7

Classical and Cascadic Multigrid—A Methodological Comparison

Folkmar A. Bornemann and Rolf Krause

1 Introduction

We consider second order elliptic boundary value problems on a polygonal domain $\Omega \subset \mathbf{R}^d$

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

Here $a(\cdot,\cdot)$ denotes a H_0^1-elliptic bilinear form and $f \in L^2(\Omega)$. Let

$$X_0 \subset X_1 \subset \ldots \subset X_j \subset \ldots \subset H_0^1(\Omega)$$

be a sequence of finite element spaces belonging to successively finer triangulations of Ω. On each level j the finite element solution u_j is given by a linear system

$$A_j u_j = f_j.$$

For its solution we think of an iterative scheme such that the result of m iterations with initial data u_j^0 will be denoted by $\mathcal{I}_j^m u_j^0$. The multilevel structure of the sequence $\{X_j\}$ suggests as a good idea to start the iteration on level j with the result of level $j-1$. Especially, this is of advantage in an adaptive setting where the space X_j is constructed only after a solution on level $j-1$ was obtained. Thus, on the final level ℓ an approximation u_ℓ^* of u_ℓ is computed via the multilevel scheme

$$u_0^* = u_0, \qquad u_j^* = \mathcal{I}_j^{m_j} u_{j-1}^* \qquad j = 1, \ldots, \ell. \tag{1}$$

If we choose for $\mathcal{I}_j$ the multigrid V-cycle on level j and the number of iterations as a constant m_*, i.e., $m_1 = \ldots = m_\ell = m_*$, this algorithm gives us the so-called *full multigrid* method [Hac85, Bra92]. Some authors call it *nested multigrid*.

Bornemann and Deuflhard [BD96a] considered standard iterative schemes like the damped Jacobi, the Gauss-Seidel, and the conjugate gradient iteration for $\mathcal{I}_j$. They showed that a proper choice of the number of iterations m_j on each level could make this method an "optimal" device—theoretically and practically. They named it the *cascadic multigrid method*. Its history can be obtained by browsing through

Ninth International Conference on Domain Decomposition Methods
Editor Petter E. Bjørstad, Magne S. Espedal and David E. Keyes

[Deu94, Sha96, BD96a, BD96b]. A distinctive feature of the cascadic multigrid method is the total absence of coarse grid corrections which means that coarse grids can be completely forgotten once they are refined. Therefore, the method is algorithmically attractive when a given finite element program cannot provide tree data structures of the refinement history but is used with a pre- and postprocessing device.

To be specific, the multilevel iteration (1) is called *optimal* with respect to an error norm $\|\cdot\|$ if we can choose for each final level ℓ a sequence of numbers of iterations $m_1, \dots, m_\ell$ such that simultaneously

- the method is *accurate*, i.e., the algebraic error on level ℓ is a fraction θ below the error of discretization,
$$\|u_\ell - u_\ell^*\| \leq \theta \|u - u_\ell\|,$$
where θ is a user given constant,
- the method has *multigrid complexity*, i.e., the total computational work for the iteration is bounded by
$$\text{work} \ \leq c \cdot n_\ell, \tag{2}$$
where c is some constant independent of ℓ, and n_ℓ denotes the number of unknowns on level ℓ.

The full multigrid method is well known to be optimal with respect to the energy norm and the L^2-norm [Hac85, Bra92]. Bornemann and Deuflhard [BD96a, BD96b] proved optimality of the cascadic multigrid method with respect to the *energy norm.* However, optimality with respect to the L^2-norm remained an open question. For *linear* finite elements simple numerical experiments suggested that the answer to that question is negative. In this paper we will give a theoretical explanation of this fact and show that the case might be different for *higher order* elements. This will be done by a careful analysis of the two grid variant of the cascadic multigrid method. Moreover we will provide a setting where one can understand the methodological difference between the cascadic multigrid method and the classical multigrid V-cycle almost immediately. As a rule of thumb we will establish that whenever the cascadic multigrid works the classical multigrid will work too, but not vice versa.

