x}^2 &:= \mathbf{x}^1 + \mathbf{p}_\ell \mathrm{MGM}_{\ell-1}^{\tau}\big(0, \mathbf{r}_\ell(\mathbf{b}_\ell - \mathbf{A}_\ell \mathbf{x}^1)\big) \\
\mathbf{x}_\ell^{\mathrm{new}} &:= \mathcal{S}_\ell^{\nu_2}(\mathbf{x}^2, \mathbf{b}_\ell).
\end{aligned}$$

From this we get

$$\begin{aligned}
\mathbf{x}_\ell^{\mathrm{new}} - \mathbf{x}_\ell^* &= \mathbf{x}_\ell^{\mathrm{new}} - \mathbf{A}_\ell^{-1}\mathbf{b}_\ell = \mathbf{S}_\ell^{\nu_2}(\mathbf{x}^2 - \mathbf{x}_\ell^*) \\
&= \mathbf{S}_\ell^{\nu_2}\big(\mathbf{x}^1 - \mathbf{x}_\ell^* + \mathbf{p}_\ell \mathrm{MGM}_{\ell-1}^{\tau}\big(0, \mathbf{r}_\ell(\mathbf{b}_\ell - \mathbf{A}_\ell \mathbf{x}^1)\big)\big).
\end{aligned}$$

Now we use the result (50) with $\mathbf{y}_{\ell-1} = 0,\ \mathbf{z}_{\ell-1} := \mathbf{r}_\ell(\mathbf{b}_\ell - \mathbf{A}_\ell \mathbf{x}^1)$. This yields

$$\begin{aligned}
\mathbf{x}_\ell^{\mathrm{new}} - \mathbf{x}_\ell^* &= \mathbf{S}_\ell^{\nu_2}\big(\mathbf{x}^1 - \mathbf{x}_\ell^* + \mathbf{p}_\ell(\mathbf{A}_{\ell-1}^{-1}\mathbf{z}_{\ell-1} - \mathbf{C}_{MG,\ell-1}^{\tau}\mathbf{A}_{\ell-1}^{-1}\mathbf{z}_{\ell-1}\big) \\
&= \mathbf{S}_\ell^{\nu_2}\big(\mathbf{I} - \mathbf{p}_\ell(\mathbf{I} - \mathbf{C}_{MG,\ell-1}^{\tau})\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell \mathbf{A}_\ell\big)(\mathbf{x}^1 - \mathbf{x}_\ell^*) \\
&= \mathbf{S}_\ell^{\nu_2}\big(\mathbf{I} - \mathbf{p}_\ell(\mathbf{I} - \mathbf{C}_{MG,\ell-1}^{\tau})\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell \mathbf{A}_\ell\big)\mathbf{S}_\ell^{\nu_1}(\mathbf{x}^{\mathrm{old}} - \mathbf{x}_\ell^*).
\end{aligned}$$

This completes the proof. ∎

The convergence analysis will be based on the following splitting of the two-grid iteration matrix, with $\nu_2 = 0$, i.e. no postsmoothing:

$$\begin{aligned}
\|\mathbf{C}_{TG,\ell}(0,\nu_1)\|_2 &= \|(\mathbf{I} - \mathbf{p}_\ell \mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell \mathbf{A}_\ell)\mathbf{S}_\ell^{\nu_1}\|_2 \\
&\le \|\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell \mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\|_2 \, \|\mathbf{A}_\ell \mathbf{S}_\ell^{\nu_1}\|_2
\end{aligned} \tag{51}$$

In section 7.2 we will prove a bound of the form $\|\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell \mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\|_2 \le C_A \|\mathbf{A}_\ell\|_2^{-1}$. This result is called the *approximation property*. In section 7.3 we derive a suitable bound for the term $\|\mathbf{A}_\ell \mathbf{S}_\ell^{\nu_1}\|_2$. This is the so-called *smoothing property*. In section 7.4 we combine these bounds with the results in (51) and in theorem 7.1. This yields bounds for the contraction number of the two-grid method and of the multigrid W-cycle. For the V-cycle a more subtle analysis is needed. This is presented in section 7.5. In the convergence analysis we need the following:

Assumption 7.2 *In the sections 7.2–7.5 we assume that the family of triangulations $\{\mathcal{T}_{h_\ell}\}$ corresponding to the finite element spaces V_ℓ, $\ell = 0, 1, \ldots$, is* quasi-uniform *and that $h_{\ell-1} \le c h_\ell$ with a constant c independent of ℓ.*

We give some results that will be used in the analysis further on. First we recall an *inverse inequality* that is known from the analysis of finite element methods:

$$|v_\ell|_1 \le c\, h_\ell^{-1} \|v_\ell\|_{L^2} \quad \text{for all } \ v_\ell \in V_\ell$$

with a constant c independent of ℓ. For this result to hold we need assumption 7.2.
We now show that, apart from a scaling factor, the isomorphism $P_\ell : (\mathbb{R}^{n_\ell}, \langle\cdot,\cdot\rangle) \to (V_\ell, \langle\cdot,\cdot\rangle_{L^2})$ and its inverse are uniformly (w.r.t. ℓ) bounded:

Lemma 7.3 *There exist constants $c_1 > 0$ and c_2 independent of ℓ such that*

$$c_1 \|P_\ell \mathbf{x}\|_{L^2} \leq h_\ell^{\frac{1}{2}d} \|\mathbf{x}\|_2 \leq c_2 \|P_\ell \mathbf{x}\|_{L^2} \quad \textit{for all } \mathbf{x} \in \mathbb{R}^{n_\ell}. \tag{52}$$

Proof: The definition of P_ℓ yields $P_\ell \mathbf{x} = \sum_{i=1}^{n_\ell} x_i \phi_i =: v_\ell \in V_\ell$ and $v_\ell(\xi_i) = x_i$, where ξ_i is the vertex in the triangulation which corresponds to the nodal basis function ϕ_i. Note that

$$\|P_\ell \mathbf{x}\|_{L^2}^2 = \|v_\ell\|_{L^2}^2 = \sum_{T \in \mathcal{T}_\ell} \|v_\ell\|_{L^2(T)}^2.$$

Since v_ℓ is linear on each simplex T in the triangulation $\mathcal{T}_\ell$ there are constants $\tilde{c}_1 > 0$ and $\tilde{c}_2$ independent of h_ℓ such that

$$\tilde{c}_1 \|v_\ell\|_{L^2(T)}^2 \leq |T| \sum_{\xi_j \in V(T)} v_\ell(\xi_j)^2 \leq \tilde{c}_2 \|v_\ell\|_{L^2(T)}^2,$$

where $V(T)$ denotes the set of vertices of the simplex T. Summation over all $T \in \mathcal{T}_\ell$, using $v_\ell(\xi_j) = x_j$ and $|T| \sim h_\ell^d$ we obtain

$$\hat{c}_1 \|v_\ell\|_{L^2}^2 \leq h_\ell^d \sum_{i=1}^{n_\ell} x_i^2 \leq \hat{c}_2 \|v_\ell\|_{L^2}^2,$$

with constants $\hat{c}_1 > 0$ and $\hat{c}_2$ independent of h_ℓ and thus we get the result in (52). ∎

The third preliminary result concerns the scaling of the stiffness matrix:

Lemma 7.4 *Let $\mathbf{A}_\ell$ be the stiffness matrix as in* (26)*. Assume that the bilinear form is such that the usual conditions* (22) *are satisfied. Then there exist constants $c_1 > 0$ and c_2 independent of ℓ such that*

$$c_1 h_\ell^{d-2} \leq \|\mathbf{A}_\ell\|_2 \leq c_2 h_\ell^{d-2}.$$

Proof: First note that

$$\|\mathbf{A}_\ell\|_2 = \max_{\mathbf{x},\mathbf{y} \in \mathbb{R}^{n_\ell}} \frac{\langle \mathbf{A}_\ell \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}.$$

Using the result in lemma 7.3, the continuity of the bilinear form and the inverse inequality we get

$$\begin{aligned} \max_{\mathbf{x},\mathbf{y} \in \mathbb{R}^{n_\ell}} \frac{\langle \mathbf{A}_\ell \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} &\leq c h_\ell^d \max_{v_\ell, w_\ell \in V_\ell} \frac{k(v_\ell, w_\ell)}{\|v_\ell\|_{L^2} \|w_\ell\|_{L^2}} \\ &\leq c h_\ell^d \max_{v_\ell, w_\ell \in V_\ell} \frac{|v_\ell|_1 |w_\ell|_1}{\|v_\ell\|_{L^2} \|w_\ell\|_{L^2}} \leq c\, h_\ell^{d-2} \end{aligned}$$

and thus the upper bound is proved. The lower bound follows from

$$\max_{\mathbf{x},\mathbf{y} \in \mathbb{R}^{n_\ell}} \frac{\langle \mathbf{A}_\ell \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \geq \max_{1 \leq i \leq n_\ell} \langle \mathbf{A}_\ell \mathbf{e}_i, \mathbf{e}_i \rangle = k(\phi_i, \phi_i) \geq c |\phi_i|_1^2 \geq c h_\ell^{d-2}$$

The last inequality can be shown by using for $T \subset \operatorname{supp}(\phi_i)$ the affine transformation from the unit simplex to T. ∎

7.2 Approximation property

In this section we derive a bound for the first factor in the splitting (51). We start with two important assumptions that are crucial for the analysis. This first one concerns *regularity of the continuous problem*, the second one is a *discretization error bound*.

Assumption 7.5 *We assume that the continuous problem in* (23) *is H^2-regular, i.e. for $f \in L^2(\Omega)$ the corresponding solution u satisfies*

$$\|u\|_{H^2} \leq c\,\|f\|_{L^2},$$

with a constant c independent of f. Furthermore we assume a finite element discretization error bound for the Galerkin discretization (25)*:*

$$\|u - u_\ell\|_{L^2} \leq c h_\ell^2 \|f\|_{L^2}$$

with c independent of f and of ℓ.

We will need the *dual* problem of (23) which is as follows: determine $\tilde{u} \in H_0^1(\Omega)$ such that $k(v, \tilde{u}) = f(v)$ for all $v \in H_0^1(\Omega)$. Note that this dual problem is obtained by interchanging the arguments in the bilinear form $k(\cdot, \cdot)$ and that the dual problem equals the original one if the bilinear form is symmetric (as for example in case of the Poisson equation).

