NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2010; **17**:273–290
Published online 21 December 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/nla.685

When does two-grid optimality carry over to the V-cycle?

Artem Napov*,† and Yvan Notay

Service de Métrologie Nucléaire, Université Libre de Bruxelles (C.P. 165/84), 50, Av. F.D. Roosevelt, B-1050 Brussels, Belgium

SUMMARY

We investigate additional condition(s) that confirm that a V-cycle multigrid method is satisfactory (say, optimal) when it is based on a two-grid cycle with satisfactory (say, level-independent) convergence properties. The main tool is McCormick's bound on the convergence factor (*SIAM J. Numer. Anal.* 1985; **22**:634–643), which we showed in previous work to be the best bound for V-cycle multigrid among those that are characterized by a constant that is the maximum (or minimum) over all levels of an expression involving only two consecutive levels; that is, that can be assessed considering only two levels at a time. We show that, given a satisfactorily converging two-grid method, McCormick's bound allows us to prove satisfactory convergence for the V-cycle if and only if the norm of a given projector is bounded at each level. Moreover, this projector norm is simple to estimate within the framework of Fourier analysis, making it easy to supplement a standard two-grid analysis with an assessment of the V-cycle potentialities. The theory is illustrated with a few examples that also show that the provided bounds may give a satisfactory sharp prediction of the actual multigrid convergence. Copyright © 2009 John Wiley & Sons, Ltd.

Received 15 May 2009; Revised 21 October 2009; Accepted 23 October 2009

KEY WORDS: multigrid; V-cycle; approximation property; two-grid; convergence analysis; Fourier analysis

1. INTRODUCTION

We consider multigrid methods for the solution of symmetric positive-definite (SPD) $n\times n$ linear systems:

$$Ax=b. \tag{1}$$

*Correspondence to: Artem Napov, Service de Métrologie Nucléaire, Université Libre de Bruxelles (C.P. 165/84), 50, Av. F.D. Roosevelt, B-1050 Brussels, Belgium.
†E-mail: anapov@ulb.ac.be

Contract/grant sponsor: Belgian FNRS ('Aspirant')
Contract/grant sponsor: Belgian FNRS ('Directeur de recherches')

Multigrid methods are based on the recursive use of a two-grid scheme. A basic two-grid method combines the action of a *smoother*, often a simple iterative method such as Gauss-Seidel, and a *coarse-grid correction*, which corresponds to the solution of the residual equation on a coarser grid. A V-cycle multigrid method is obtained when the residual equation is solved approximately with one application of the two-grid scheme on that level, and so on, until the coarsest level, where an exact solve is performed. Other cycles may be defined, including the W-cycle based on two recursive applications of the two-grid scheme on each level, see, e.g. [1].

If there are only two levels, accurate bounds may be obtained either by means of Fourier analysis [1–3], or by using some appropriate algebraic tools [4–8]. This focus on two-grid schemes is motivated by the fact that, 'if the two-grid method converges sufficiently well, then the multigrid method with W-cycle will have similar convergence properties' [1, p. 77] (see also [9, pp. 226–228] and [10]). This is not the case for the V-cycle since there are known examples where the two-grid method converges relatively well, whereas the multigrid method with V-cycle scales poorly with the number of levels [11]. Hence, V-cycle analysis has to be, at some point, essentially different from two-grid analysis.

In this paper, we investigate additional condition(s) for obtaining an optimal V-cycle method from an optimal[‡] two-grid method. Note that we do not base our work on a new analysis of the V-cycle. Several analyses are indeed available, which, however, have a common gap: the conditions for proving that the V-cycle converges nicely have not been compared with the two-grid convergence factor, and it is so far unclear how they are related. In fact, a number of results relate the V-cycle convergence to *sufficient* conditions for two-grid convergence; see, e.g. the two conditions (8) in [12], the first of which is sufficient for two-grid. Or, simply, consider V-cycle analysis particularized to the two-level case. Such sufficient conditions are, however, often stronger than needed for just two-level convergence, and, as far as we know, no comparison has been made with *necessary and sufficient* conditions or with two-grid convergence factor.

To analyze the V-cycle, one possibility consists of defining an appropriate subspace decomposition and then applying successive subspace correction (SSC) theory [13–18]. Another possibility consists in checking the so-called smoothing and approximation properties [19–25]. Regarding the latter approach, the best results for SPD matrices have been obtained by Hackbusch [22, Theorem 7.2.2] and McCormick [24]. In a previous paper [26], we show that these results are qualitatively equivalent with McCormick's bound being always the sharpest. Note that, in both cases, the bound is characterized by a constant that is the minimum/maximum over all levels of an expression involving only two consecutive levels. This last property is important in the context of this study, since it seems at first sight not possible to compare with the two-grid convergence rate a global expression that would involve simultaneously all levels.

On the other hand, we also consider in [26] the classical formulation of the SSC theory (as stated in [16, 17]), and discuss how to obtain a bound that could also be assessed considering only two levels at a time. It turns out that this requires the use of the so-called a-orthogonal decomposition, which corresponds to the choice most frequently made when applying the SSC theory to multigrid methods for H^2-regular problems. Then, the analysis in [26] shows that this approach is also qualitatively equivalent to the Hackbusch and McCormick ones, the latter remaining the sharpest.

[‡]By 'optimal', for a two-grid method, we mean 'having level-independent convergence properties'; that is, referring to a situation where the two-grid method is defined at different levels of a multigrid hierarchy, it is considered optimal if there is a level-independent bound on the convergence factor that is uniform with respect to the problem size.

DOI: 10.1002/nla

Hence, regarding the goal pursued in this work, all exploitable results are superseded by (but qualitatively equivalent to) McCormick's bound, which is characterized by the constant δ; in this work, we relate this constant to the two-grid convergence factor. This reveals that a satisfactory (optimal) two-grid cycle on each level leads to a satisfactory estimate of δ if and only if a given norm of an exact coarse-grid correction (projection) operator remains bounded at each level. Moreover, it turns out that this norm is easy to assess within the framework of a Fourier analysis.

Eventually, we consider several examples, illustrating the sharpness of the bound based on two-grid convergence rates and the projector norm. It further turns out that both of these ingredients are independent and play an important role in the V-cycle convergence behavior.