2 Two-grid Methods — the Abstract Setting

In the Galerkin framework, a typical setting for two grid methods is given by two finite dimensional spaces
$$X_{2h} \subset X_h$$
parametrized by some discretization parameter h. These spaces should be provided for at least a sequence of parameters h converging to zero. They itself are subspaces of certain function spaces which measure *smoothness.* In the following we have to compare two *different* measures of smoothness, given by the spaces X_+, X_-,
$$X_h \subset X_+ \hookrightarrow X_-.$$

The Galerkin method is given by a projection

$$P_h : X_+ \to X_h,$$

which obeys an approximation property, or *Jackson inequality*,[1]

$$\|u - P_h u\|_{X_-} \le c h^{\sigma} \|u\|_{X_+} \qquad \forall u \in X_+.$$

Here, $\sigma > 0$ denotes some positive constant. The Galerkin method is called *optimal* if this property is complemented by an inverse inequality, or *Bernstein inequality*,

$$\|u_h\|_{X_+} \le c h^{-\sigma} \|u_h\|_{X_-} \qquad \forall u_h \in X_h.$$

Notice that the Jackson inequality implies [Bra92] the compactness of the embedding $X_+ \hookrightarrow X_-$ thus making the two measures of smoothness really different and therefore showing the necessity of introducing them. The Bernstein inequality implies that the order σ of convergence as stated in the Jackson inequality is the best possible one [Bor94, Bra92].

Moreover, the two measures of smoothness allow an abstraction of the notion of low and high frequency in the space X_h. A function $v_h \in X_h$ is of high frequency if the X_+-norm gives much larger values than the X_--norm, i.e., taking the Bernstein inequality into account,

$$\|v_h\|_{X_+} \approx h^{-\sigma} \|v_h\|_{X_-}.$$

It is of low frequency if both norms give roughly the same value,

$$\|v_h\|_{X_+} \approx \|v_h\|_{X_-}.$$

Using this abstract notion of frequency one can easily understand and formulate the properties of basic iterative schemes for the solution of the computational problem,

$$u_h = P_h u,$$

which constitutes a large, badly conditioned linear system. The error propagation after m steps of such an iterative scheme,

$$u_h - u_h^m = S^m (u_h - u_h^0),$$

will be characterized by the *smoothing property* [BD96a, BD96b]

$$\|S^m v_h\|_{X_+} \le \frac{c h^{-\sigma}}{\phi(m)} \|v_h\|_{X_-} \qquad \forall v_h \in X_h.$$

Here, we suppose $\phi(m) \to \infty$ for $m \to \infty$. With respect to the X_+-norm such a smoothing iteration reduces high frequency errors with a h-independent rate whereas low frequency error components are handled increasingly less efficient for smaller and smaller h.

---

1 Here and in what follows we denote by c a generic constant independent of h.

Both the classical and the cascadic two grid method are designed to handle low frequency errors in a better way. For a given initial value $u_h^0 \in X_h$ the *classical* two grid method first performs m smoothing iterations,

$$u_h - u_h^m = S^m(u_h - u_h^0)$$

followed by a *coarse grid correction*,

$$u_h^* = u_h^m + P_{2h}(u_h - u_h^m).$$

The combined error of these two steps is given by

$$\begin{aligned}\|u_h - u_h^*\|_{X_-} &= \|(u_h - u_h^m) - P_{2h}(u_h - u_h^m)\|_{X_-} \leq ch^{\sigma}\|u_h - u_h^m\|_{X_+} \\ &\leq \frac{c}{\phi(m)}\|u_h - u_h^0\|_{X_-},\end{aligned}$$

where we have used first the Jackson inequality and second the smoothing property. Thus, we end up with an error reduction in the X_--norm which is independent of the discretization parameter h.

The *cascadic* two grid method changes somewhat the order of the two steps of the classical method. It first performs a coarse grid projection

$$u_h^0 = P_{2h}u_h$$

followed by m smoothing iterations,

$$u_h - u_h^* = S^m(u_h - u_h^0).$$

Here the combined error is given by

$$\begin{aligned}\|u_h - u_h^*\|_{X_+} &\leq \frac{ch^{-\sigma}}{\phi(m)}\|u_h - P_{2h}u_h\|_{X_-} = \frac{ch^{-\sigma}}{\phi(m)}\|(u_h - u_h^0) - P_{2h}(u_h - u_h^0)\|_{X_-} \\ &\leq \frac{c}{\phi(m)}\|u_h - u_h^0\|_{X_+},\end{aligned}$$

using first the smoothing property and then the Jackson inequality. Here we end up with an error reduction in the X_+-norm which is independent of the discretization parameter h.