In the analysis we will use the adjoint operator $P_\ell^* : V_\ell \to \mathbb{R}^{n_\ell}$ which satisfies $\langle P_\ell \mathbf{x}, v_\ell \rangle_{L^2} = \langle \mathbf{x}, P_\ell^* v_\ell \rangle$ for all $\mathbf{x} \in \mathbb{R}^{n_\ell},\ v_\ell \in V_\ell$. As a direct consequence of lemma 7.3 we obtain

$$c_1 \|P_\ell^* v_\ell\|_2 \leq h_\ell^{\frac{1}{2}d} \|v_\ell\|_{L^2} \leq c_2 \|P_\ell^* v_\ell\|_2 \quad \text{for all } \ v_\ell \in V_\ell \tag{53}$$

with constants $c_1 > 0$ and c_2 independent of ℓ. We now formulate a main result for the convergence analysis of multigrid methods:

Theorem 7.6 (Approximation property.) *Consider $\mathbf{A}_\ell,\ \mathbf{p}_\ell,\ \mathbf{r}_\ell$ as defined in* (26), (29),(30)*. Assume that the variational problem* (23) *is such that the usual conditions* (22) *are satisfied. Moreover, the problem* (23) *and the corresponding dual problem are assumed to be H^2-regular. Then there exists a constant C_A independent of ℓ such that*

$$\|\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell \mathbf{A}_{\ell-1}^{-1} \mathbf{r}_\ell\|_2 \leq C_A \|\mathbf{A}_\ell\|_2^{-1} \quad \textit{for } \ \ell = 1, 2, \ldots \tag{54}$$

Proof: Let $\mathbf{b}_\ell \in \mathbb{R}^{n_\ell}$ be given. The constants in the proof are independent of $\mathbf{b}_\ell$ and of ℓ. Consider the variational problems:

$$\begin{aligned}
u \in H_0^1(\Omega): &\quad k(u, v) = \langle (P_\ell^*)^{-1} \mathbf{b}_\ell, v \rangle_{L^2} \quad \text{for all } \ v \in H_0^1(\Omega) \\
u_\ell \in V_\ell: &\quad k(u_\ell, v_\ell) = \langle (P_\ell^*)^{-1} \mathbf{b}_\ell, v_\ell \rangle_{L^2} \quad \text{for all } \ v_\ell \in V_\ell \\
u_{\ell-1} \in V_{\ell-1}: &\quad k(u_{\ell-1}, v_{\ell-1}) = \langle (P_\ell^*)^{-1} \mathbf{b}_\ell, v_{\ell-1} \rangle_{L^2} \quad \text{for all } \ v_{\ell-1} \in V_{\ell-1}.
\end{aligned}$$

Then

$$\mathbf{A}_\ell^{-1} \mathbf{b}_\ell = P_\ell^{-1} u_\ell \ \text{ and } \ \mathbf{A}_{\ell-1}^{-1} \mathbf{r}_\ell \mathbf{b}_\ell = P_{\ell-1}^{-1} u_{\ell-1}$$

hold. Hence we obtain, using lemma 7.3,

$$\|(\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell \mathbf{A}_{\ell-1}^{-1} \mathbf{r}_\ell) \mathbf{b}_\ell\|_2 = \|P_\ell^{-1}(u_\ell - u_{\ell-1})\|_2 \leq c\, h_\ell^{-\frac{1}{2}d} \|u_\ell - u_{\ell-1}\|_{L^2}. \tag{55}$$

Now we use the assumptions on the discretization error bound and on the H^2-regularity of the problem. This yields

$$\begin{aligned} \|u_\ell - u_{\ell-1}\|_{L^2} &\leq \|u_\ell - u\|_{L^2} + \|u_{\ell-1} - u\|_{L^2} \\ &\leq c h_\ell^2 |u|_2 + + c h_{\ell-1}^2 |u|_2 \leq c h_\ell^2 \|(P_\ell^*)^{-1} \mathbf{b}_\ell\|_{L^2} \end{aligned} \tag{56}$$

We combine (55) with (56) and use (53), and get

$$\|(\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell \mathbf{A}_{\ell-1}^{-1} \mathbf{r}_\ell)\mathbf{b}_\ell\|_2 \leq c\, h_\ell^{2-d} \|\mathbf{b}_\ell\|_2$$

and thus $\|\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell \mathbf{A}_{\ell-1}^{-1} \mathbf{r}_\ell\|_2 \leq c\, h_\ell^{2-d}$. The proof is completed if we use lemma 7.4. ■

Note that in the proof of the approximation property we use the underlying continuous problem.

7.3 Smoothing property

In this section we derive inequalities of the form

$$\|\mathbf{A}_\ell \mathbf{S}_\ell^\nu\|_2 \leq g(\nu) \|\mathbf{A}_\ell\|_2$$

where $g(\nu)$ is a monotonically decreasing function with $\lim_{\nu \to \infty} g(\nu) = 0$. In the first part of this section we derive results for the case that $\mathbf{A}_\ell$ is symmetric positive definite. In the second part we discuss the general case.

Smoothing property for the symmetric positive definite case
We start with an elementary lemma:
Lemma 7.7 *Let $\mathbf{B} \in \mathbb{R}^{m \times m}$ be a symmetric positive definite matrix with $\sigma(\mathbf{B}) \subset (0, 1]$. Then we have*

$$\|\mathbf{B}(\mathbf{I} - \mathbf{B})^\nu\|_2 \leq \frac{1}{2(\nu + 1)} \quad \textit{for } \nu = 1, 2, \ldots$$

Proof: Note that

$$\|\mathbf{B}(\mathbf{I} - \mathbf{B})^\nu\|_2 = \max_{x \in (0,1]} x(1 - x)^\nu = \frac{1}{\nu + 1} \big(\frac{\nu}{\nu + 1}\big)^\nu.$$

A simple computation shows that $\nu \to \big(\frac{\nu}{\nu+1}\big)^\nu$ is decreasing on $[1, \infty)$. ■

Below for a few basic iterative methods we derive the smoothing property for the symmetric case, i.e., $\mathbf{b} = 0$ in the bilinear form $k(\cdot, \cdot)$. We first consider the Richardson method:
Theorem 7.8 *Assume that in the bilinear form we have $\mathbf{b} = 0$ and that the usual conditions* (22) *are satisfied. Let $\mathbf{A}_\ell$ be the stiffness matrix in* (26)*. For $c_0 \in (0, 1]$ we have the smoothing property*

$$\|\mathbf{A}_\ell (\mathbf{I} - \frac{c_0}{\rho(\mathbf{A}_\ell)} \mathbf{A}_\ell)^\nu\|_2 \leq \frac{1}{2 c_0 (\nu + 1)} \|\mathbf{A}_\ell\|_2 \,, \quad \nu = 1, 2, \ldots$$

holds.

Proof: Note that $\mathbf{A}_\ell$ is symmetric positive definite. Apply lemma 7.7 with $\mathbf{B} := \omega_\ell \mathbf{A}_\ell$, $\omega_\ell := c_0\, \rho(\mathbf{A}_\ell)^{-1}$. This yields

$$\|\mathbf{A}_\ell(\mathbf{I} - \omega_\ell \mathbf{A}_\ell)^\nu\|_2 \leq \omega_\ell^{-1} \frac{1}{2(\nu+1)} \leq \frac{1}{2c_0(\nu+1)} \rho(\mathbf{A}_\ell) = \frac{1}{2c_0(\nu+1)} \|\mathbf{A}_\ell\|_2$$

and thus the result is proved. ∎

A similar result can be shown for the damped Jacobi method:

Theorem 7.9 *Assume that in the bilinear form we have* $\mathbf{b} = 0$ *and that the usual conditions* (22) *are satisfied. Let* $\mathbf{A}_\ell$ *be the stiffness matrix in* (26) *and* $\mathbf{D}_\ell := \mathrm{diag}(\mathbf{A}_\ell)$. *There exists an* $\omega \in (0, \rho(\mathbf{D}_\ell^{-1}\mathbf{A}_\ell)^{-1}]$, *independent of* ℓ, *such that the smoothing property*

$$\|\mathbf{A}_\ell(\mathbf{I} - \omega \mathbf{D}_\ell^{-1}\mathbf{A}_\ell)^\nu\|_2 \leq \frac{1}{2\omega(\nu+1)} \|\mathbf{A}_\ell\|_2\,, \quad \nu = 1, 2, \ldots$$

holds.

Proof: Define the symmetric positive definite matrix $\tilde{\mathbf{B}} := \mathbf{D}_\ell^{-\frac{1}{2}} \mathbf{A}_\ell \mathbf{D}_\ell^{-\frac{1}{2}}$. Note that

$$(D_\ell)_{ii} = (A_\ell)_{ii} = k(\phi_i, \phi_i) \geq c\, |\phi_i|_1^2 \geq c\, h_\ell^{d-2}\,, \tag{57}$$

with $c > 0$ independent of ℓ and i. Using this in combination with lemma 7.4 we get

$$\|\tilde{\mathbf{B}}\|_2 \leq \frac{\|\mathbf{A}_\ell\|_2}{\lambda_{\min}(\mathbf{D}_\ell)} \leq c\,, \quad c \text{ independent of } \ell.$$

