Cascadic Multigrid Methods

FOLKMAR A. BORNEMANN AND PETER DEUFLHARD

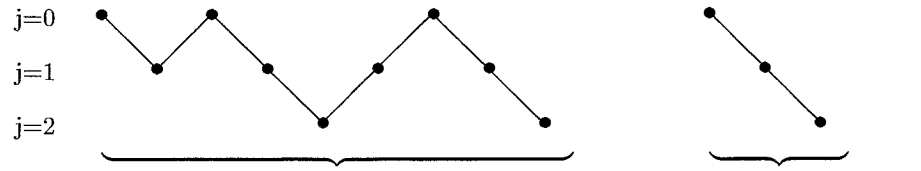
1 Introduction

In this contribution we consider linear scalar elliptic problems on general domains with space dimension $d \leq 3$. For the numerical solution of such problems with finite elements, *multigrid methods* are a both popular and efficient choice, cf. Hackbusch [Hac85]. Such methods work on a sequence of grid levels $j = 0, 1, \ldots \ell$, where in our notation $j = 0$ denotes the coarse grid level and $j > 0$ the refinement levels. For the case of *adaptive* grids, the local multigrid method suggested by [BPWX91] turned out to be the method of choice. In the adaptive setting it is favorable to consider *nested iterations*, wherein the computed solution on the previous level serves as starting point for the iteration on the new level. A typical distinction of different types of nested multigrid methods is made by the number p of correction cycles on each level: W-cycles are characterized by $p = 2$, V-cycles by $p = 1$.

For the first time the case $p = 0$, i.e., performing *no* coarse grid correction at all, was seriously considered by Deuflhard [Deu94]. He pointed out that the use of an a posteriori algorithmic control of this kind of nested iteration in combination with the conjugate gradient method gave reasonable results in practice. He called the method *cascadic conjugate gradient method.* As a distinctive feature this method performs *more iterations on coarser levels* so as to obtain less iterations on finer levels. Shaidurov [Sha94] as well as Bornemann and Deuflhard [BD96] proved accuracy and optimality of this approach with respect to the *energy norm.* The latter two authors could also show, that the conjugate gradient method can be replaced by any smoothing iteration like the traditional candidates SSOR or damped Jacobi. Because of this fact, they called this kind of iteration *cascadic multigrid methods.*

Domain Decomposition Methods in Sciences and Engineering, edited by R. Glowinski *et al.*

In order to convey the basic structure, we give a schematic comparison of the cascadic multigrid method with the nested V-cycle multigrid method:



The purpose of the present contribution is to survey the main methods of the proof, the main results for the case of adaptive triangulations and to include some numerical comparisons with nested multigrid.

2 Analysis for general smoothers

In this section, we analyze the cascadic multigrid method with respect to accuracy and computational complexity using a general smoother as iterative method on each discretization level.

Let $\Omega \subset \mathbf{R}^d$ be a polygonal Lipschitz domain. We consider an elliptic Dirichlet problem on Ω in the weak formulation:

$$u \in H_0^1(\Omega): \qquad a(u,v) = (f,v)_{L^2} \quad \forall v \in H_0^1(\Omega).$$

For the sake of a clear notation we only consider the H^2-regular case, i.e.,

$$\|u\|_{H^2} \leq c\,\|f\|_{L^2} \qquad \forall f \in L^2(\Omega).$$

The induced energy-norm will be denoted by

$$\|u\|_a^2 = a(u,u) \qquad \forall u \in H_0^1(\Omega).$$

Given a nested family of triangulations $(\mathcal{T}_j)_{j=1}^{\ell}$, the spaces of linear finite elements are

$$X_j = \{u \in C(\bar{\Omega}) : u|_T \in P_1(T) \quad \forall T \in \mathcal{T}_j, \quad u|_{\partial\Omega} = 0\},$$

where $P_1(T)$ denotes the linear functions on the triangle T. We have

$$X_0 \subset X_1 \subset \ldots \subset X_\ell \subset H_0^1(\Omega).$$

The finite element approximations are given by

$$u_j \in X_j: \qquad a(u_j, v_j) = (f, v_j)_{L^2} \qquad \forall v_j \in X_j.$$

In this section we consider *quasi-uniform* triangulations with meshsize parameter

$$\frac{1}{c}2^{-j} \leq h_j = \max_{T \in \mathcal{T}_j} \operatorname{diam} T \leq c\,2^{-j}.$$

For ease of notation, we will use the symbol c for any positive constant, that only depends on the bilinear form $a(\cdot,\cdot)$, on Ω and the shape regularity as well as the quasi-uniformity of the triangulations. All other dependencies will be stated explicitly.