Notice, however, some important differences between the two methods. Since the order of the smoothing iterations and the coarse grid problem are interchanged, and therefore the Jackson inequality and the smoothing property (which resembles the Bernstein inequality) are applied in reverse order, the error reduction occurs in *different norms*: For the classical two grid in X_-, for the cascadic two grid in X_+. As we will see later on this is the reason for essentially different behavior in certain settings. A second methodological difference is of algorithmic nature: Unlike the classical two grid method, the cascadic two grid method has no choice of an initial value, which precludes it from being part of an iterative scheme itself.

3 Application to Finite Element Spaces

Here, we apply the abstract theory of the preceding section to k-th order finite elements. We consider two choices for the norm in which error reduction of the iterative scheme is measured: energy norm and L^2-norm. These two examples clearly reveal the general principle.

Energy Norm and Classical Multigrid

Here, error reduction is obtained in the X_--norm and we thus put $X_- = H^1(\Omega)$. The more smooth space is set to $X_+ = H^{1+\alpha}(\Omega)$ where $\alpha > 0$ is chosen such that the corresponding Jackson and Bernstein inequalities hold:

$$\|u - P_h u\|_{H^1} \leq ch^{\alpha}\|u\|_{H^{1+\alpha}}, \qquad \|u_h\|_{H^{1+\alpha}} \leq ch^{-\alpha}\|u_h\|_{H^1} \tag{3}$$

for all $u \in H^{1+\alpha}$ and $u_h \in X_h$. The smoothness parameter[2] α is restricted by the regularity of the elliptic problem and by the order of the chosen finite elements: For $H^{1+\mu}$-regular elliptic problems we can take any α as large as

$$\alpha \leq \min(\mu, k).$$

For $\alpha \geq 1/2$, the term $\|u_h\|_{H^{1+\alpha}}$ is meant to denote the corresponding *discrete* Sobolev norm [Bra92]. Now, our abstract theory yields the h-independent error reduction

$$\|u_h - u_h^*\|_{H^1} \leq \frac{c}{\phi(m)}\|u_h - u_h^0\|_{H^1}.$$

Energy Norm and Cascadic Multigrid

Here, error reduction is obtain in the X_+-norm and we thus put $X_+ = H^1(\Omega)$. Whenever the Jackson and Bernstein inequalities (3) of the $(H^1, H^{1+\alpha})$ pair holds we get by a Aubin-Nitsche type of duality argument corresponding Jackson- and Bernstein inequalities for the pair $(H^{1-\alpha}, H^1)$,

$$\|u - P_h u\|_{H^{1-\alpha}} \leq ch^{\alpha}\|u\|_{H^1}, \qquad \|u_h\|_{H^1} \leq ch^{-\alpha}\|u_h\|_{H^{1-\alpha}} \tag{4}$$

for all $u \in H^1$ and $u_h \in X_h$. As in the case of the classical multigrid method we thus get the h-independent error reduction

$$\|u_h - u_h^*\|_{H^1} \leq \frac{c}{\phi(m)}\|u_h - u_h^0\|_{H^1}.$$

L^2-Norm and Classical Multigrid

Here, we put $X_- = L^2(\Omega)$ and assume enough regularity for $\alpha = 1$ in (3) and (4). Taking $X_+ = H^1(\Omega)$ the Jackson and Bernstein inequalities (4) lead to the h-independent convergence rate

$$\|u_h - u_h^*\|_{L^2} \leq \frac{c}{\phi(m)}\|u_h - u_h^0\|_{L^2}.$$

2 For a polygonal domain Ω and piecewise smooth coefficients of the elliptic operator we always get some $\alpha > 0$.

L^2-Norm and Cascadic Multigrid

We have to put $X_+ = L^2(\Omega)$. Thus, X_- must be a Sobolev space of *negative* order, say $X_- = H^{-\epsilon}(\Omega)$, $\epsilon > 0$. Our argument would work, if we had a Jackson inequality like

$$\|u - P_h u\|_{H^{-\epsilon}} \leq ch^{\epsilon}\|u\|_{L^2}. \tag{5}$$

Assuming the same regularity as for the classical two grid method, i.e., at least $\alpha = 1$, and replacing u by $u - P_h u$ in (5) we get

$$\|u - P_h u\|_{H^{-\epsilon}} \leq ch^{\epsilon}\|u - P_h u\|_{L^2} \leq ch^{1+\epsilon}\|u - P_h u\|_{H^1}.$$