Hence for $\omega \in (0, \frac{1}{c}] \subset (0, \rho(\mathbf{D}_\ell^{-1}\mathbf{A}_\ell)^{-1}]$ we have $\sigma(\omega\tilde{\mathbf{B}}) \subset (0, 1]$. Application of lemma 7.7, with $\mathbf{B} = \omega\tilde{\mathbf{B}}$, yields

$$\begin{aligned}\|\mathbf{A}_\ell(\mathbf{I} - \omega \mathbf{D}_\ell^{-1}\mathbf{A}_\ell)^\nu\|_2 &\leq \omega^{-1} \|\mathbf{D}_\ell^{\frac{1}{2}}\|_2 \|\omega\tilde{\mathbf{B}}(\mathbf{I} - \omega\tilde{\mathbf{B}})^\nu\|_2 \|\mathbf{D}_\ell^{\frac{1}{2}}\|_2 \\ &\leq \frac{\|\mathbf{D}_\ell\|_2}{2\omega(\nu+1)} \leq \frac{1}{2\omega(\nu+1)} \|\mathbf{A}_\ell\|_2\end{aligned}$$

and thus the result is proved. ∎

Remark 7.10 The value of the parameter ω used in theorem 7.9 is such that $\omega\rho(\mathbf{D}_\ell^{-1}\mathbf{A}_\ell) = \omega\rho(\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{A}_\ell\mathbf{D}_\ell^{-\frac{1}{2}}) \leq 1$ holds. Note that

$$\rho(\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{A}_\ell\mathbf{D}_\ell^{-\frac{1}{2}}) = \max_{\mathbf{x}\in\mathbb{R}^{n_\ell}} \frac{\langle \mathbf{A}_\ell\mathbf{x}, \mathbf{x}\rangle}{\langle \mathbf{D}_\ell\mathbf{x}, \mathbf{x}\rangle} \geq \max_{1\leq i\leq n_\ell} \frac{\langle \mathbf{A}_\ell\mathbf{e}_i, \mathbf{e}_i\rangle}{\langle \mathbf{D}_\ell\mathbf{e}_i\mathbf{e}_i\rangle} = 1$$

and thus we have $\omega \leq 1$. This explains why in multigrid methods one usually uses a *damped* Jacobi method as a smoother. □

We finally consider the symmetric Gauss-Seidel method. If $\mathbf{A}_\ell = \mathbf{A}_\ell^T$ this method has an iteration matrix

$$\mathbf{S}_\ell = \mathbf{I} - \mathbf{M}_\ell^{-1}\mathbf{A}_\ell, \quad \mathbf{M}_\ell = (\mathbf{D}_\ell - \mathbf{L}_\ell)\mathbf{D}_\ell^{-1}(\mathbf{D}_\ell - \mathbf{L}_\ell^T)\,, \tag{58}$$

where we use the decomposition $\mathbf{A}_\ell = \mathbf{D}_\ell - \mathbf{L}_\ell - \mathbf{L}_\ell^T$ with $\mathbf{D}_\ell$ a diagonal matrix and $\mathbf{L}_\ell$ a strictly lower triangular matrix.

Theorem 7.11 *Assume that in the bilinear form we have* $\mathbf{b} = 0$ *and that the usual conditions* (22) *are satisfied. Let* $\mathbf{A}_\ell$ *be the stiffness matrix in* (26) *and* $\mathbf{M}_\ell$ *as in* (58)*. The smoothing property*

$$\|\mathbf{A}_\ell(\mathbf{I} - \mathbf{M}_\ell^{-1}\mathbf{A}_\ell)^\nu\|_2 \leq \frac{c}{\nu+1}\|\mathbf{A}_\ell\|_2\,, \quad \nu = 1, 2, \ldots$$

holds with a constant c *independent of* ν *and* ℓ*.*

Proof: Note that $\mathbf{M}_\ell = \mathbf{A}_\ell + \mathbf{L}_\ell\mathbf{D}_\ell^{-1}\mathbf{L}_\ell^T$ and thus $\mathbf{M}_\ell$ is symmetric positive definite. Define the symmetric positive definite matrix $\mathbf{B} := \mathbf{M}_\ell^{-\frac{1}{2}}\mathbf{A}_\ell\mathbf{M}_\ell^{-\frac{1}{2}}$. From

$$0 < \max_{\mathbf{x}\in\mathbb{R}^{n_\ell}} \frac{\langle \mathbf{B}\mathbf{x}, \mathbf{x}\rangle}{\langle \mathbf{x}, \mathbf{x}\rangle} = \max_{\mathbf{x}\in\mathbb{R}^{n_\ell}} \frac{\langle \mathbf{A}_\ell\mathbf{x}, \mathbf{x}\rangle}{\langle \mathbf{M}_\ell\mathbf{x}, \mathbf{x}\rangle} = \max_{\mathbf{x}\in\mathbb{R}^{n_\ell}} \frac{\langle \mathbf{A}_\ell\mathbf{x}, \mathbf{x}\rangle}{\langle \mathbf{A}_\ell\mathbf{x}, \mathbf{x}\rangle + \langle \mathbf{D}_\ell^{-1}\mathbf{L}_\ell^T\mathbf{x}, \mathbf{L}_\ell^T\mathbf{x}\rangle} \leq 1$$

it follows that $\sigma(\mathbf{B}) \subset (0, 1]$. Application of lemma 7.7 yields

$$\|\mathbf{A}_\ell(\mathbf{I} - \mathbf{M}_\ell^{-1}\mathbf{A}_\ell)^\nu\|_2 \leq \|\mathbf{M}_\ell^{\frac{1}{2}}\|_2^2\,\|\mathbf{B}(\mathbf{I} - \mathbf{B})^\nu\|_2 \leq \|\mathbf{M}_\ell\|_2 \frac{1}{2(\nu+1)}.$$

From (57) we have $\|\mathbf{D}_\ell^{-1}\|_2 \leq c\,h_\ell^{2-d}$. Using the sparsity of $\mathbf{A}_\ell$ we obtain

$$\|\mathbf{L}_\ell\|_2\|\mathbf{L}_\ell^T\|_2 \leq \|\mathbf{L}_\ell\|_\infty\|\mathbf{L}_\ell\|_1 \leq c(\max_{i,j}|(A_\ell)_{ij}|)^2 \leq c\|\mathbf{A}_\ell\|_2^2.$$

In combination with lemma 7.4 we then get

$$\|\mathbf{M}_\ell\|_2 \leq \|\mathbf{D}_\ell^{-1}\|_2\|\mathbf{L}_\ell\|_2\|\mathbf{L}_\ell^T\|_2 \leq c\,h_\ell^{2-d}\|\mathbf{A}_\ell\|_2^2 \leq c\|\mathbf{A}_\ell\|_2 \tag{59}$$

and this completes the proof. ∎

For the symmetric positive definite case smoothing properties have also been proved for other iterative methods. For example, in Wittum[15,16] a smoothing property is proved for a variant of the ILU method and in Bröker et al.[17] it is shown that the SPAI (sparse approximate inverse) preconditioner satisfies a smoothing property.

Smoothing property for the nonsymmetric case

For the analysis of the smoothing property in the general (possibly nonsymmetric) case we can not use lemma 7.7. Instead the analysis will be based on the following lemma (cf. Reusken[18,19]):

Lemma 7.12 *Let* $\|\cdot\|$ *be any induced matrix norm and assume that for* $\mathbf{B} \in \mathbb{R}^{m\times m}$ *the inequality* $\|\mathbf{B}\| \leq 1$ *holds. The we have*

$$\|(\mathbf{I} - \mathbf{B})(\mathbf{I} + \mathbf{B})^\nu\| \leq 2^{\nu+1}\sqrt{\frac{2}{\pi\nu}}\,, \quad \textit{for } \nu = 1, 2, \ldots$$

Proof: Note that

$$(\mathbf{I} - \mathbf{B})(\mathbf{I} + \mathbf{B})^\nu = (\mathbf{I} - \mathbf{B})\sum_{k=0}^{\nu}\binom{\nu}{k}\mathbf{B}^k = \mathbf{I} - \mathbf{B}^{\nu+1} + \sum_{k=1}^{\nu}\left(\binom{\nu}{k} - \binom{\nu}{k-1}\right)\mathbf{B}^k.$$

This yields

$$\|(\mathbf{I} - \mathbf{B})(\mathbf{I} + \mathbf{B})^\nu\| \leq 2 + \sum_{k=1}^{\nu}\left|\binom{\nu}{k} - \binom{\nu}{k-1}\right|.$$

Using $\binom{\nu}{k} \geq \binom{\nu}{k-1} \;\Leftrightarrow\; k \leq \frac{1}{2}(\nu+1)$ and $\binom{\nu}{k} \geq \binom{\nu}{\nu-k}$ we get (with $[\,\cdot\,]$ the round down operator):

$$\begin{aligned}
&\sum_{k=1}^{\nu} \left| \binom{\nu}{k} - \binom{\nu}{k-1} \right| \\
&= \sum_{1}^{[\frac{1}{2}(\nu+1)]} \left(\binom{\nu}{k} - \binom{\nu}{k-1} \right) + \sum_{[\frac{1}{2}(\nu+1)]+1}^{\nu} \left(\binom{\nu}{k-1} - \binom{\nu}{k} \right) \\
&= \sum_{1}^{[\frac{1}{2}\nu]} \left(\binom{\nu}{k} - \binom{\nu}{k-1} \right) + \sum_{m=1}^{[\frac{1}{2}\nu]} \left(\binom{\nu}{m} - \binom{\nu}{m-1} \right) \\
&= 2 \sum_{k=1}^{[\frac{1}{2}\nu]} \left(\binom{\nu}{k} - \binom{\nu}{k-1} \right) = 2 \left(\binom{\nu}{[\frac{1}{2}\nu]} - \binom{\nu}{0} \right).
\end{aligned}$$

An elementary analysis yields (cf., for example, Reusken[19])

$$\binom{\nu}{[\frac{1}{2}\nu]} \leq 2^{\nu} \sqrt{\frac{2}{\pi\nu}} \quad \text{for } \nu \geq 1.$$

Thus we have proved the bound. ∎

Corollary 7.13 *Let $\|\cdot\|$ be any induced matrix norm. Assume that for a linear iterative method with iteration matrix $\mathbf{I} - \mathbf{M}_\ell^{-1}\mathbf{A}_\ell$ we have*

$$\|\mathbf{I} - \mathbf{M}_\ell^{-1}\mathbf{A}_\ell\| \leq 1 \tag{60}$$

Then for $\mathbf{S}_\ell := \mathbf{I} - \frac{1}{2}\mathbf{M}_\ell^{-1}\mathbf{A}_\ell$ the following smoothing property holds:

$$\|\mathbf{A}_\ell \mathbf{S}_\ell^{\nu}\| \leq 2\sqrt{\frac{2}{\pi\nu}}\, \|\mathbf{M}_\ell\|\,, \quad \nu = 1, 2, \ldots$$

Proof: Define $\mathbf{B} = \mathbf{I} - \mathbf{M}_\ell^{-1}\mathbf{A}_\ell$ and apply lemma 7.12:

$$\|\mathbf{A}_\ell \mathbf{S}_\ell^{\nu}\| \leq \|\mathbf{M}_\ell\| \left(\frac{1}{2}\right)^{\nu} \|(\mathbf{I}-\mathbf{B})(\mathbf{I}+\mathbf{B})^{\nu}\| \leq 2\sqrt{\frac{2}{\pi\nu}} \|\mathbf{M}_\ell\|.$$