Denoting the *basic iterative procedure* on each level by $\mathcal{I}$, the cascadic multigrid method can be written as:

$$\begin{aligned} &\text{(i)} \quad u_0^* = u_0 \\ &\text{(ii)} \quad j = 1, \ldots, \ell: \quad u_j^* = \mathcal{I}_{j,m_j} u_{j-1}^*. \end{aligned} \tag{1}$$

Here $\mathcal{I}_{j,m_j}$ denotes m_j steps of the basic iteration applied on level j.

We call a cascadic multigrid method *optimal* for level ℓ (with respect to the energy norm), if we obtain *both accuracy*

$$\|u_\ell - u_\ell^*\|_a \approx \|u - u_\ell\|_a,$$

which means that the iteration error is comparable to the approximation error, *and multigrid complexity*

$$\text{amount of work } = O(n_\ell),$$

where $n_\ell = \dim X_\ell$.

We consider the following type of basic iterations for the finite-element problem on level j started with $u_j^0 \in X_j$:

$$\|u_j - \mathcal{I}_{j,m_j} u_j^0\|_a \leq \|\mathcal{S}_{j,m_j}(u_j - u_j^0)\|_a$$

with a *linear* mapping $\mathcal{S}_{j,m_j} : X_j \to X_j$ for the error propagation. We call the basic iteration an *energy reducing smoother*, if it obeys the *smoothing properties*

$$\begin{aligned} &\text{(i)} \quad \|\mathcal{S}_{j,m_j} v_j\|_a \leq c \frac{h_j^{-1}}{m_j^\gamma} \|v_j\|_{L^2} \qquad \forall v_j \in X_j, \\ &\text{(ii)} \quad \|\mathcal{S}_{j,m_j} v_j\|_a \leq \|v_j\|_a \end{aligned} \tag{2}$$

with a parameter $0 < \gamma \leq 1$. As is shown in [Hac85] the symmetric Gauß-Seidel, the SSOR and the damped Jacobi iteration are smoothers in the sense of (2) with parameter $\gamma = 1/2$, for which the operator $\mathcal{S}_{j,m_j}$ is simply defined by

$$u_j - \mathcal{I}_{j,m_j} u_j^0 = \mathcal{S}_{j,m_j}(u_j - u_j^0).$$

This simple construction does not work for the conjugate gradient method because of its well-known *nonlinearity*. However, the conjugate gradient method can be put into our framework by means of a nontrivial construction as given by the authors in [BD96] which yields $\gamma = 1$.

The smoothing property (2(i)) resembles an *inverse* inequality which is exactly accompanied by the following *approximation property*

$$\|u_j - u_{j-1}\|_{L^2} \leq c\, h_j \|u_j - u_{j-1}\|_a \qquad j = 1, \ldots, \ell, \tag{3}$$

which can be proved by the standard Aubin-Nitsche duality argument.

The smoothing property and the approximation property are now the main building blocks of the central convergence estimate for the cascadic multigrid method (1).

Theorem 1 *The error of the cascadic multigrid method with a smoother as basic iteration can be estimated by*

$$\|u_\ell - u_\ell^*\|_a \le c \sum_{j=1}^{\ell} \frac{1}{m_j^\gamma} \|u_j - u_{j-1}\|_a \le c \sum_{j=1}^{\ell} \frac{h_j}{m_j^\gamma} \|f\|_{L^2}.$$

Proof. For $j = 1, \ldots, \ell$ we get by the linearity of $\mathcal{S}_{j,m_j}$

$$\begin{aligned} \|u_j - u_j^*\|_a &= \|u_j - \mathcal{I}_{j,m_j} u_{j-1}^*\|_a \le \|\mathcal{S}_{j,m_j}(u_j - u_{j-1}^*)\|_a \\ &\le \|\mathcal{S}_{j,m_j}(u_j - u_{j-1})\|_a + \|\mathcal{S}_{j,m_j}(u_{j-1} - u_{j-1}^*)\|_a. \end{aligned}$$

The first term can be estimated by the smoothing property (2(i)) and the approximation property (3):

$$\|\mathcal{S}_{j,m_j}(u_j - u_{j-1})\|_a \le c \frac{h_j^{-1}}{m_j^\gamma} \|u_j - u_{j-1}\|_{L^2} \le c \frac{1}{m_j^\gamma} \|u_j - u_{j-1}\|_a.$$

If we estimate the second term by property (2(ii)) of a smoother, we thus get

$$\|u_j - u_j^*\|_a \le \frac{c}{m_j^\gamma} \|u_j - u_{j-1}\|_a + \|u_{j-1} - u_{j-1}^*\|_a. \tag{4}$$