The reminder of this paper is organized as follows. In Section 2 we state the general setting of this study and gather the needed assumptions. The relation between the McCormick constant δ and the two-grid convergence factor is established in Section 3. Illustrative examples are discussed in Section 4.

Notation

Let I denote the identity matrix and O the zero matrix. When the dimensions are not obvious from the context, we write more specifically I_m for the $m\times m$ identity matrix.

For any rectangular matrix B, B^{T} stands for its transpose and B^H for its transpose complex conjugate. For any square matrix C, $\rho(C)$ is its spectral radius (that is, its largest eigenvalue in modulus), $\|C\|=\sqrt{\rho(C^TC)}$ is the usual 2-norm and, for an SPD matrix D, $\|C\|_D=\|D^{1/2}CD^{-1/2}\|$ is the D-norm (if $D=A$, it is also called energy norm).

2. GENERAL SETTING

We consider a multigrid method with $J+1$ levels ($J\geqslant 1$); index J refers to the finest level (on which the system (1) is to be solved), and index 0 to the coarsest level. The number of unknowns at level k, $0\leqslant k\leqslant J$, is noted n_k (with thus $n_J=n$).

Our analysis applies to symmetric multigrid schemes based on the Galerkin principle for the SPD system (1); that is, restriction is the transpose of prolongation and the matrix A_k at level k, $k=J-1,\dots,0$, is given by $A_k=P_k^TA_{k+1}P_k$, where P_k is the prolongation operator from level k to level $k+1$; we also assume that the smoother R_k is SPD and that the number of pre-smoothing steps v ($v>0$) is equal to the number of post-smoothing steps. The algorithm for V-cycle multigrid is then as follows.

Multigrid with V-cycle at level $k: x_{n+1}=\mathrm{MG}(b, A_k, x_n, k)$

(1) Relax v times with smoother $R_k: \bar{x}_n=$ Smooth (x_n, A_k, R_k, v, b)
(2) Compute residual: $r_k=b-A_k\bar{x}_n$
(3) Restrict residual: $r_{k-1}={P_{k-1}}^{\mathrm{T}}r_k$
(4) Corse-grid correction: **if** $k=1$, $e_0=A_0^{-1}r_0$ **else** $e_{k-1}=\mathrm{MG}$ $(r_{k-1}, A_{k-1}, 0, k-1)$
(5) Prolongate coarse-grid correction: $\hat{\bar{x}}_n=\bar{x}_n+P_{k-1}e_{k-1}$
(6) Relax v times with smoother $R_k: x_{n+1}=$ Smooth $(\hat{\bar{x}}_n, A_k, R_k, v, b)$

When applying this algorithm, the error satisfies

$$A_k^{-1}b-x_{n+1}=E_{MG}^{(k)}(A_k^{-1}b-x_n),$$

DOI: 10.1002/nla

where the iteration matrix $E_{MG}^{(k)}$ is recursively defined from

$$\begin{aligned} &E_{MG}^{(0)} = 0 \quad \text{and for } k = 1, 2, \ldots, J: \\ &E_{MG}^{(k)} = (I - R_k^{-1} A_k)^{\nu} (I - P_{k-1}(I - E_{MG}^{(k-1)}) A_{k-1}^{-1} P_{k-1}{}^{\mathrm{T}} A_k)(I - R_k^{-1} A_k)^{\nu} \end{aligned} \tag{2}$$

(see, e.g. [1, p. 48]). Our main objective is the analysis of the spectral radius of $E_{MG}^{(J)}$, which governs convergence on the finest level. Our analysis makes use of the following general assumptions.

General assumptions

- $n = n_J > n_{J-1} > \cdots > n_0$;
- P_k is an $n_{k+1} \times n_k$ matrix of rank n_k, $k = J-1, \ldots, 0$;
- $A_J = A$ and $A_k = P_k^T A_{k+1} P_k$, $k = J-1, \ldots, 0$;
- R_k is SPD and such that $\rho(I - R_k^{-1} A_k) < 1$, $k = J, \ldots, 1$.

In what follows, we make use of the two-grid cycle involving two consecutive levels k and $k-1$, which corresponds to the following iteration matrix:

$$E_{TG}^{(k)} = (I - R_k^{-1} A_k)^{\nu} (I - P_{k-1} A_{k-1}^{-1} P_{k-1}{}^{\mathrm{T}} A_k)(I - R_k^{-1} A_k)^{\nu}, \quad k = 1, \ldots, J. \tag{3}$$

Most of our results do not refer explicitly to the smoother R_k, but are stated with respect to the matrices $M_k^{(\nu)}$ defined from

$$I - M_k^{(\nu)\,-1} A_k = (I - R_k^{-1} A_k)^{\nu}. \tag{4}$$

That is, $M_k^{(\nu)}$ is the smoother that provides in one step the same effect as ν steps with R_k. The results stated with respect to $M_k^{(\nu)}$ may then be seen as the results stated for the case of one pre-and one post-smoothing step, which can be extended to the general case via the relations (4).

We close this subsection by introducing the projector π_{A_k}, which plays an important role throughout this paper:

$$\pi_{A_k} = P_{k-1} A_{k-1}^{-1} P_{k-1}{}^{\mathrm{T}} A_k. \tag{5}$$

Note that $I - \pi_{A_k}$ is the (exact) coarse-grid correction matrix at level k.

3. THEORETICAL ANALYSIS

3.1. McCormick's bound

We recall in the following theorem the bound obtained in [24, Lemma 2.3, Theorem 3.4 and Section 5] (see also [23, 25] for an alternative proof). The equivalence of (8) with the definition (7) is proved in [26].

Note that the convergence estimates based on regularity assumptions are also considered in [23]. These estimates are obtained when Theorem 3.1 below is applied to the discretized Partial differential equations (PDEs). However, Theorem 3.1 on its own is a purely algebraic result that may by applied to any multigrid method satisfying the general assumptions in Section 2, without reference to a PDE context. Hence, there is no need for regularity assumptions to apply here, as may be further confirmed by the purely algebraic proof in [25].