∎

Remark 7.14 Note that in the smoother in corollary 7.13 we use damping with a factor $\frac{1}{2}$. Generalizations of the results in lemma 7.12 and corollary 7.13 are given in Nevanlinna[20], Hackbusch[21], Zulehner[22]. In Nevanlinna[20], Zulehner[22] it is shown that the damping factor $\frac{1}{2}$ can be replaced by an arbitrary damping factor $\omega \in (0, 1)$. Also note that in the smoothing property in corollary 7.13 we have a ν-dependence of the form $\nu^{-\frac{1}{2}}$, whereas in the symmetric case this is of the form ν^{-1}. It Hackbusch[21] it is shown that this loss of a factor $\nu^{\frac{1}{2}}$ when going to the nonsymmetric case is due to the fact that complex eigenvalues may occur. □

To verify the condition in (60) we will use the following elementary result:

Lemma 7.15 *If $\mathbf{E} \in \mathbb{R}^{m\times m}$ is such that there exists a $c > 0$ with*

$$\|\mathbf{E}\mathbf{x}\|_2^2 \le c\langle \mathbf{E}\mathbf{x}, \mathbf{x}\rangle \quad \textit{for all } \mathbf{x} \in \mathbb{R}^m$$

then we have $\|\mathbf{I} - \omega\mathbf{E}\|_2 \le 1$ for all $\omega \in [0, \frac{2}{c}]$.
Proof: Follows from:

$$\begin{aligned}\|(\mathbf{I} - \omega\mathbf{E})\mathbf{x}\|_2^2 &= \|\mathbf{x}\|_2^2 - 2\omega\langle \mathbf{E}\mathbf{x}, \mathbf{x}\rangle + \omega^2\|\mathbf{E}\mathbf{x}\|_2^2 \\ &\le \|\mathbf{x}\|_2^2 - \omega(\frac{2}{c} - \omega)\|\mathbf{E}\mathbf{x}\|_2^2 \\ &\le \|\mathbf{x}\|_2^2 \quad \text{if } \omega(\frac{2}{c} - \omega) \ge 0.\end{aligned}$$

■

We now use these results to derive a smoothing property for the Richardson method.
Theorem 7.16 *Assume that the bilinear form satisfies the usual conditions* (22)*. Let $\mathbf{A}_\ell$ be the stiffness matrix in* (26)*. There exist constants $\omega > 0$ and c independent of ℓ such that the following smoothing property holds:*

$$\|\mathbf{A}_\ell(\mathbf{I} - \omega h_\ell^{2-d}\mathbf{A}_\ell)^\nu\|_2 \le \frac{c}{\sqrt{\nu}}\|\mathbf{A}_\ell\|_2\,, \quad \nu = 1, 2, \ldots.$$

Proof: Using lemma 7.3, the inverse inequality and the ellipticity of the bilinear form we get, for arbitrary $\mathbf{x} \in \mathbb{R}^{n_\ell}$:

$$\begin{aligned}\|\mathbf{A}_\ell\mathbf{x}\|_2 &= \max_{\mathbf{y}\in\mathbb{R}^{n_\ell}} \frac{\langle \mathbf{A}_\ell\mathbf{x}, \mathbf{y}\rangle}{\|\mathbf{y}\|_2} \le c\, h_\ell^{\frac{1}{2}d} \max_{v_\ell \in V_\ell} \frac{k(P_\ell\mathbf{x}, v_\ell)}{\|v_\ell\|_{L^2}} \\ &\le c\, h_\ell^{\frac{1}{2}d} \max_{v_\ell \in V_\ell} \frac{|P_\ell\mathbf{x}|_1|v_\ell|_1}{\|v_\ell\|_{L^2}} \le c\, h_\ell^{\frac{1}{2}d-1}|P_\ell\mathbf{x}|_1 \\ &\le c\, h_\ell^{\frac{1}{2}d-1}k(P_\ell\mathbf{x}, P_\ell\mathbf{x})^{\frac{1}{2}} = c\, h_\ell^{\frac{1}{2}d-1}\langle \mathbf{A}_\ell\mathbf{x}, \mathbf{x}\rangle^{\frac{1}{2}}.\end{aligned}$$

From this and lemma 7.15 it follows that there exists a constant $\omega > 0$ such that

$$\|\mathbf{I} - 2\omega h_\ell^{2-d}\mathbf{A}_\ell\|_2 \le 1 \quad \text{for all } \ell. \tag{61}$$

Define $\mathbf{M}_\ell := \frac{1}{2\omega}h_\ell^{d-2}\mathbf{I}$. From lemma 7.4 it follows that there exists a constant c_M independent of ℓ such that $\|\mathbf{M}_\ell\|_2 \le c_M\|\mathbf{A}_\ell\|_2$. Application of corollary 7.13 proves the result of the lemma. ■

We now consider the damped Jacobi method.
Theorem 7.17 *Assume that the bilinear form satisfies the usual conditions* (22)*. Let $\mathbf{A}_\ell$ be the stiffness matrix in* (26) *and $\mathbf{D}_\ell = \operatorname{diag}(\mathbf{A}_\ell)$. There exist constants $\omega > 0$ and c independent of ℓ such that the following smoothing property holds:*

$$\|\mathbf{A}_\ell(\mathbf{I} - \omega\mathbf{D}_\ell^{-1}\mathbf{A}_\ell)^\nu\|_2 \le \frac{c}{\sqrt{\nu}}\|\mathbf{A}_\ell\|_2\,, \quad \nu = 1, 2, \ldots$$

Proof: We use the matrix norm induced by the vector norm $\|\mathbf{y}\|_D := \|\mathbf{D}_\ell^{\frac{1}{2}}\mathbf{y}\|_2$ for $\mathbf{y} \in \mathbb{R}^{n_\ell}$. Note that for $\mathbf{B} \in \mathbb{R}^{n_\ell\times n_\ell}$ we have $\|\mathbf{B}\|_D = \|\mathbf{D}_\ell^{\frac{1}{2}}\mathbf{B}\mathbf{D}_\ell^{-\frac{1}{2}}\|_2$. The inequalities

$$\|\mathbf{D}_\ell^{-1}\|_2 \le c_1\, h_\ell^{2-d}, \quad \kappa(\mathbf{D}_\ell) \le c_2 \tag{62}$$

hold with constants c_1, c_2 independent of ℓ. Using this in combination with lemma 7.3, the inverse inequality and the ellipticity of the bilinear form we get, for arbitrary $\mathbf{x} \in \mathbb{R}^{n_\ell}$:

$$\begin{aligned}
\|\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{A}_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x}\|_2 &= \max_{\mathbf{y}\in\mathbb{R}^{n_\ell}} \frac{\langle \mathbf{A}_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x}, \mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{y}\rangle}{\|\mathbf{y}\|_2} = \max_{\mathbf{y}\in\mathbb{R}^{n_\ell}} \frac{k(P_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x}, P_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{y})}{\|\mathbf{y}\|_2} \\
&\le c\, h_\ell^{-1} \max_{\mathbf{y}\in\mathbb{R}^{n_\ell}} \frac{|P_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x}|_1 \|P_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{y}\|_{L^2}}{\|\mathbf{y}\|_2} \\
&\le c\, h_\ell^{\frac{1}{2}d-1} |P_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x}|_1 \|\mathbf{D}_\ell^{-\frac{1}{2}}\|_2 \le c\, |P_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x}|_1 \\
&\le c\, k(P_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x}, P_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x})^{\frac{1}{2}} = c\, \langle \mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{A}_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x}, \mathbf{x}\rangle^{\frac{1}{2}}.
\end{aligned}$$

From this and lemma 7.15 it follows that there exists a constant $\omega > 0$ such that

$$\|\mathbf{I} - 2\omega\mathbf{D}_\ell^{-1}\mathbf{A}_\ell\|_D = \|\mathbf{I} - 2\omega\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{A}_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\|_2 \le 1 \quad \text{for all } \ell.$$

Define $\mathbf{M}_\ell := \frac{1}{2\omega}\mathbf{D}_\ell$. Application of corollary 7.13 with $\|\cdot\| = \|\cdot\|_D$ in combination with (62) yields

$$\begin{aligned}
\|\mathbf{A}_\ell(\mathbf{I} - \omega h_\ell \mathbf{D}_\ell^{-1}\mathbf{A}_\ell)^\nu\|_2 &\le \kappa(\mathbf{D}_\ell^{\frac{1}{2}})\, \|\mathbf{A}_\ell(\mathbf{I} - \frac{1}{2}\mathbf{M}_\ell^{-1}\mathbf{A}_\ell)^\nu\|_D \\
&\le \frac{c}{\sqrt{\nu}}\|\mathbf{M}_\ell\|_D = \frac{c}{2\omega\sqrt{\nu}}\|\mathbf{D}_\ell\|_2 \le \frac{c}{\sqrt{\nu}}\|\mathbf{A}_\ell\|_2
\end{aligned}$$

and thus the result is proved. ∎

7.4 Multigrid contraction number

In this section we prove a bound for the contraction number in the Euclidean norm of the multigrid algorithm (31) with $\tau \ge 2$. We follow the analysis introduced by Hackbusch[1, 14]. Apart from the approximation and smoothing property that have been proved in the sections 7.2 and 7.3 we also need the following stability bound for the iteration matrix of the smoother:

$$\exists\, C_S \,:\, \|\mathbf{S}_\ell^\nu\|_2 \le C_S \quad \text{for all } \ell \text{ and } \nu. \tag{63}$$

Lemma 7.18 *Consider the Richardson method as in theorem 7.8 or theorem 7.16. In both cases* (63) *holds with* $C_S = 1$.
Proof: In the symmetric case (theorem 7.8) we have

$$\|\mathbf{S}_\ell\|_2 = \|\mathbf{I} - \frac{c_0}{\rho(\mathbf{A}_\ell)}\mathbf{A}_\ell\|_2 = \max_{\lambda\in\sigma(\mathbf{A}_\ell)} \big|1 - c_0\, \frac{\lambda}{\rho(\mathbf{A}_\ell)}\big| \le 1.$$