However, for a right hand side $f \in H^{\epsilon}(\Omega)$ this would be possible *only if* we could impose enough regularity: By duality

$$\|u - P_h u\|_{H^{-\epsilon}} \geq \frac{\langle u - P_h u, f\rangle_{L^2}}{\|f\|_{H^{\epsilon}}} = \frac{a(u - P_h u, u - P_h u)}{\|f\|_{H^{\epsilon}}} \geq c\,\frac{\|u - P_h u\|^2_{H^1}}{\|f\|_{H^{\epsilon}}}$$

we would get the H^1-estimate

$$\|u - P_h u\|_{H^1} \leq ch^{1+\epsilon}\,\|f\|_{H^{\epsilon}},$$

which means that $\alpha = 1 + \epsilon$ would be admissible, cf. [Bor94].

In particular, the hypothetical Jackson inequality (5) does *not* hold for *linear* finite elements, $k = 1$, where we are restricted to $\alpha \leq 1$. In this case the only estimate we can prove for the cascadic two grid method is

$$\|u_h - u_h^*\|_{L^2} \leq \frac{c_\epsilon h^{-\epsilon}}{\phi_\epsilon(m)}\|u_h - u_h^0\|_{L^2}.$$

Using a damped Jacobi or Gauss-Seidel iteration, we have $\phi_\epsilon(m) = m^{\epsilon/2}$ as shown in [BD96a]. *Assuming* that our estimate is essentially a sharp one we would expect that the number of iterations m_h, which is needed to reduce the algebraic error a given amount, *increases* like

$$m_h \propto h^{-2}.$$

This was observed in several numerical experiments. Hence our estimates appear to be rather sharp and there is not much to improve.

Discussion

The Aubin-Nitsche duality argument and the reverse duality argument of the last paragraph show that Jackson and Bernstein inequalities for the finite element spaces are located exactly in the smoothness range between $H^{1-\alpha}$ and $H^{1+\alpha}$ which is *symmetric* with respect to H^1. For this reason the classical and the cascadic two grid method have the same chance to locate the partner space $X_\pm$ of the space $X_\mp$ that measures the energy norm. However, since the space L^2 is located in this smoothness range in a way leaving more place for more smooth spaces than for less smooth spaces it gives preference to the *classical* two grid method which puts $X_- = L^2$ and needs a more smooth partner space X_+.

4 Remarks on Optimality in the L^2-NORM

As we have seen, the cascadic two grid method works for the L^2-norm if we have $\alpha > 1$. *Assuming* this regularity we will discuss shortly whether we can prove optimality of the cascadic multigrid method with respect to L^2 on uniform triangulations. We follow quite closely the proof given in [BD96a, BD96b] for the energy norm case.

By linearity the basic error estimate governing the multilevel iteration (1) is now given by

$$\|u_j - u_j^*\|_{L^2} \leq \|S^{m_j}(u_j - u_{j-1})\|_{L^2} + \|S^{m_j}(u_{j-1} - u_{j-1}^*)\|_{L^2}.$$

Applying that recursively we get by setting $M_j = m_\ell + \ldots + m_j$ the estimate

$$\begin{aligned} \|u_\ell - u_\ell^*\|_{L^2} &\leq \sum_{j=1}^{\ell} \|S^{M_j}(u_j - u_{j-1})\|_{L^2} \leq c \sum_{j=1}^{\ell} \frac{h_j^{1-\alpha}}{M_j^{\gamma(\alpha-1)}} \|u_j - u_{j-1}\|_{H^{1-\alpha}} \\ &\leq c \sum_{j=1}^{\ell} \frac{1}{M_j^{\gamma(\alpha-1)}} \|u_j - u_{j-1}\|_{L^2}, \end{aligned}$$