The start $u_0^* = u_0$ and induction yields the assertion. □

Since we have

$$2^{\ell-j} h_\ell / c \le h_j \le c\, 2^{\ell-j} h_\ell$$

Theorem 1 leads us to consider sequences $m_1, \ldots, m_\ell$ of the kind

$$m_j = \lceil \beta^{\ell-j} m_\ell \rceil, \tag{5}$$

for some fixed $\beta > 0$. Now, a simple evaluation of geometric sums gives conditions for accuracy and optimality:

Lemma 2 *Let the number m_j of iterations on level j be given by* (5). *The cascadic multigrid method yields the error*

$$\|u_\ell - u_\ell^*\|_a \le c \cdot \frac{1}{1 - (2/\beta^\gamma)} \cdot \frac{h_\ell}{m_\ell^\gamma} \|f\|_{L^2}, \qquad \text{for} \quad \beta > 2^{1/\gamma},$$

and a computational cost proportional to

$$\sum_{j=1}^{\ell} m_j n_j \le c \cdot \frac{1}{1 - \beta/2^d} \cdot m_\ell n_\ell, \qquad \text{for} \quad \beta < 2^d.$$

This Lemma shows that the two goals accuracy and multigrid complexity are not contradicting each other as long as

$$\gamma > 1/d.$$

As shown in [BD96] either the accuracy or the complexity has to deteriorate logarithmically for $\gamma = 1/d$.

Summarizing, we have proved that a plain symmetric Gauß-Seidel, SSOR or damped Jacobi iteration as basic iteration is optimal for $d = 3$, whereas the conjugate gradient method is optimal for $d \ge 2$.

3 Adaptive cascadic multigrid methods

In this section, we develop an a posteriori control for the number m_j of iteration on each level. This will be done for adaptively chosen triangulations, thus dropping the assumption of quasi-uniformity.

Using a *diagonally preconditioned* stiffness matrix the authors show in [BD96] for the smoothing iterations considered

$$\|\mathcal{S}_{j,m_j} v_j\|_a \leq \frac{c}{m_j^\gamma} \left(\sum_{T \in \mathcal{T}_j} h_T^{-2} \|v_j\|^2_{L^2(T)} \right)^{1/2}.$$

In order to compensate the hidden local inverse inequality we make a local and a global approximation assumption:

$$\begin{aligned} &\text{(i)} \quad h_T^{-2} \|u_j - u_{j-1}\|^2_{L^2(T)} \leq c \|u_j - u_{j-1}\|^2_{H^1(T)}, \qquad \forall T \in \mathcal{T}_j \\ &\text{(ii)} \quad \|u - u_j\| \leq c\, n_j^{-1/d} \|f\|_{L^2}, \end{aligned} \tag{6}$$

which are heuristically justified for *adaptive* triangulations. Note that quasi-uniform triangulations do not satisfy assumption (ii) for problems which are not H^2-regular.

The proof of Theorem 1 gives the estimate

$$\|u_\ell - u_\ell^*\|_a \leq c \sum_{j=1}^{\ell} \frac{1}{m_j^\gamma} \|u_j - u_{j-1}\|_a \leq c \sum_{j=1}^{\ell} \frac{1}{m_j^\gamma\, n_j^{1/d}} \|f\|_{L^2}.$$

We can now extend the accuracy and optimality result Lemma 2 to the adaptive case. Here we have to use additionally, that the sequence of number of unknowns belongs to a geometric progression:

$$n_j < \sigma_0 n_j \leq n_{j+1} \leq \sigma_1 n_j \qquad j = 0, 1, \ldots.$$

With the choice of the iteration numbers m_j on level j as

$$m_j = \left\lceil m_\ell \left(\frac{n_\ell}{n_j} \right)^{(d+1)/2d\gamma} \right\rceil, \tag{7}$$

we get for $d > 1$ under assumption (6) the final error

$$\|u_\ell - u_\ell^*\|_a \leq \frac{c}{m_\ell^\gamma\, n_\ell^{1/d}} \|f\|_{L^2}$$

and for $\gamma = 1$ the complexity

$$\sum_{j=1}^{\ell} m_j\, n_j \leq c\, m_\ell\, n_\ell.$$

However, at the intermediate level j we do *not* know the number n_ℓ of nodal points at the final level ℓ, which means that so far our iteration is not yet implementable.