For the general case (theorem 7.16) we have, using (61):

$$\begin{aligned}
\|\mathbf{S}_\ell\|_2 &= \|\mathbf{I} - \omega h_\ell^{2-d}\mathbf{A}_\ell\|_2 = \|\frac{1}{2}\mathbf{I} + \frac{1}{2}(\mathbf{I} - 2\omega h_\ell^{2-d}\mathbf{A}_\ell)\|_2 \\
&\le \frac{1}{2} + \frac{1}{2}\|\mathbf{I} - 2\omega h_\ell^{2-d}\mathbf{A}_\ell\|_2 \le 1.
\end{aligned}$$

∎

Lemma 7.19 *Consider the damped Jacobi method as in theorem 7.9 or theorem 7.17. In both cases* (63) *holds.*
Proof: Both in the symmetric and nonsymmetric case we have

$$\|\mathbf{S}_\ell\|_D = \|\mathbf{D}_\ell^{\frac{1}{2}}(\mathbf{I} - \omega\mathbf{D}_\ell^{-1}\mathbf{A}_\ell)\mathbf{D}_\ell^{-\frac{1}{2}}\|_2 \leq 1$$

and thus

$$\|\mathbf{S}_\ell^\nu\|_2 \leq \|\mathbf{D}_\ell^{-\frac{1}{2}}(\mathbf{D}_\ell^{\frac{1}{2}}\mathbf{S}_\ell\mathbf{D}_\ell^{-\frac{1}{2}})^\nu\mathbf{D}_\ell^{\frac{1}{2}}\|_2 \leq \kappa(\mathbf{D}_\ell^{\frac{1}{2}})\,\|\mathbf{S}_\ell\|_D^\nu \leq \kappa(\mathbf{D}_\ell^{\frac{1}{2}})$$

Now note that $\mathbf{D}_\ell$ is uniformly (w.r.t. ℓ) well-conditioned. ∎

Using lemma 7.3 it follows that for $\mathbf{p}_\ell = P_\ell^{-1}P_{\ell-1}$ we have

$$C_{p,1}\|\mathbf{x}\|_2 \leq \|\mathbf{p}_\ell\mathbf{x}\|_2 \leq C_{p,2}\|\mathbf{x}\|_2 \quad \text{for all } \mathbf{x} \in \mathbb{R}^{n_{\ell-1}}. \tag{64}$$

with constants $C_{p,1} > 0$ and $C_{p,2}$ independent of ℓ.
We now formulate a main convergence result for the multigrid method.

Theorem 7.20 *Consider the multigrid method with iteration matrix given in* (49) *and parameter values $\nu_2 = 0$, $\nu_1 = \nu > 0$, $\tau \geq 2$. Assume that there are constants C_A, C_S and a monotonically decreasing function $g(\nu)$ with $g(\nu) \to 0$ for $\nu \to \infty$ such that for all ℓ:*

$$\|\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\|_2 \leq C_A\|\mathbf{A}_\ell\|_2^{-1} \tag{65a}$$

$$\|\mathbf{A}_\ell\mathbf{S}_\ell^\nu\|_2 \leq g(\nu)\,\|\mathbf{A}_\ell\|_2\,, \quad \nu \geq 1 \tag{65b}$$

$$\|\mathbf{S}_\ell^\nu\|_2 \leq C_S\,, \quad \nu \geq 1. \tag{65c}$$

For any $\xi^* \in (0,1)$ there exists a ν^* such that for all $\nu \geq \nu^*$

$$\|\mathbf{C}_{MG,\ell}\|_2 \leq \xi^*\,, \quad \ell = 0, 1, \ldots$$

holds.

Proof: For the two-grid iteration matrix we have

$$\|\mathbf{C}_{TG,\ell}\|_2 \leq \|\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\|_2\|\mathbf{A}_\ell\mathbf{S}_\ell^\nu\|_2 \leq C_A g(\nu).$$

Define $\xi_\ell = \|\mathbf{C}_{MG.\ell}\|_2$. From (49) we obtain $\xi_0 = 0$ and for $\ell \geq 1$:

$$\begin{aligned}
\xi_\ell &\leq C_A g(\nu) + \|\mathbf{p}_\ell\|_2\xi_{\ell-1}^\tau\|\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\mathbf{S}_\ell^\nu\|_2 \\
&\leq C_A g(\nu) + C_{p,2}C_{p,1}^{-1}\xi_{\ell-1}^\tau\|\mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\mathbf{S}_\ell^\nu\|_2 \\
&\leq C_A g(\nu) + C_{p,2}C_{p,1}^{-1}\xi_{\ell-1}^\tau\big(\|(\mathbf{I} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell)\mathbf{S}_\ell^\nu\|_2 + \|\mathbf{S}_\ell^\nu\|_2\big) \\
&\leq C_A g(\nu) + C_{p,2}C_{p,1}^{-1}\xi_{\ell-1}^\tau\big(C_A g(\nu) + C_S\big) \leq C_A g(\nu) + C^*\xi_{\ell-1}^\tau
\end{aligned}$$

with $C^* := C_{p,2}C_{p,1}^{-1}(C_A g(1) + C_S)$. Elementary analysis shows that for $\tau \geq 2$ and any $\xi^* \in (0,1)$ the sequence $x_0 = 0,\ x_i = C_A g(\nu) + C^* x_{i-1}^\tau,\ i \geq 1$, is bounded by ξ^* for $g(\nu)$ sufficiently small. ∎

Remark 7.21 Consider $\mathbf{A}_\ell,\ \mathbf{p}_\ell,\ \mathbf{r}_\ell$ as defined in (26), (29),(30). Assume that the variational problem (23) is such that the usual conditions (22) are satisfied. Moreover, the problem (23) and the corresponding dual problem are assumed to be H^2-regular. In the

multigrid method we use the Richardson or the damped Jacobi method described in section 7.3. *Then the assumptions* (65) *are fulfilled and thus for* $\nu_2 = 0$ *and* ν_1 *sufficiently large the multigrid W-cylce has a contractrion number smaller than one indpendent of* ℓ. □

Remark 7.22 Let $\mathbf{C}_{MG,\ell}(\nu_2, \nu_1)$ be the iteration matrix of the multigrid method with ν_1 pre- and ν_2 postsmoothing iterations. With $\nu := \nu_1 + \nu_2$ we have

$$\rho\big(\mathbf{C}_{MG,\ell}(\nu_2, \nu_1)\big) = \rho\big(\mathbf{C}_{MG,\ell}(0, \nu)\big) \leq \|\mathbf{C}_{MG,\ell}(0, \nu)\|_2$$

Using theorem 7.20 we thus get, for $\tau \geq 2$, a bound for the *spectral radius* of the iteration matrix $\mathbf{C}_{MG,\ell}(\nu_2, \nu_1)$. □

Remark 7.23 The multigrid convergence analysis presented above assumes sufficient regularity (namely H^2-regularity) of the elliptic boundary value problem. There have been developed convergence analyses in which this regularity assumption is avoided and an h-independent convergence rate of multigrid is proved. These analyses are based on so-called subspace decomposition techniques. Two review papers on multigrid convergence proofs are Yserentant[23] and Xu[24]. □

7.5 Convergence analysis for symmetric positive definite problems

In this section we analyze the convergence of the multigrid method for the symmetric positive definite case, i.e., the stiffness matrix $\mathbf{A}_\ell$ is assumed to be symmetric positive definite. This property allows a refined analysis which proves that the contraction number of the multigrid method with $\tau \geq 1$ (the V-cycle is included !) and $\nu_1 = \nu_2 \geq 1$ pre- and postsmoothing iterations is bounded by a constant smaller than one independent of ℓ. The basic idea of this analysis is due to Braess[25] and is further simplified by Hackbusch[1, 14].

Throughout this section we make the following

Assumption 7.24 *In the bilinear form* $k(\cdot, \cdot)$ *in* (23) *we have* $\mathbf{b} = 0$ *and the conditions* (22) *are satisfied.*

Due to this the stiffness matrix $\mathbf{A}_\ell$ is symmetric positive definite and we can define the energy scalar product and corresponding norm:

$$\langle \mathbf{x}, \mathbf{y} \rangle_A := \langle \mathbf{A}_\ell \mathbf{x}, \mathbf{y} \rangle \ , \quad \|\mathbf{x}\|_A := \langle \mathbf{x}, \mathbf{x} \rangle_A^{\frac{1}{2}} \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^{n_\ell}.$$

We only consider smoothers with an iteration matrix $\mathbf{S}_\ell = \mathbf{I} - \mathbf{M}_\ell^{-1} \mathbf{A}_\ell$ in which $\mathbf{M}_\ell$ is symmetric positive definite. Important examples are the smoothers analyzed in section 7.3:

$$\text{Richardson method : } \ \mathbf{M}_\ell = c_0^{-1} \rho(\mathbf{A}_\ell) \mathbf{I} \ , \quad c_0 \in (0, 1] \tag{66a}$$

$$\text{Damped Jacobi : } \ \mathbf{M}_\ell = \omega^{-1} \mathbf{D}_\ell, \quad \omega \ \text{ as in thm. 7.9} \tag{66b}$$

$$\text{Symm. Gauss-Seidel : } \ \mathbf{M}_\ell = (\mathbf{D}_\ell - \mathbf{L}_\ell) \mathbf{D}_\ell^{-1} (\mathbf{D}_\ell - \mathbf{L}_\ell^T). \tag{66c}$$

For symmetric matrices $\mathbf{B}, \mathbf{C} \in \mathbb{R}^{m \times m}$ we use the notation $\mathbf{B} \leq \mathbf{C}$ iff $\langle \mathbf{B}\mathbf{x}, \mathbf{x} \rangle \leq \langle \mathbf{C}\mathbf{x}, \mathbf{x} \rangle$ for all $\mathbf{x} \in \mathbb{R}^m$.