Theorem 3.1
Let $E_{MG}^{(J)}$, $M_k^{(\nu)}$, and π_{A_k}, $k=1,\ldots,J$, be defined, respectively, by (2), (4), and (5), with P_k, $k=0,\ldots,J-1$, A_k, $k=0,\ldots,J$, and R_k, $k=1,\ldots,J$, satisfying the general assumptions stated in Section 2.

Then

$$\rho(E_{MG}^{(J)})\leqslant 1-\delta^{(\nu)}\ , \tag{6}$$

where

$$\delta^{(\nu)} = \min_{1\leqslant k\leqslant J}\min_{v_k\in\mathbb{R}^{n_k}} \frac{\|v_k\|_{A_k}^2-\|(I-M_k^{(\nu)\,-1}A_k)v_k\|_{A_k}^2}{\|(I-\pi_{A_k})v_k\|_{A_k}^2} \tag{7}$$

$$= \min_{1\leqslant k\leqslant J}\min_{v_k\in\mathbb{R}^{n_k}} \frac{v_k^{\mathrm{T}}A_kv_k}{v_k^{\mathrm{T}}(I-\pi_{A_k})^{\mathrm{T}}M_k^{(2\nu)}(I-\pi_{A_k})v_k} \tag{8}$$

3.2. Relationship to the two-grid convergence rate

We first recall, in the following lemma, a useful characterization of the two-grid rate obtained in [6, p. 480].

Lemma 3.1
Let $E_{TG}^{(k)}$, $M_k^{(\nu)}$, and π_{A_k}, $k=1,\ldots,J$, be defined, respectively, by (3), (4), and (5), with P_k, $k=0,\ldots,J-1$, A_k, $k=0,\ldots,J$, and R_k, $k=1,\ldots,J$, satisfying the general assumptions stated in Section 2.

Then

$$1-\rho(E_{TG}^{(k)})=\min_{v_k\in\mathbb{R}^{n_k}}\frac{v_k^{\mathrm{T}}(I-\bar{\pi}_{A_k})A_k^{1/2}M_k^{(2\nu)^{-1}}A_k^{1/2}(I-\bar{\pi}_{A_k})v_k}{v_k^{\mathrm{T}}(I-\bar{\pi}_{A_k})v_k}, \tag{9}$$

with $\bar{\pi}_{A_k}=A_k^{1/2}\pi_{A_k}A_k^{-1/2}$.

The next theorem contains our main result.

Theorem 3.2
Let $E_{TG}^{(k)}$, $M_k^{(\nu)}$, and π_{A_k}, $k=1,\ldots,J$, be defined, respectively, by (3), (4), and (5), with P_k, $k=0,\ldots,J-1$, A_k, $k=0,\ldots,J$, and R_k, $k=1,\ldots,J$, satisfying the general assumptions stated in Section 2. Let $\delta^{(\nu)}$ be defined by (7).

Then

$$\delta^{(\nu)}\geqslant\min_{1\leqslant k\leqslant J}\frac{1-\rho(E_{TG}^{(k)})}{\|I-\pi_{A_k}\|_{M_k^{(2\nu)}}^2}=\min_{1\leqslant k\leqslant J}\frac{1-\rho(E_{TG}^{(k)})}{\|\pi_{A_k}\|_{M_k^{(2\nu)}}^2}. \tag{10}$$

Moreover,

$$\delta^{(\nu)}\leqslant\min_{1\leqslant k\leqslant J}\min\left(1-\rho(E_{TG}^{(k)}),\ \frac{1}{\|\pi_{A_k}\|_{M_k^{(2\nu)}}^2}\right). \tag{11}$$

DOI: 10.1002/nla

Proof
Let ξ_k be defined by

$$\xi_k = \min_{v\in\mathbb{R}^k} \frac{v^{\mathrm{T}} A_k v}{v^{\mathrm{T}}(I-\pi_{A_k})^{\mathrm{T}} M_k^{(2v)}(I-\pi_{A_k})v}.$$

From (8), there holds

$$\delta^{(v)} = \min_{1\leqslant k\leqslant J} \xi_k. \tag{12}$$

On the other hand, Lemma 3.1 implies (since $A_k(I-\pi_{A_k}) = (I-\pi_{A_k})^{\mathrm{T}} A_k$ and $(I-\pi_{A_k}) = (I-\pi_{A_k})^2$)

$$\begin{aligned}
1-\rho(E_{TG}^{(k)}) &= \min_{v_k\in\mathbb{R}^{n_k}} \frac{v_k^{\mathrm{T}} A_k^{1/2}(I-\pi_{A_k}) M_k^{(2v)^{-1}} A_k (I-\pi_{A_k}) A_k^{-1/2} v_k}{v_k^{\mathrm{T}} A_k^{1/2} (I-\pi_{A_k}) A_k^{-1/2} v_k} \\
&= \min_{v_k\in\mathbb{R}^{n_k}} \frac{v_k^{\mathrm{T}} (I-\pi_{A_k}) M_k^{(2v)^{-1}} A_k (I-\pi_{A_k}) A_k^{-1} v_k}{v_k^{\mathrm{T}} (I-\pi_{A_k})(I-\pi_{A_k}) A_k^{-1} v_k} \\
&= \min_{v_k\in\mathbb{R}^{n_k}} \frac{v_k^{\mathrm{T}} (I-\pi_{A_k}) M_k^{(2v)^{-1}} (I-\pi_{A_k})^{\mathrm{T}} v_k}{v_k^{\mathrm{T}} (I-\pi_{A_k}) A_k^{-1} (I-\pi_{A_k})^{\mathrm{T}} v_k}.
\end{aligned} \tag{13}$$