where we have used the smoothing property and the Jackson inequality (5) with $\epsilon = \alpha - 1$. Moreover we specified the function ϕ of the smoothing property by $\phi(m) = m^{\epsilon\gamma}$, where $\gamma = 1/2$ for the damped Jacobi or Gauss-Seidel iteration and $\gamma = 1$ for the CG-iteration [BD96a, BD96b]. As discussed in these references we set

$$m_j = \lceil \beta^{\ell-j} m_\ell \rceil.$$

Taking into account the L^2-error estimate

$$\|u_j - u_{j-1}\|_{L^2} \leq c h_j^{1+\alpha} \|u\|_{H^{1+\alpha}},$$

we finally get the estimate

$$\|u_\ell - u_\ell^*\|_{L^2} \leq \frac{c h_\ell^{1+\alpha}}{m_\ell^{\gamma(\alpha-1)}} \sum_{j=0}^{\ell-1} \left(\frac{2^{\alpha+1}}{\beta^{\gamma(\alpha-1)}} \right)^j \|u\|_{H^{1+\alpha}}.$$

Thus, *accuracy* of the method is guaranteed if the sum can be bounded independently of the final level ℓ. This is the case if and only if

$$\beta > 2^{\frac{1}{\gamma}\frac{\alpha+1}{\alpha-1}}.$$

As shown in [BD96a, BD96b] multigrid complexity is obtained if and only if $\beta < 2^d$, where d is the dimension of the domain Ω. Thus, a sufficient condition for the cascadic multigrid to be optimal with respect to the L^2-norm is

$$d > \frac{1}{\gamma} \cdot \frac{\alpha + 1}{\alpha - 1}. \tag{6}$$

If we relax the demand (2) for multigrid complexity to

$$\text{work} \leq c \cdot n_\ell \log^p n_\ell, \tag{7}$$

for some $p > 0$, one can show exactly in the same way as in [BD96a] for the energy norm that equality is admissible in condition (6).

Examples

This restrictive condition will be illuminated by several specific cases:

- $d = 2$, $\gamma = 1$ (CG-iteration): Condition (6) is equivalent to $\alpha > 3$, which means at least H^σ-regularity with $\sigma > 4$ and order $k > 3$ finite elements.
- $d = 3$, $\gamma = 1$ (CG-iteration): Condition (6) is equivalent to $\alpha > 2$, which means at least H^σ-regularity with $\sigma > 3$ and order $k > 2$ finite elements.
- $d = 2$, $\gamma = 1/2$ (damped Jacobi or Gauss-Seidel iteration): Condition (6) is *not* satisfied for any $\alpha > 1$. In this case the cascadic multigrid method is not optimal with respect to the L^2-norm for any regularity and any order of finite elements!
- $d = 3$, $\gamma = 1/2$ (damped Jacobi or Gauss-Seidel iteration): Condition (6) is equivalent to $\alpha > 5$ which means at least H^σ-regularity with $\sigma > 6$ and order $k > 5$ finite elements.

For the relaxed multigrid complexity (7) equality is admissible in all cases.

REFERENCES

[BD96a] Bornemann F. A. and Deuflhard P. (1996) The cascadic multigrid method for elliptic problems. *Numer. Math.* 75: 135–152.

[BD96b] Bornemann F. A. and Deuflhard P. (1996) Cascadic Multigrid Methods. In Glowinski R., Périaux J., Shi Z., and Widlund O. (eds) *Domain Decomposition Methods in Sciences and Engineering.* John Wiley & Sons, Chichester, New York.

[Bor94] Bornemann F. A. (1994) Interpolation spaces and optimal multilevel preconditioners. In Keyes D. and Xu J. (eds) *Proceedings of the 7th International Conference on Domain Decomposition Methods 1993*, pages 3–8. AMS, Providence.

[Bra92] Braess D. (1992) *Finite Elemente.* Springer-Verlag, Berlin, Heidelberg, New York.

[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*, pages 29–42. AMS, Providence.

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

[Sha96] Shaidurov V. V. (1996) Some estimates of the rate of convergence for the cascadic conjugate-gradient method. *Computers Math. Applic.* 31: 161–171.