Lemma 7.25 *For* $\mathbf{M}_\ell$ *as in* (66) *the following properties hold:*

$$\mathbf{A}_\ell \leq \mathbf{M}_\ell \quad \textit{for all } \ell \tag{67a}$$

$$\exists\, C_M : \quad \|\mathbf{M}_\ell\|_2 \leq C_M \|\mathbf{A}_\ell\|_2 \quad \textit{for all } \ell. \tag{67b}$$

Proof: For the Richardson method the result is trivial. For the damped Jacobi method we have $\omega \in (0, \rho(\mathbf{D}_\ell^{-1}\mathbf{A}_\ell)^{-1}]$ and thus $\omega\rho(\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{A}_\ell\mathbf{D}_\ell^{-\frac{1}{2}}) \leq 1$. This yields $\mathbf{A}_\ell \leq \omega^{-1}\mathbf{D}_\ell = \mathbf{M}_\ell$. The result in (67b) follows from $\|\mathbf{D}_\ell\|_2 \leq \|\mathbf{A}_\ell\|_2$. For the symmetric Gauss-Seidel method the results (67a) follows from $\mathbf{M}_\ell = \mathbf{A}_\ell + \mathbf{L}_\ell\mathbf{D}_\ell^{-1}\mathbf{L}_\ell^T$ and the result in (67b) is proved in (59). ∎

We introduce the following *modified approximation property*:

$$\exists\, \tilde{C}_A: \quad \|\mathbf{M}_\ell^{\frac{1}{2}}\big(\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\big)\mathbf{M}_\ell^{\frac{1}{2}}\|_2 \leq \tilde{C}_A \quad \text{for } \ell = 1, 2, \ldots \tag{68}$$

We note that the standard approximation property (54) implies the result (68) if we consider the smoothers in (66):

Lemma 7.26 *Consider* $\mathbf{M}_\ell$ *as in* (66) *and assume that the approximation property* (54) *holds. Then* (68) *holds with* $\tilde{C}_A = C_M C_A$.

Proof: Trivial. ∎

One easily verifies that for the smoothers in (66) the modified approximation property (68) implies the standard approximation property (54) if $\kappa(\mathbf{M}_\ell)$ is uniformly (w.r.t. ℓ) bounded. The latter property holds for the Richardson and the damped Jacobi method.

We will analyze the convergence of the two-grid and multigrid method using the energy scalar product. For matrices $\mathbf{B},\ \mathbf{C} \in \mathbb{R}^{n_\ell \times n_\ell}$ that are symmetric w.r.t. $\langle\cdot,\cdot\rangle_A$ we use the notation $\mathbf{B} \leq_A \mathbf{C}$ iff $\langle\mathbf{B}\mathbf{x},\mathbf{x}\rangle_A \leq \langle\mathbf{C}\mathbf{x},\mathbf{x}\rangle_A$ for all $\mathbf{x} \in \mathbb{R}^{n_\ell}$. Note that $\mathbf{B} \in \mathbb{R}^{n_\ell \times n_\ell}$ is symmetric w.r.t. $\langle\cdot,\cdot\rangle_A$ iff $(\mathbf{A}_\ell\mathbf{B})^T = \mathbf{A}_\ell\mathbf{B}$ holds. We also note the following elementary property for symmetric matrices $\mathbf{B},\ \mathbf{C} \in \mathbb{R}^{n_\ell \times n_\ell}$:

$$\mathbf{B} \leq \mathbf{C} \ \Leftrightarrow\ \mathbf{B}\mathbf{A}_\ell \leq_A \mathbf{C}\mathbf{A}_\ell. \tag{69}$$

We now turn to the two-grid method. For the coarse grid correction we introduce the notation $\mathbf{Q}_\ell := \mathbf{I} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell$. For symmetry reasons we only consider $\nu_1 = \nu_2 = \frac{1}{2}\nu$ with $\nu > 0$ even. The iteration matrix of the two-grid method is given by

$$\mathbf{C}_{TG,\ell} = \mathbf{C}_{TG,\ell}(\nu) = \mathbf{S}_\ell^{\frac{1}{2}\nu}\mathbf{Q}_\ell\mathbf{S}_\ell^{\frac{1}{2}\nu}.$$

Due the symmetric positive definite setting we have the following fundamental property:

Theorem 7.27 *The matrix* $\mathbf{Q}_\ell$ *is an orthogonal projection w.r.t.* $\langle\cdot,\cdot\rangle_A$.

Proof: Follows from

$$\mathbf{Q}_\ell^2 = \mathbf{Q}_\ell \quad \text{and} \quad (\mathbf{A}_\ell\mathbf{Q}_\ell)^T = \mathbf{A}_\ell\mathbf{Q}_\ell.$$

∎

As an direct consequence we have

$$0 \leq_A \mathbf{Q}_\ell \leq_A \mathbf{I}. \tag{70}$$

The next lemma gives another characterization of the modified approximation property:

Lemma 7.28 *The property* (68) *is equivalent to*

$$0 \leq_A \mathbf{Q}_\ell \leq_A \tilde{C}_A\mathbf{M}_\ell^{-1}\mathbf{A}_\ell \quad \textit{for } \ell = 1, 2, \ldots \tag{71}$$

Proof: Using (69) we get

$$\begin{aligned}
&\|\mathbf{M}_\ell^{\frac{1}{2}}\big(\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\big)\mathbf{M}_\ell^{\frac{1}{2}}\|_2 \le \tilde{C}_A \quad \text{for all } \ell \\
\Leftrightarrow\ & -\tilde{C}_A\mathbf{I} \le \mathbf{M}_\ell^{\frac{1}{2}}\big(\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\big)\mathbf{M}_\ell^{\frac{1}{2}} \le \tilde{C}_A\mathbf{I} \quad \text{for all } \ell \\
\Leftrightarrow\ & -\tilde{C}_A\mathbf{M}_\ell^{-1} \le \mathbf{A}_\ell^{-1} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell \le \tilde{C}_A\mathbf{M}_\ell^{-1} \quad \text{for all } \ell \\
\Leftrightarrow\ & -\tilde{C}_A\mathbf{M}_\ell^{-1}\mathbf{A}_\ell \le_A \mathbf{Q}_\ell \le_A \tilde{C}_A\mathbf{M}_\ell^{-1}\mathbf{A}_\ell \quad \text{for all } \ell.
\end{aligned}$$

In combination with (70) this proves the result. ∎

We now present a convergence result for the two-grid method:

Theorem 7.29 *Assume that* (67a) *and* (68) *hold. Then we have*

$$\begin{aligned}
\|\mathbf{C}_{TG,\ell}(\nu)\|_A &\le \max_{y\in[0,1]} y(1-\tilde{C}_A^{-1}y)^\nu \\
&= \begin{cases} (1-\tilde{C}_A^{-1})^\nu & \text{if } \ \nu \le \tilde{C}_A - 1 \\ \frac{\tilde{C}_A}{\nu+1}\big(\frac{\nu}{\nu+1}\big)^\nu & \text{if } \ \nu \ge \tilde{C}_A - 1. \end{cases}
\end{aligned} \tag{72}$$

Proof: Define $\mathbf{X}_\ell := \mathbf{M}_\ell^{-1}\mathbf{A}_\ell$. This matrix is symmetric w.r.t. the energy scalar product and from (67a) it follows that

$$0 \le_A \mathbf{X}_\ell \le_A \mathbf{I} \tag{73}$$

holds. From lemma 7.28 we obtain $0 \le_A \mathbf{Q}_\ell \le_A \tilde{C}_A\mathbf{X}_\ell$. Note that due to this, (73) and the fact that $\mathbf{Q}_\ell$ is an A-orthogonal projection which is not identically zero we get

$$\tilde{C}_A \ge 1. \tag{74}$$

Using (70) we get

$$0 \le_A \mathbf{Q}_\ell \le_A \alpha\tilde{C}_A\mathbf{X}_\ell + (1-\alpha)\mathbf{I} \quad \text{for all } \ \alpha \in [0,1]. \tag{75}$$

Hence, using $\mathbf{S}_\ell = \mathbf{I} - \mathbf{X}_\ell$ we have

$$0 \le_A \mathbf{C}_{TG,\ell}(\nu) \le_A (\mathbf{I}-\mathbf{X}_\ell)^{\frac{1}{2}\nu}\big(\alpha\tilde{C}_A\mathbf{X}_\ell + (1-\alpha)\mathbf{I}\big)(\mathbf{I}-\mathbf{X}_\ell)^{\frac{1}{2}\nu}$$

for all $\alpha \in [0,1]$, and thus

$$\|\mathbf{C}_{TG,\ell}(\nu)\|_A \le \min_{\alpha\in[0,1]}\max_{x\in[0,1]}\big(\alpha\tilde{C}_A x + (1-\alpha)\big)(1-x)^\nu.$$

A minimax result (cf. Sion[26]) implies that in the previous expression the min and max operations can be interchanged. A simple computation yields

$$\begin{aligned}
&\max_{x\in[0,1]}\min_{\alpha\in[0,1]}\big(\alpha\tilde{C}_A x + (1-\alpha)\big)(1-x)^\nu \\
&= \max\Big\{\max_{x\in[0,\tilde{C}_A^{-1}]}\tilde{C}_A x(1-x)^\nu\,,\ \max_{x\in[\tilde{C}_A^{-1},1]}(1-x)^\nu\Big\} \\
&= \max_{x\in[0,\tilde{C}_A^{-1}]}\tilde{C}_A x(1-x)^\nu = \max_{y\in[0,1]} y(1-\tilde{C}_A^{-1}y)^\nu.
\end{aligned}$$

This proves the inequality in (72). An elementary computation shows that the equality in (72) holds. ∎

We now show that the approach used in the convergence analysis of the two-grid method in theorem 7.29 can also be used for the multigrid method.
We start with an elementary result concerning a fixed point iteration that will be used in theorem 7.31.