In what follows, we omit the subscripts k, as well as the superscript (k) and $(2v)$ in E_{TG} and M, respectively, when they are obvious from context. Using (13), one obtains

$$\begin{aligned}
\xi^{-1} &= \max_{v\in\mathbb{R}^n} \frac{v^{\mathrm{T}}(I-\pi_A)^{\mathrm{T}} M (I-\pi_A) v}{v^{\mathrm{T}} A v} \\
&= \max_{v\in\mathbb{R}^n} \frac{v^{\mathrm{T}} A^{-1/2} (I-\pi_A)^{\mathrm{T}} M^{1/2} M^{1/2} (I-\pi_A) A^{-1/2} v}{v^{\mathrm{T}} v} \\
&= \max_{v\in\mathbb{R}^n} \frac{v^{\mathrm{T}} M^{1/2} (I-\pi_A) A^{-1/2} A^{-1/2} (I-\pi_A)^{\mathrm{T}} M^{1/2} v}{v^{\mathrm{T}} v} \\
&= \max_{v\in\mathbb{R}^n} \frac{v^{\mathrm{T}} (I-\pi_A) A^{-1} (I-\pi_A)^{\mathrm{T}} v}{v^{\mathrm{T}} M^{-1} v} \qquad (14) \\
&\leqslant \max_{v\in\mathbb{R}^n} \frac{v^{\mathrm{T}} (I-\pi_A) A^{-1} (I-\pi_A)^{\mathrm{T}} v}{v^{\mathrm{T}} (I-\pi_A) M^{-1} (I-\pi_A)^{\mathrm{T}} v} \max_{v\in\mathbb{R}^n} \frac{v^{\mathrm{T}} (I-\pi_A) M^{-1} (I-\pi_A)^{\mathrm{T}} v}{v^{\mathrm{T}} M^{-1} v} \\
&= \frac{1}{1-\rho(E_{TG})} \max_{v\in\mathbb{R}^n} \frac{v^{\mathrm{T}} M^{1/2} (I-\pi_A) M^{-1/2} M^{-1/2} (I-\pi_A)^{\mathrm{T}} M^{1/2} v}{v^{\mathrm{T}} v} \\
&= \frac{1}{1-\rho(E_{TG})} \max_{v\in\mathbb{R}^n} \frac{v^{\mathrm{T}} M^{-1/2} (I-\pi_A)^{\mathrm{T}} M^{1/2} M^{1/2} (I-\pi_A) M^{-1/2} v}{v^{\mathrm{T}} v}
\end{aligned}$$

DOI: 10.1002/nla

$$= \frac{1}{1-\rho(E_{TG})} \max_{v\in\mathbb{R}^n} \frac{v^{\mathrm{T}}(I-\pi_A)^{\mathrm{T}} M (I-\pi_A) v}{v^{\mathrm{T}} M v}$$
$$= \frac{1}{1-\rho(E_{TG})} \|I-\pi_A\|_M^2.$$

The result (10) follows directly, using Kato's lemma (e.g. [27, Lemma 3.6]) which implies $\|I-\pi_A\|_M = \|\pi_A\|_M$, since $\pi_A \neq O, I$ by virtue of our general assumptions.

In addition, using (14) together with Lemma 3.1, one also has

$$\begin{aligned}\xi &= \min_{v\in\mathbb{R}^n} \frac{v^{\mathrm{T}} M^{-1} v}{v^{\mathrm{T}}(I-\pi_A)A^{-1}(I-\pi_A)^{\mathrm{T}} v}\\ &\leqslant \min_{v=(I-\pi_A)^{\mathrm{T}} w,\, w\in\mathbb{R}^n} \frac{v^{\mathrm{T}} M^{-1} v}{v^{\mathrm{T}}(I-\pi_A)A^{-1}(I-\pi_A)^{\mathrm{T}} v}\\ &= 1-\rho(E_{TG}),\end{aligned}$$

which gives the first term in the right-hand side of (11).

On the other hand, since

$$v^{\mathrm{T}} A^{1/2} M^{(2\nu)^{-1}} A^{1/2} v = v^{\mathrm{T}} v - v^{\mathrm{T}} (I - A^{1/2} M^{(\nu)^{-1}} A^{1/2})^2 v^{\mathrm{T}} \leqslant v^{\mathrm{T}} v, \quad \forall v \in \mathbb{R}^n,$$

there holds

$$v^{\mathrm{T}} A v \leqslant v^{\mathrm{T}} M v, \quad \forall v \in \mathbb{R}^n.$$

Hence,

$$\begin{aligned}\xi &= \min_{v\in\mathbb{R}^n} \frac{v^{\mathrm{T}} A v}{v^{\mathrm{T}}(I-\pi_A)^{\mathrm{T}} M (I-\pi_A) v}\\ &\leqslant \min_{v\in\mathbb{R}} \frac{v^{\mathrm{T}} M v}{v^{\mathrm{T}}(I-\pi_A)^{\mathrm{T}} M (I-\pi_A) v}\\ &= \frac{1}{\|I-\pi_A\|_M^2},\end{aligned} \tag{15}$$

which, combined with Kato's lemma $\|I-\pi_A\|_M = \|\pi_A\|_M$, gives the second term in the right-hand side of (11). □

Theorem 3.2 shows that McCormick's bound proves a satisfactory convergence rate for the V-cycle if and only if, at each level, the two-grid method converges fast enough and $\|\pi_{A_k}\|_{M_k^{(2\nu)}} = \|M_k^{(2\nu)^{1/2}} \pi_{A_k} M_k^{(2\nu)^{-1/2}}\|$ is nicely bounded. We can further show the following corollary.

Corollary 3.1

Let the assumptions of Theorem 3.2 hold and let $E_{MG}^{(J)}$ be defined by (2).