To make it implementable, we *define* the final level ℓ as the first level on which the approximation error is below some user given tolerance TOL. Hence assumption (6) gives us the relation

$$\frac{\|u - u_j\|_a}{\mathrm{TOL}} \approx \left(\frac{n_\ell}{n_j}\right)^{1/d}, \tag{8}$$

which leads us to replace (7) by

$$m_j = \left\lceil m_\ell \left(\frac{\|u - u_j\|_a}{\mathrm{TOL}}\right)^{(d+1)/2\gamma} \right\rceil. \tag{9}$$

This algorithm is closest to the a priori choice of the parameters m_j. However, in an actual computation, the basic iteration can be accurate enough much earlier than stated by theory. Therefore, we now go back to the crucial recursion (4), i.e.,

$$\|u_j - u_j^*\|_a \leq \frac{c}{m_j^\gamma}\|u - u_{j-1}\|_a + \|u_{j-1} - u_{j-1}^*\|_a,$$

which we simply turn into a termination criterion for the basic iteration by inserting (9). We thus end up with the termination criterion

$$\|u_j - u_j^*\|_a \leq \rho \left(\frac{\mathrm{TOL}}{\|u - u_j\|_a}\right)^{(d+1)/2} \|u - u_{j-1}\|_a + \|u_{j-1} - u_{j-1}^*\|_a, \tag{10}$$

where $0 < \rho \leq 1$ is some safety factor. Note that the smoothing parameter γ dropped out since we stress the accuracy aspect of our analysis for the adaptive control. Herein the approximation error $\|u - u_j\|_a$ is not known, but can be replaced by the estimate

$$\|u - u_j\|_a \approx \|u - u_{j-1}\|_a \left(\frac{n_{j-1}}{n_j}\right)^{1/d} \approx \epsilon_{j-1} \left(\frac{n_{j-1}}{n_j}\right)^{1/d},$$

where ϵ_{j-1} denotes some estimate of the discretization error on the previous level, which is certainly provided by any adaptive algorithm; cf. [DLY89, BEK96]. If we replace the iteration errors $\|u_j - u_j^*\|_a$, $\|u_{j-1} - u_{j-1}^*\|_a$ by appropriate estimates δ_j, δ_{j-1}, the design of an implementable control strategy for the adaptive cascadic multigrid method is completed:

$$\delta_j \leq \rho \left(\frac{\mathrm{TOL}}{\epsilon_{j-1}} \left(\frac{n_j}{n_{j-1}}\right)^{1/d}\right)^{(d+1)/2} \epsilon_{j-1} + \delta_{j-1}. \tag{11}$$

4 A numerical experiment

The above designed adaptive cascadic multigrid method has been implemented in four different variants regarding to the choice of the underlying basic iteration:

- **CCG:** the conjugate gradient method, i.e., no coarse grid correction (cascadic conjugate gradient method).

- **CSGS:** the symmetric Gauß-Seidel iteration, i.e., again no coarse grid correction.
- **BPX:** the conjugate gradient method with the so-called BPX multilevel preconditioner due to Bramble, Pasciak and Xu [BPX90, Xu92, BY93]. This implementation is closest to the original "cascade principle" of Deuflhard, Leinen and Yserentant [DLY89, BEK93].
- **V-cycle:** a multigrid V-cycle using a Jacobi smoother, local — i.e., on new nodal points and neighbors only — in the case of locally refined triangulations. This implementation can be viewed as a robust and efficient *automatic choice* of the number m_j of V-cycles on level j for the usual *nested multigrid method.*

The variants were tested using the elliptic problem

$$-\Delta u = 0, \qquad u|_{\Gamma_1} = 10^3, \quad u|_{\Gamma_2} = 0, \quad \partial_n u|_{\Gamma_3} = 0$$

on a domain Ω which is a unit square with slit

$$\Omega = \{x \in \mathbf{R}^2 : |x|_\infty \leq 1\} \cap \{x \in \mathbf{R}^2 : |x_2| \geq 0.03 x_1\}.$$

The boundary pieces are

$$\Gamma_1 = \{x \in \Omega : x_1 = 1, x_2 \geq 0.03\}, \quad \Gamma_2 = \{x \in \Omega : x_1 = 1, x_2 \leq -0.03\},$$

and $\Gamma_3 = \partial\Omega \setminus (\Gamma_1 \cup \Gamma_2)$.