Lemma 7.30 *For given constants $c > 1$, $\nu \geq 1$ define $g : [0,1) \to \mathbb{R}$ by*

$$g(\xi) = \begin{cases} (1 - \frac{1}{c})^{\nu} & \text{if } \ 0 \leq \xi < 1 - \frac{\nu}{c-1} \\ \frac{c}{\nu+1}\big(\frac{\nu}{\nu+1}\big)^{\nu}(1-\xi)\big(1 + \frac{1}{c}\frac{\xi}{1-\xi}\big)^{\nu+1} & \text{if } \ 1 - \frac{\nu}{c-1} \leq \xi < 1. \end{cases} \tag{76}$$

For $\tau \in \mathbb{N}$, $\tau \geq 1$, define the sequence $\xi_{\tau,0} = 0, \quad \xi_{\tau,i+1} = g(\xi_{\tau,i}^{\tau})$ for $i \geq 1$. The following holds:

- *$\xi \to g(\xi)$ is continuous and increasing on $[0,1)$.*
- *For $c = \tilde{C}_A$, $g(0)$ coincides with the upper bound in* (72)*.*
- *$g(\xi) = \xi$ iff $\xi = \dfrac{c}{c+\nu}$.*
- *The sequence $(\xi_{\tau,i})_{i \geq 0}$ is monotonically increasing, and $\xi_{\tau}^{*} := \lim\limits_{i \to \infty} \xi_{\tau,i} < 1$.*
- *$\big((\xi_{\tau}^{*})^{\tau}, \xi_{\tau}^{*}\big)$ is the first intersection point of the graphs of $g(\xi)$ and $\xi^{\frac{1}{\tau}}$.*
- *$\dfrac{c}{c+\nu} = \xi_1^{*} \geq \xi_2^{*} \geq \ldots \geq \xi_{\infty}^{*} = g(0)$.*

Proof: Elementary calculus. ∎

As an illustration for two pairs (c, ν) we show the graph of the function g in Fig. 9.

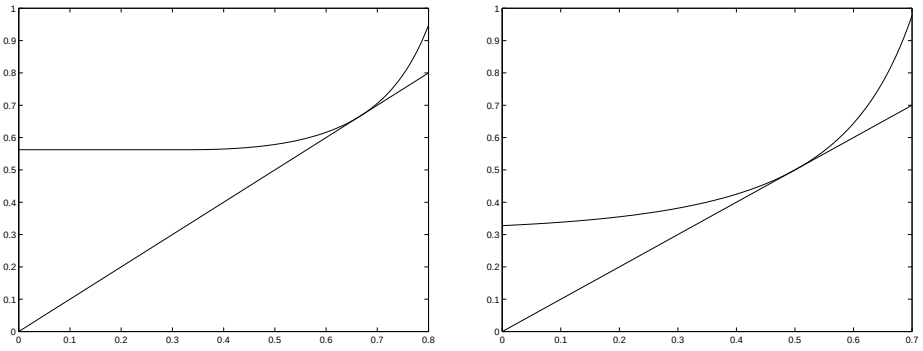


Figure 9. Function $g(\xi)$ for $\nu = 2,\ c = 4$ (left) and $\nu = 4,\ c = 4$ (right).

Theorem 7.31 *We take $\nu_1 = \nu_2 = \nu$ and consider the multigrid algorithm with iteration matrix $\mathbf{C}_{MG,\ell} = \mathbf{C}_{MG,\ell}(\nu,\tau)$ as in* (49)*. Assume that* (67a) *and* (68) *hold. For $c = \tilde{C}_A$, $\nu \geq 2$ and τ as in* (49) *let $\xi_\tau^* \leq \frac{c}{c+\nu}$ be the fixed point defined in lemma 7.30. Then*

$$\|\mathbf{C}_{MG,\ell}\|_A \leq \xi_\tau^*$$

holds.

Proof: From (49) we have

$$\begin{aligned}\mathbf{C}_{MG,\ell} &= \mathbf{S}_\ell^{\frac{1}{2}\nu}\big(\mathbf{I} - \mathbf{p}_\ell(\mathbf{I} - \mathbf{C}_{MG,\ell-1}^\tau)\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\big)\mathbf{S}_\ell^{\frac{1}{2}\nu} \\ &= \mathbf{S}_\ell^{\frac{1}{2}\nu}(\mathbf{Q}_\ell + \mathbf{R}_\ell)\mathbf{S}_\ell^{\frac{1}{2}\nu}\,, \quad \mathbf{R}_\ell := \mathbf{p}_\ell\mathbf{C}_{MG,\ell-1}^\tau\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell.\end{aligned}$$

The matrices $\mathbf{S}_\ell$ and $\mathbf{Q}_\ell$ are symmetric w.r.t. $\langle\cdot,\cdot\rangle_A$. If $\mathbf{C}_{MG,\ell-1}$ is symmetric w.r.t. $\langle\cdot,\cdot\rangle_{A_{\ell-1}}$ then from

$$(\mathbf{A}_\ell\mathbf{R}_\ell)^T = \big[(\mathbf{A}_\ell\mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1})(\mathbf{A}_{\ell-1}\mathbf{C}_{MG,\ell-1}^\tau)(\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell)\big]^T = \mathbf{A}_\ell\mathbf{R}_\ell$$

it follows that $\mathbf{R}_\ell$ is symmetric w.r.t. $\langle\cdot,\cdot\rangle_A$, too. By induction we conclude that for all ℓ the matrices $\mathbf{R}_\ell$ and $\mathbf{C}_{MG,\ell}$ are symmetric w.r.t. $\langle\cdot,\cdot\rangle_A$. Note that

$$0 \leq_A \mathbf{C}_{MG,\ell-1}^\tau \;\Leftrightarrow\; 0 \leq \mathbf{C}_{MG,\ell-1}^\tau\mathbf{A}_{\ell-1}^{-1} \;\Leftrightarrow\; 0 \leq \mathbf{p}_\ell\mathbf{C}_{MG,\ell-1}^\tau\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell \;\Leftrightarrow\; 0 \leq_A \mathbf{R}_\ell$$

holds. Thus, by induction and using $0 \leq_A \mathbf{Q}_\ell$ we get

$$0 \leq_A \mathbf{Q}_\ell + \mathbf{R}_\ell\,, \quad 0 \leq_A \mathbf{C}_{MG,\ell} \quad \text{for all } \ell. \tag{77}$$

For $\ell \geq 0$ define $\xi_\ell := \|\mathbf{C}_{MG,\ell}\|_A$. Hence, $0 \leq_A \mathbf{C}_{MG,\ell} \leq_A \xi_\ell\mathbf{I}$ holds. For arbitrary $\mathbf{x} \in \mathbb{R}^{n_\ell}$ we have

$$\begin{aligned}\langle\mathbf{R}_\ell\mathbf{x},\mathbf{x}\rangle_A &= \langle\mathbf{C}_{MG,\ell-1}^\tau\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\mathbf{x}, \mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\mathbf{x}\rangle_{A_{\ell-1}} \\ &\leq \xi_{\ell-1}^\tau\langle\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\mathbf{x}, \mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\mathbf{x}\rangle_{A_{\ell-1}} = \xi_{\ell-1}^\tau\langle\mathbf{x},(\mathbf{I}-\mathbf{Q}_\ell)\mathbf{x}\rangle_A\end{aligned}$$

and thus

$$\mathbf{R}_\ell \leq_A \xi_{\ell-1}^\tau(\mathbf{I} - \mathbf{Q}_\ell) \tag{78}$$

holds. Define $\mathbf{X}_\ell := \mathbf{M}_\ell^{-1}\mathbf{A}_\ell$. Using (75), (77) and (78) we get

$$\begin{aligned}0 \leq_A \mathbf{Q}_\ell + \mathbf{R}_\ell &\leq_A (1-\xi_{\ell-1}^\tau)\mathbf{Q}_\ell + \xi_{\ell-1}^\tau\mathbf{I} \\ &\leq_A (1-\xi_{\ell-1}^\tau)\big(\alpha\tilde{C}_A\mathbf{X}_\ell + (1-\alpha)\mathbf{I}\big) + \xi_{\ell-1}^\tau\mathbf{I} \quad \text{for all } \alpha \in [0,1].\end{aligned}$$

Hence, for all $\alpha \in [0,1]$ we have

$$0 \leq_A \mathbf{C}_{MG,\ell} \leq_A (\mathbf{I}-\mathbf{X}_\ell)^{\frac{1}{2}\nu}\big[(1-\xi_{\ell-1}^\tau)\big(\alpha\tilde{C}_A\mathbf{X}_\ell + (1-\alpha)\mathbf{I}\big) + \xi_{\ell-1}^\tau\mathbf{I}\big](\mathbf{I}-\mathbf{X}_\ell)^{\frac{1}{2}\nu}.$$

This yields

$$\xi_\ell \leq \min_{\alpha\in[0,1]}\max_{x\in[0,1]}\big[(1-\xi_{\ell-1}^\tau)\big(\alpha\tilde{C}_Ax + 1 - \alpha\big) + \xi_{\ell-1}^\tau\big](1-x)^\nu.$$

As in the proof of theorem 7.29 we can interchange the min and max operations in the previous expression. A simple computation shows that for $\xi \in [0,1]$ we have

$$\begin{aligned}&\max_{x\in[0,1]}\min_{\alpha\in[0,1]}\big[(1-\xi)\big(\alpha\tilde{C}_Ax + 1 - \alpha\big) + \xi\big](1-x)^\nu \\ &= \max\Big\{\max_{x\in[0,\tilde{C}_A^{-1}]}\big((1-\xi)\tilde{C}_Ax + \xi\big)(1-x)^\nu\,,\ \max_{x\in[\tilde{C}_A^{-1},1]}(1-x)^\nu\Big\} = g(\xi)\end{aligned}$$

where $g(\xi)$ is the function defined in lemma 7.30 with $c = \tilde{C}_A$. Thus ξ_ℓ satisfies $\xi_0 = 0$ and $\xi_\ell \leq g(\xi_{\ell-1}^\tau)$ for $\ell \geq 1$. Application of the results in lemma 7.30 completes the proof. ■

The bound ξ_τ^* for the multigrid contraction number in theorem 7.31 decreases if τ increases. Moreover, for $\tau \to \infty$ the bound converges to the bound for the two-grid contraction number in theorem 7.29.

Corollary 7.32 *Consider* $\mathbf{A}_\ell,\ \mathbf{p}_\ell,\ \mathbf{r}_\ell$ *as defined in* (26), (29),(30)*. Assume that the variational problem* (23) *is such that* $\mathbf{b} = 0$ *and that the usual conditions* (22) *are satisfied. Moreover, the problem is assumed to be* H^2*-regular. In the multigrid method we use one of the smoothers* (66)*. Then the assumptions* (67a) *and* (68) *are satisfied and thus for* $\nu_1 = \nu_2 \geq 1$ *the multigrid V-cycle has a contraction number (w.r.t.* $\|\cdot\|_A$*) smaller than one independent of* ℓ*.* □

8 Convergence analysis for Stokes problems

The multigrid method for the Stokes problem can be analyzed along the same lines as in section 7.4, i.e., based on a smoothing and approximation property. For the Stokes problem an analysis which proves convergence of the V-cycle is *not* known. In other words, results as presented for scalar elliptic problems in section 7.5 are not known for the Stokes equation.