Then

$$\rho(E_{TG}^{(J)})\leqslant\rho(E_{MG}^{(J)})\leqslant 1-\delta^{(\nu)}\leqslant 1-\min_{1\leqslant k\leqslant J}\frac{1-\rho(E_{TG}^{(k)})}{\|\pi_{A_k}\|^2_{M_k^{(2\nu)}}}. \qquad (16)$$

Proof

The proof of $\rho(E_{TG}^{(k)})\leqslant\rho(E_{MG}^{(k)})$ can be deduced from the relation (7.2.2a) in [22] combined with (7.2.4a) from the same reference, which proves that

$$A^{1/2}E_{MG}^{(k)}A^{-1/2}\leqslant A^{1/2}E_{TG}^{(k)}A^{-1/2}.$$

The other results follow from Theorems 3.1 and 3.2. □

Note that the V-cycle convergence factor is bounded below by the two-grid convergence factor on the finest grid only. Indeed, $\max_{1\leqslant k\leqslant J}\rho(E_{TG}^{(k)})$ can be close to 1 even when $\rho(E_{MG}^{(J)})$ is not, for instance, when the smoother alone is efficient enough on the finest level, so that poor two-grid ingredients on coarser levels will not significantly affect the convergence. In practice, however, one has often $\max_{1\leqslant k\leqslant J}\rho(E_{TG}^{(k)})\approx\rho(E_{TG}^{(J)})$ (e.g. consider the discrete Poisson equation on many simple geometries with uniform meshes). Then (16) defines an interval, containing both $1-\delta^{(\nu)}$ and $\rho(E_{MG}^{(J)})$, that is narrow if and only if $\max_{1\leqslant k\leqslant J}\|\pi_{A_k}\|_{M_k^{(2\nu)}}$ is not much larger than 1.

3.3. Fourier analysis

Often, a multigrid method is assessed by estimating the two-grid convergence rate with Fourier analysis [1–3]. This means that one considers a model constant-coefficient PDE for which the eigenvectors of the discrete matrix are explicitly known at all levels. Simple smoothers have the same set of eigenvectors and, hence, the matrices A_k and R_k are both diagonal whenever expressed in the corresponding basis (the Fourier basis). In more complicated situations, R_k may be only block-diagonal with small diagonal blocks; A_k may also have a block diagonal structure in case of coupled systems of PDEs. Note that $M_k^{(2\nu)}$, expressed in the Fourier basis, will then have the same block diagonal structure as A_k and R_k, and will be pointwise diagonal if A_k and R_k are pointwise diagonal.

Let

$$A_k=\begin{pmatrix}\Lambda_1^{(k)} & & & \\ & \Lambda_2^{(k)} & & \\ & & \ddots & \\ & & & \Lambda_{l_k}^{(k)}\end{pmatrix},\quad M_k^{(2\nu)}=\begin{pmatrix}\Sigma_1^{(k)} & & & \\ & \Sigma_2^{(k)} & & \\ & & \ddots & \\ & & & \Sigma_{l_k}^{(k)}\end{pmatrix}$$

be this (block) diagonal representation of A_k and $M_k^{(2\nu)}$, where the ith block has size $m_i^{(k)}\times m_i^{(k)}$, $i=1,\ldots,l_k$. Technically, the Fourier analysis of a two-grid method at level k characterized by a given prolongation P_{k-1} is possible if there exists a basis of the coarse space (the coarse Fourier basis) such that the expression of P_{k-1} in both this basis and the (fine grid) Fourier basis has the structure

$$P_{k-1}=\begin{pmatrix} p_1^{(k-1)} & & & \\ & p_2^{(k-1)} & & \\ & & \ddots & \\ & & & p_{l_k}^{(k-1)} \end{pmatrix},$$

where $p_i^{(k-1)}$ are (possibly complex) rectangular matrices of size $m_i^{(k)}\times m_i^{(k-1)}$.

Here, we observe that, in this context, $M_k^{(2\nu)^{1/2}}\pi_{A_k}M_k^{(2\nu)^{-1/2}}$ is also block diagonal with diagonal blocks of the form

$$\Sigma_i^{(k)^{1/2}}p_i^{(k-1)}\left(p_i^{(k-1)^H}\Lambda_i^{(k)}p_i^{(k-1)}\right)^{-1}p_i^{(k-1)^H}\Lambda_i^{(k)}\Sigma_i^{(k)^{-1/2}}. \tag{17}$$

Hence, $\|\pi_{A_k}\|^2_{M_k^{(2\nu)}}$ is the maximal norm of all these $m_i^{(k)}\times m_i^{(k)}$ blocks. Further, the matrices (17) are the product of rectangular matrices; taking the product of their norms gives an easy-to-assess upper bound:

$$\|\pi_{A_k}\|_{M_k^{(2\nu)}}\leqslant\max_i\|\Sigma_i^{(k)^{1/2}}p_i^{(k-1)}\|\|(p_i^{(k-1)^H}\Lambda_i^{(k)}p_i^{(k-1)})^{-1}p_i^{(k-1)^H}\Lambda_i^{(k)}\Sigma_i^{(k)^{-1/2}}\|. \tag{18}$$

It is worth noting that the latter inequality becomes an equality when $m_i^{(k-1)}=1$ for all i; that is, when the rectangular blocks $p_i^{(k-1)}$ are all simple vectors, as most often arises when analyzing scalar PDEs.

3.4. *Finite element setting*

Consider a finite element discretization of the Poisson boundary value problem on a bounded domain. Such a domain is first approximated by an appropriate polygonal or polyhedral mesh, which is then refined several times. These refinements naturally induce a multigrid hierarchy (including inter-grid transfer operators P_k). It then can be shown (see [28, Theorem 4.2]) that $\|\pi_{A_k}\|$ are bounded on all levels if and only if the underlying problem possesses (full) elliptic regularity. Since $\|\cdot\|$ behaves similar to $\|\cdot\|_{M_k^{(2\nu)}}$ for a number of smoothers, essentially the same conclusions hold with respect to $\|\pi_{A_k}\|_{M_k^{(2\nu)}}$.

With regards to Theorem 3.2, these observations show that level-independent two-grid convergence implies, in this context, a level-independent bound for V-cycle multigrid if and only if the problem has full elliptic regularity. Hence, it follows that McCormick's analysis cannot prove optimal bounds for the V-cycle if the problem does not possess full regularity. Considering the results in [26], the same conclusions hold for Hackbusch's analysis [22, Section 7.2], and the SSC theory with a-orthogonal decomposition [16, 17]. Thus, for the case when $\|\pi_{A_k}\|$ and $\|\pi_{A_k}\|_{M_k^{(2\nu)}}$ behave similarly with respect to the problem size, we show here that another type of analysis, as developed in, e.g. [13–18], is really needed to get uniform results for the V-cycle for problems with less than full regularity.