The accuracy parameter was set to a relative $\mathrm{TOL}_{\mathrm{rel}} = 2.24 \cdot 10^{-2}$, i.e., $\mathrm{TOL} = \|u\|_a \cdot \mathrm{TOL}_{\mathrm{rel}} = 412.52 \cdot \mathrm{TOL}_{\mathrm{rel}}$. Throughout the safety factor was set to $\rho = 0.4$. These choices yield 15 refinement levels with a final triangulation of roughly 4400 nodal points, which allows one to estimate the *discretization error* using two further *uniform* refinements:

$$\frac{\|u - u_{15}\|_a}{\|u\|_a} \approx 2 \cdot 10^{-2}, \qquad \frac{\|u - u_{15}\|_{L^2}}{\|u\|_{L^2}} \approx 1 \cdot 10^{-4},$$

where $\|u\|_a = 412.52$ and $\|u\|_{L^2} = 1121.36$.

Table 1 Algebraic errors of the variants for $\mathrm{TOL}_{\mathrm{rel}} = 2.24 \cdot 10^{-2}$

variant	$\frac{\|u_{15} - u_{15}^*\|_a}{\|u_{15}\|_a}$	$\frac{\|u_{15} - u_{15}^*\|_{L^2}}{\|u_{15}\|_{L^2}}$	$\frac{\sum \text{CPU-time}}{n_{15}}$
CCG	$6 \cdot 10^{-2}$	$1 \cdot 10^{-2}$	2.14 ms
CSGS	$6 \cdot 10^{-2}$	$1 \cdot 10^{-2}$	2.70 ms
BPX	$4 \cdot 10^{-3}$	$2 \cdot 10^{-4}$	3.53 ms
V-cycle (adap.)	$2 \cdot 10^{-3}$	$1 \cdot 10^{-4}$	4.27 ms

Table 1 shows the behavior of the (purely) *algebraic error* for the different variants. In *energy norm* they are comparable to the required accuracy. However, the variants without coarse grid corrections are slightly less accurate by a factor of three, whereas

the nested multigrid like variants BPX and V-cycle stay nicely below the discretization error.

With respect to the L^2-*norm* only the variants which include in some sense a coarse grid correction (BPX and V-cycle) give satisfactory results. This fact points out that the given termination criterion (11) can be viewed as a robust tool to control the number of iterations in nested multigrid methods. In terms of Schwarz methods [Xu92] one observes that the *additive* variant BPX is by a factor of two less accurate than the *multiplicative* V-cycle, but slightly faster.

Remark. The reader should note, that the term *optimality* has been used in this paper with respect to the *energy norm*. As Table 1 already indicates things are totally different for the L^2- or L^∞-norm. By means of simple examples one can *prove* that the cascadic multigrid method *cannot* be optimal with respect to the L^2-norm. There is a rather precise theoretical understanding of this phenomenon, which will be subject of a forthcoming publication [BK95].

REFERENCES

[BD96] Bornemann F. A. and Deuflhard P. (1996) The cascadic multigrid method for elliptic problems. *Numer. Math.* to appear.

[BEK93] Bornemann F. A., Erdmann B., and Kornhuber R. (1993) Adaptive multilevel methods in three space dimensions. *Inter. J. Numer. Meth. Engrg.* 36: 3187–3202.

[BEK96] Bornemann F. A., Erdmann B., and Kornhuber R. (1996) A posteriori error estimates for elliptic problems in two and three space dimensions. *SIAM J. Numer. Anal.* to appear.

[BK95] Bornemann F. A. and Krause R. (1995) manuscript in preparation.

[BY93] Bornemann F. A. and Yserentant H. (1993) A basic norm equivalence for the theory of multilevel methods. *Numer. Math.* 64: 455–476.

[BPWX91] Bramble J. H., Pasciak J. E., Wang J., and Xu J. (1991) Convergence estimates for multigrid algorithms without regularity assumptions. *Math. Comp.* 57: 23–45.

[BPX90] Bramble J. H., Pasciak J. E., and Xu J. (1990) Parallel multilevel preconditioners. *Math. Comp.* 55: 1–22.

[Deu94] Deuflhard P. (1994) Cascadic conjugate gradient methods for elliptic partial differential equations. Algorithm and numerical results. In Keyes D. and Xu J. (eds) *Proceedings of the 7th International Conference on Domain Decomposition Methods 1993.* AMS, Providence.

[DLY89] Deuflhard P., Leinen P., and Yserentant H. (1989) Concepts of an adaptive hierarchical finite element code. *IMPACT Comput. Sci. Engrg.* 1: 3–35.

[Hac85] Hackbusch W. (1985) *Multi-Grid Methods and Applications.* Springer-Verlag, Berlin, Heidelberg, New York.

[Sha94] Shaidurov V. V. (1994) *Some estimates of the rate of convergence for the cascadic conjugate-gradient method.* Otto-von-Guericke-Universität, Magdeburg. Preprint.

[Xu92] Xu J. (1992) Iterative methods by space decomposition and subspace correction. *SIAM Review* 34: 581–613.