We briefly outline the convergence results available for multigrid applied to the Stokes problem. For a detailed treatment we refer to the literature, for example to Verfürth[27], Larin[28], Zulehner[11]. As in section 7 we assume that the family of triangulations $\{\mathcal{T}_{h_\ell}\}$ is quasi-uniform and that $h_{\ell-1}/h_\ell$ is uniformly bounded w.r.t. ℓ. We assume H^2-regularity of the Stokes problem, i.e., for the solution $(\vec{u}, p)$ of (38) we have

$$\|\vec{u}\|_{H^2} + \|p\|_{H^1} \leq c\|\vec{f}\|_{L^2}$$

with a constant c independent of $\vec{f} \in L^2(\Omega)^d$. The finite element spaces $\mathbf{V}_\ell,\ Q_\ell$ should have the approximation property

$$\inf_{\vec{v}\in\mathbf{V}_\ell} \|\vec{u} - \vec{v}\|_{H^1} + \inf_{q\in Q_\ell} \|p - q\|_{L^2} \leq c\, h_\ell \big(\|\vec{u}\|_{H^2} + \|p\|_{H^1}\big),$$

for all $\vec{u} \in (H^2(\Omega) \cap H_0^1(\Omega))^d,\ p \in H^1(\Omega) \cap L_0^2(\Omega)$. This holds, for example, for the Hood-Taylor pair of finite element spaces. Let $\mathcal{A}_\ell$ be the Stokes stiffness matrix as in (40) and $\mathcal{S}_\ell$ the iteration matrix of the smoother. The prolongation P_ℓ is as in (41). For the restriction R_ℓ we take the adjoint of the prolongation. The iteration matrix of the two-grid method with $\nu = \nu_1$ pre-smoothing and $\nu_2 = 0$ post-smoothing iterations is given by

$$\mathcal{M}_\ell = (I - P_\ell \mathcal{A}_{\ell-1}^{-1} R_\ell \mathcal{A}_\ell)\mathcal{S}_\ell^\nu.$$

For the analysis we have to introduce a suitable scaled Euclidean norm defined by

$$\left\| \begin{pmatrix} \mathbf{u}_\ell \\ \mathbf{p}_\ell \end{pmatrix} \right\|_h^2 := \|\mathbf{u}_\ell\|^2 + h_\ell^2 \|\mathbf{p}_\ell\|^2 = \left\| \Lambda_\ell \begin{pmatrix} \mathbf{u}_\ell \\ \mathbf{p}_\ell \end{pmatrix} \right\|^2 \quad \text{with } \Lambda_\ell := \begin{pmatrix} I_{n_\ell} & 0 \\ 0 & h_\ell I_{m_\ell} \end{pmatrix}. \tag{79}$$

Furthermore we introduce the scaled matrices

$$\tilde{\mathcal{A}}_\ell := \Lambda_\ell^{-1} \mathcal{A}_\ell \Lambda_\ell^{-1} = \begin{pmatrix} A_\ell & h_\ell^{-1} B_\ell^T \\ h_\ell^{-1} B_\ell & 0 \end{pmatrix}, \quad \tilde{\mathcal{S}}_\ell := \Lambda_\ell \mathcal{S}_\ell \Lambda_\ell^{-1}.$$

Using these definitions we obtain

$$\begin{aligned}\|\mathcal{M}_\ell\|_h &= \|\Lambda_\ell(\mathcal{A}_\ell^{-1} - P_\ell\mathcal{A}_{\ell-1}^{-1}R_\ell)\Lambda_\ell\Lambda_\ell^{-1}\mathcal{A}_\ell\mathcal{S}_\ell^\nu\Lambda_\ell^{-1}\| \\ &\le \|\Lambda_\ell(\mathcal{A}_\ell^{-1} - P_\ell\mathcal{A}_{\ell-1}^{-1}R_\ell)\Lambda_\ell\|\,\|\tilde{\mathcal{A}}_\ell\tilde{\mathcal{S}}_\ell^\nu\|.\end{aligned}$$

In Larin[28] the *approximation property*

$$\|\Lambda_\ell(\mathcal{A}_\ell^{-1} - P_\ell\mathcal{A}_{\ell-1}^{-1}R_\ell)\Lambda_\ell\| \le ch_\ell^2 \tag{80}$$

is proved. In that paper it is also shown (using an analysis along the same lines as in section 7.3) that for the Braess-Sarazin method in which the system in (45) is solved exactly, we have a smoothing property

$$\|\tilde{\mathcal{A}}_\ell\tilde{\mathcal{S}}_\ell^\nu\| \le \frac{ch_\ell^{-2}}{e(\nu-2)+1} \quad \text{for } \nu \ge 2. \tag{81}$$

In Zulehner[11] a smoothing property for the Braess-Sarazin method with an *inexact* (but sufficiently accurate) inner solve for the system (45) is proved:

$$\|\tilde{\mathcal{A}}_\ell\tilde{\mathcal{S}}_\ell^\nu\| \le \frac{ch_\ell^{-2}}{\nu-1} \quad \text{for } \nu \ge 2. \tag{82}$$

Combining the approximation property in (80) with the smoothing property (81) or (82) we obtain a bound for the contraction number of the two-grid iteration matrix:

$$\|\mathcal{M}_\ell\|_h \le \frac{c_A}{\nu-1} \quad \text{for } \nu \ge 2$$

with a constant c_A independent of ℓ and ν. Thus we have a two-grid convergence with a rate independent of ℓ if the number of smoothing iterations ν is sufficiently high. Using an analysis as in section 7.4 one can derive a convergence result for the multigrid W-cycle method.

A smoothing property of the form (81), (82) for the Vanka smoother is *not* known in the literature. A theoretical analysis which proves convergence of the multigrid method with a Vanka smoother for the Stokes equations is not available.

References

1. Wolfgang Hackbusch, *Multigrid Methods and Applications*, vol. 4 of *Springer Series in Computational Mathematics*, Springer, Berlin, Heidelberg, 1985.
2. P. Wesseling, *An introduction to Multigrid Methods*, Wiley, Chichester, 1992.
3. W. L. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial (2nd ed.)*, SIAM, Philadelphia, 2000.
4. U. Trottenberg C. W. Oosterlee and A. Schüler, (Eds.), *Multigrid*, Academic Press, London, 2001.
5. J. H. Bramble, *Multigrid Methods*, Longman, Harlow, 1993.
6. Wolfgang Hackbusch, *Elliptic Differential Equations: Theory and Numerical Treatment*, vol. 18 of *Springer Series in Computational Mathematics*, Springer, Berlin, 1992.
7. Christian Großmann and H. G. Roos, *Numerik Partieller Differentialgleichungen*, Teubner, Stuttgart, 2. edition, 1994.

8. Jürgen Bey, *Tetrahedral Grid Refinement*, Computing, **55**, no. 4, 355–378, 1995.
9. Dietrich Braess, M. Dryja, and Wolfgang Hackbusch, *A multigrid method for nonconforming FE-discretisations with application to non-matching grids*, Computing, **63**, 1–25, 1999.
10. Dietrich Braess and R. Sarazin, *An efficient smoother for the Stokes problem*, Applied Numerical Mathematics, **23**, 3–19, 1997.
11. Walter Zulehner, *A class of smoothers for saddle point problems*, Computing, **65**, 227–246, 2000.
12. S. Vanka, *Block-implicit multigrid calculation of two-dimensional recirculating flows*, Computer Methods in Appl. Mech. and Eng., **59**, 29–48, 1986.
13. Volker John and G. Matthies, *Higher order finite element discretizations in a benchmark problem for incompressible flows*, Int. J. Num. Meth. Fluids, **37**, 885–903, 2001.
14. Wolfgang Hackbusch, *Iterative Solution of Large Sparse Systems of Equations*, vol. 95 of *Applied Mathematical Sciences*, Springer, New York, 1994.
15. Gabriel Wittum, *On the Robustness of ILU-Smoothing*, SIAM J. Sci. Stat. Comp., **10**, 699–717, 1989.
16. Gabriel Wittum, *Linear iterations as smoothers in multigrid methods : Theory with applications to incomplete decompositions*, Impact Comput. Sci. Eng., **1**, 180–215, 1989.
17. O. Bröker, M. Grote, C. Mayer, and A. Reusken, *Robust parallel smoothing for multigrid via sparse approximate inverses*, SIAM J. Sci. Comput., **32**, 1395–1416, 2001.
18. Arnold Reusken, *On maximum norm convergence of multigrid methods for two-point boundary value problems*, SIAM J. Numer. Anal., **29**, 1569–1578, 1992.
19. Arnold Reusken, *The smoothing property for regular splittings*, in: Incomplete Decompositions : (ILU)-Algorithms, Theory and Applications, W. Hackbusch and G. Wittum, (Eds.), vol. 41 of *Notes on Numerical Fluid Mechanics*, pp. 130–138, Vieweg, Braunschweig, 1993.
20. Olavi Nevanlinna, *Convergence of Iterations for Linear Equations*, Birkhäuser, Basel, 1993.
21. W. Hackbusch, *A note on Reusken's Lemma*, Computing, **55**, 181–189, 1995.
22. E. Ecker and W. Zulehner, *On the smoothing property of multi-grid methods in the non-symmetric case*, Numerical Linear Algebra with Applications, **3**, 161–172, 1996.
23. Harry Yserentant, *Old and new convergence proofs of multigrid methods*, Acta Numerica, pp. 285–326, 1993.
24. Jinchao Xu, *Iterative methods by space decomposition and subspace correction*, SIAM Review, **34**, 581–613, 1992.
25. Dietrich Braess and Wolfgang Hackbusch, *A New Convergence Proof for the Multigrid Method including the V-Cycle*, SIAM J. Numer. Anal., **20**, 967–975, 1983.
26. Maurice Sion, *On general minimax theorems*, Pacific J. of Math., **8**, 171–176, 1958.
27. Rüdiger Verfürth, *Error estimates for a mixed finite element approximation of Stokes problem*, RAIRO Anal. Numer., **18**, 175–182, 1984.
28. Maxim Larin and Arnold Reusken, *A comparative study of efficient iterative solvers for generalized Stokes equations*, Numer. Linear Algebra Appl., **15**, 13–34, 2008.