DOI: 10.1002/nla

4. EXAMPLES

We consider three examples that represent three possible different practical situations. In the first, both $\rho(E_{TG}^{(k)})$ and $\|\pi_A\|^2_{M^{(2)}}$ are nicely bounded above. In the second example, $\rho(E_{TG}^{(k)})$ remains bounded away from one, whereas $\|\pi_A\|^2_{M^{(2)}}$ increases rapidly with the problem size. The third example is the other way around: $\|\pi_A\|^2_{M^{(2)}}$ is nicely bounded, whereas $\rho(E_{TG}^{(k)})$ is far from being optimal.

4.1. Standard multigrid with 2D Poisson

We consider the linear system resulting from the bilinear finite element discretization of the two-dimensional Poisson problem

$$-\Delta u = f \quad \text{in } \Omega = (0,1)\times(0,1)$$
$$u = 0 \quad \text{in } \partial\Omega$$

on a uniform grid of mesh size $h = 1/N_J$ in both the directions. The matrix corresponds then to the following nine point stencil:

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}. \tag{19}$$

Up to some scaling factor, this is also the stencil obtained with nine-point finite difference discretization. We assume $N_J = 2^J N_0$ for some integer N_0, allowing J steps of regular geometric coarsening. We consider the standard prolongation operator

$$P_k = \begin{pmatrix} J_k \\ I_{n_k} \end{pmatrix},$$

where J_k corresponds to the natural interpolation associated with bilinear finite element basis functions. The restriction P_k^{T} corresponds then to 'full weighting', as defined in, e.g. [1].[§] We consider damped Jacobi smoothing: $R_k = \omega_{\mathrm{Jac}}^{-1}\mathrm{diag}(A_k)$. Since the stencil is preserved on all levels, it is sufficient to consider only two successive grids; to alleviate notation, we therefore let $N = N_k$, $A = A_k$, $R = R_k$, $M = M_k^{(v)}$, $P = P_{k-1}$, $A_c = A_{k-1} = P^{\mathrm{T}} A P$, and $\pi_A = \pi_{A_k} = P A_c^{-1} P^{\mathrm{T}} A$.

We now use the Fourier analysis to asses $\|\pi_A\|_{M^{(2v)}}$ via (18). The eigenvectors of A are, for $i, j = 1, \ldots, N-1$, the functions

$$u_{i,j}^{(N)} = \sin(\mathrm{i}\pi x)\sin(\mathrm{j}\pi y)$$

evaluated at the grid points. The eigenvalue corresponding to $u_{i,j}^{(N)}$ is

$$\lambda_{i,j}^{(N)} = 4(3s_i + 3s_j - 4s_i s_j), \tag{20}$$

[§] Up to some scaling factor, the scalings of the prolongation and restriction are unimportant when using coarse-grid matrices of the Galerkin type.

DOI: 10.1002/nla

where

$$s_i = \sin^2\left(\frac{\mathrm{i}\pi}{2N}\right), \quad s_j = \sin^2\left(\frac{\mathrm{j}\pi}{2N}\right). \tag{21}$$

Hence, the eigenvalues of $I - R^{-1}A$ are in the interval $[1-\omega_{\mathrm{Jac}}\frac{3}{2}, 1)$. One has therefore $\rho(I-R^{-1}A)\leqslant 1$, as required by our general assumptions if $\omega_{\mathrm{Jac}} \in (0, \frac{4}{3})$. The prolongation P satisfies (see, e.g. [1, p. 87])

$$P^{\mathrm{T}} \begin{Bmatrix} u_{i,j}^{(N)} \\ u_{N-i,N-j}^{(N)} \\ -u_{N-i,j}^{(N)} \\ -u_{i,N-j}^{(N)} \end{Bmatrix} = 4 \begin{Bmatrix} (1-s_i)(1-s_j) \\ s_i s_j \\ s_i(1-s_j) \\ (1-s_i)s_j \end{Bmatrix} u_{i,j}^{(N/2)}$$

for $1 \leqslant i, j \leqslant N/2-1$, with $P^{\mathrm{T}} u_{i,j}^{(N)} = 0$ for $i = N/2$ or $j = N/2$. Using

$$p_{i,j} = 4((1-s_i)(1-s_j) \quad s_i s_j \quad s_i(1-s_j) \quad (1-s_i)s_j)^{\mathrm{T}},$$

$$\Lambda_{i,j} = \mathrm{diag}(\lambda_{i,j}^{(N)}, \lambda_{N-i,N-j}^{(N)}, \lambda_{N-i,j}^{(N)}, \lambda_{i,N-j}^{(N)}),$$

$$\Sigma_{i,j}^{(v)} = \mathrm{diag}\left(\left\{ \sigma^{(v)}(\lambda_{c,s}^{(N)}) \,\middle|\, \sigma^{(v)}(\lambda) = \frac{\lambda}{1-\left(1-\frac{\omega_{\mathrm{Jac}}\lambda}{8}\right)^{v}} \right\}_{(c,s)=(i,j),(N-i,N-j),(N-i,j),(i,N-j)}\right),$$

we can rewrite (18):

$$\|\pi_A\|_{M^{(2v)}}^2 = \max_{i,j=1,\ldots,N-1} g^{(v)}(s_i, s_j),$$

where

$$g^{(v)}(s_i, s_j) = \frac{\|\Sigma_{i,j}^{(2v)^{1/2}} p_{i,j}\|^2 \|p_{i,j}^{\mathrm{T}} \Lambda_{i,j} \Sigma_{i,j}^{(2v)^{-1/2}}\|^2}{(p_{i,j}^{\mathrm{T}} \Lambda_{i,j} p_{i,j})^2}. \tag{22}$$

One also has

$$\max_{i,j=1,\ldots,N-1} g^{(v)}(s_i, s_j) \leqslant \sup_{(s_i,s_j)\in(0,1)\times(0,1)} g^{(v)}(s_i, s_j).$$

For all v, $g^{(v)}(s_i, s_m)$ exhibits the following symmetries: $g^{(v)}(s_i, s_j) = g^{(v)}(1-s_i, s_j) = g^{(v)}(s_i, 1-s_j) = g^{(v)}(1-s_i, 1-s_j)$. Further, numerical investigations reveal that the maximum on the considered domain is located at the boundary, i.e. corresponds to, e.g. $s_j = 0$ or, equivalently, $j = 0$ (such index values represent asymptotic behavior and do not correspond to any Fourier block).

DOI: 10.1002/nla

Because of the symmetries, it is sufficient to analyze this latter case. Next, since

$$
\begin{aligned}
& g^{(v)}(s_i,0) \\
&= \frac{\left((p_{i,0})_1^2\sigma^{(2v)}(\lambda_{i,0}^{(N)})+(p_{i,0})_3^2\sigma^{(2v)}(\lambda_{N-i,0}^{(N)})\right)\left(\dfrac{(p_{i,0})_1^2(\lambda_{i,0}^{(N)})^2}{\sigma^{(2v)}(\lambda_{i,0}^{(N)})}+\dfrac{(p_{i,0})_3^2(\lambda_{N-i,0}^{(N)})^2}{\sigma^{(2v)}(\lambda_{N-i,0}^{(N)})}\right)}{((p_{i,0})_1^2\lambda_{i,0}^{(N)}+(p_{i,0})_3^2\lambda_{N-i,0}^{(N)})^2} \\
&= 1+\frac{(p_{i,0})_1^2(p_{i,0})_3^2\left(\dfrac{\sigma^{(2v)}(\lambda_{i,0}^{(N)})}{\sigma^{(2v)}(\lambda_{N-i,0}^{(N)})}(\lambda_{N-i,0}^{(N)})^2+\dfrac{\sigma^{(2v)}(\lambda_{N-i,0}^{(N)})}{\sigma^{(2v)}(\lambda_{i,0}^{(N)})}(\lambda_{i,0}^{(N)})^2-2\lambda_{i,0}^{(N)}\lambda_{N-i,0}^{(N)}\right)}{((p_{i,0})_1^2\lambda_{i,0}+(p_{i,0})_3^2\lambda_{N-i,0})^2} \\
&= 1+s_i(1-s_i)\left(\frac{1-\left(1-\frac{3}{2}\omega_{\text{Jac}}s_i\right)^{2v}}{1-\left(1-\frac{3}{2}\omega_{\text{Jac}}(1-s_i)\right)^{2v}}+\frac{1-\left(1-\frac{3}{2}\omega_{\text{Jac}}(1-s_i)\right)^{2v}}{1-\left(1-\frac{3}{2}\omega_{\text{Jac}}s_i\right)^{2v}}-2\right),
\end{aligned}
\tag{23}
$$

we obtain (see Appendix A for details)

$$
\|\pi_A\|_{M^{(2v)}}^2 \leqslant \sup_{(s_i,s_j)\in(0,1)\times(0,1)} g^{(v)}(s_i,s_j)=\sup_{s_i\in(0,1)} g^{(v)}(s_i,0)\leqslant \begin{cases} 2-\dfrac{3\omega_{\text{Jac}}}{4} & \text{if } v=1 \\[2ex] 1+\dfrac{1}{3v\omega_{\text{Jac}}} & \text{if } v>1 \end{cases}
$$

Note that this bound is asymptotically sharp for $N\to\infty$ when $v=1$, since $\lim_{s\to 0} g^{(1)}(s,0)=2-3\omega_{\text{Jac}}/4$. In Tables I and II, we use this bound and the asymptotically sharp estimate

$$
\delta^{(v)\,-1}\leqslant \frac{1}{3v\omega_{\text{Jac}}}+\frac{1}{1-\left(1-\dfrac{3\omega_{\text{Jac}}}{2}\right)^{2v}}, \quad \forall v=1,2,
$$

obtained in [26] to illustrate inequalities (16), with two-grid and V-cycle multigrid convergence factors numerically assessed for $N_0=2$ and $J=7$ (hence $N=256$). Note that $\rho(E_{TG}^{(k)})$ increases with the mesh size, so that $\max_{1\leqslant k\leqslant J}\rho(E_{TG}^{(k)})$ corresponds to the value on the finest grid, which is close to the asymptotic one. Observe that the interval containing both $\rho(E_{MG}^{(J)})$ and $1-\delta^{(1)}$ is sharp enough. On the other hand, $1-1/\|\pi_A\|_{M^{(2)}}^2$ is also a lower bound on $1-\delta^{(1)}$ by (11), but in general not a lower bound on the effective convergence factor.

DOI: 10.1002/nla

Table I. The estimates of main convergence parameters for $v=1$ and for different damping factors ω_{Jac}.

ω_{Jac}	$1-\frac{1}{\|\pi_A\|^2_{M^{(2)}}}$	$\rho(E_{TG}^{(J)})$	$\rho(E_{MG}^{(J)})$	$1-\delta^{(1)}$	$1-\frac{(1-\rho(E_{TG}^{(J)}))}{\|\pi_A\|^2_{M^{(2)}}}$
$\frac{1}{2}$	0.385	0.391	0.398	0.423	0.625
$\frac{2}{3}$	0.333	0.25	0.271	0.333	0.5
1	0.2	0.25	0.251	0.4	0.4

Table II. The estimates of main convergence parameters for $v=2$ and for different damping factors ω_{Jac}.

ω_{Jac}	$1-\frac{1}{\|\pi_A\|^2_{M^{(4)}}}$	$\rho(E_{TG}^{(J)})$	$\rho(E_{MG}^{(J)})$	$1-\delta^{(2)}$	$1-\frac{(1-\rho(E_{TG}^{(J)}))}{\|\pi_A\|^2_{M^{(4)}}}$
$\frac{1}{2}$	0.25	0.153	0.187	0.252	0.365
$\frac{2}{3}$	0.2	0.083	0.121	0.2	0.266
1	0.143	0.068	0.091	0.189	0.2

4.2. Aggregation-based multigrid for 1D Poisson

We consider $N\times N$ linear system associated with $A=A(\varepsilon)$, where

$$A(\varepsilon)=\begin{pmatrix} 2 & -1 & & \cdots & -1 \\ -1 & 2 & -1 & & \\ & -1 & 2 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & -1 \\ -1 & & \cdots & -1 & 2 \end{pmatrix}+\varepsilon N^{-1} I_N, \tag{24}$$

with $N=2^J N_0$ and $\varepsilon>0$. We also assume piecewise constant prolongation of the form

$$P=\begin{pmatrix} 1 & 1 & & & & & \\ & & 1 & 1 & & & \\ & & & & \ddots & & \\ & & & & & 1 & 1 \end{pmatrix}^{\mathrm{T}}.$$

Note that, with this prolongation, the successive coarse-grid matrices $A_k=A_k(\varepsilon)$ are also given by (24) with N replaced by $N_k=2^k N_0$, where we consider $N_0\geqslant 2$. Hence, we can omit the subscript k (or $k-1$), let $A_c=A_{k-1}=P^{\mathrm{T}}AP$, and set $\pi_A=\pi_{A_k}=PA_c^{-1}P^{\mathrm{T}}A$.

Note that this is a 1D-like problem that could be solved more efficiently using a tri-diagonal solver. The analysis below can however be easily repeated in more dimensions, leading essentially to the same conclusions. We therefore continue with the 1D variant for the sake of simplicity.

DOI: 10.1002/nla

The eigenvectors of $A(\varepsilon)$ are, for $j=0,\ldots,N-1$, the functions

$$u_j^{(N)}=\frac{1}{\sqrt{N}}\exp(\mathrm{i}j\pi x)$$

evaluated at the grid points, with $\mathrm{i}=\sqrt{-1}$. The eigenvalue corresponding to $u_j^{(N)}$ is

$$\lambda_j^{(N)}(\varepsilon)=4\sin^2(\mathrm{j}\pi N^{-1})+\varepsilon N^{-1}.$$

The prolongation P satisfies (see [11, p. 1087])

$$P^{\mathrm{T}}\begin{Bmatrix}u_j^{(N)}\\ u_{j+N/2}^{(N)}\end{Bmatrix}=\sqrt{2}\mathrm{e}^{\mathrm{i}j\pi N^{-1}}\begin{Bmatrix}\cos(\mathrm{j}\pi N^{-1})\\ \mathrm{i}\sin(\mathrm{j}\pi N^{-1})\end{Bmatrix}u_j^{(N/2)}.$$

We consider damped Jacobi smoother $R=2\ \mathrm{diag}(A)$. Hence, the eigenvalues of $I-R^{-1}A$ are in the interval

$$\left[1-\frac{\varepsilon N^{-1}}{4+2\varepsilon N^{-1}},1-\frac{4+\varepsilon N^{-1}}{4+2\varepsilon N^{-1}}\right)=[\omega,1-\omega)\quad\text{with }\omega=\frac{4+\varepsilon N^{-1}}{4+2\varepsilon N^{-1}}\in(0,1).$$

One therefore has $\rho(I-R^{-1}A)\leqslant 1$, as required by our general assumptions.

Letting

$$p_j=\sqrt{2}\mathrm{e}^{\mathrm{i}j\pi N^{-1}}(\cos(\mathrm{j}\pi N^{-1})\mathrm{i}\sin(\mathrm{j}\pi N^{-1}))^H,$$

$$\Lambda_j(\varepsilon)=\mathrm{diag}(\lambda_j^{(N)}(\varepsilon),\lambda_{j+N/2}^{(N)}(\varepsilon)),$$

$$\Sigma_j^{(v)}(\varepsilon)=\mathrm{diag}\left(\left\{\sigma^{(v)}(\lambda_c^{(N)}(\varepsilon))\,\middle|\,\sigma^{(v)}(\lambda)=\frac{\lambda}{1-\left(1-\dfrac{\omega\lambda}{4+\varepsilon N^{-1}}\right)^v}\right\}_{(c)=(j),(j+N/2)}\right),$$

we can rewrite (18) as:

$$\|\pi_A\|_{M^{(2v)}}=\max_{j=0,\ldots,N/2-1}\frac{\|\Sigma_j^{(2v)}(\varepsilon)^{1/2}p_j\|\|p_j^H\Lambda_j(\varepsilon)\Sigma_j^{(2v)}(\varepsilon)^{-1/2}\|}{p_j^H\Lambda_j(\varepsilon)p_j}. \tag{25}$$

First observe that $\sigma^{(2v)}(\lambda)$ is an increasing function of λ since $t(1-(1-t)^{2v})^{-1}$ is an increasing function of t on the interval $(0,1)$. Hence, since $\lambda_1^{(N)}(\varepsilon)\leqslant\lambda_{1+N/2}^{(N)}(\varepsilon)$ for $N\geqslant 2N_0\geqslant 4$, we have

$$\begin{aligned}\|\pi_A\|_{M^{(2v)}}&\geqslant\frac{\|\Sigma_1^{(2v)}(\varepsilon)^{1/2}p_1\|\|p_1^H\Lambda_1(\varepsilon)\Sigma_1^{(2v)}(\varepsilon)^{-1/2}\|}{p_1^H\Lambda_1(\varepsilon)p_1}\\&=\frac{\sqrt{|(p_1)_1|^2\dfrac{\sigma^{(2v)}(\lambda_1^{(N)}(\varepsilon))}{\sigma^{(2v)}(\lambda_{1+N/2}^{(N)}(\varepsilon))}+|(p_1)_2|^2}\sqrt{|(p_1)_1|^2\lambda_1^{(N)}(\varepsilon)^2\dfrac{\sigma^{(2v)}(\lambda_{1+N/2}^{(N)}(\varepsilon))}{\sigma^{(2v)}(\lambda_1^{(N)}(\varepsilon))}+|(p_1)_2|^2\lambda_{1+N/2}^{(N)}(\varepsilon)^2}}{|(p_1)_1|^2\lambda_1^{(N)}(\varepsilon)+|(p_1)_2|^2\lambda_{1+N/2}^{(N)}(\varepsilon)}\end{aligned}$$

DOI: 10.1002/nla

$$\geqslant \sqrt{\frac{\sigma^{(2v)}(\lambda_1^{(N)}(\varepsilon))}{\sigma^{(2v)}(\lambda_{1+N/2}^{(N)}(\varepsilon))}} \frac{\sqrt{|(p_1)_1|^2+|(p_1)_2|^2}\sqrt{|(p_1)_1|^2\lambda_1^{(N)}(\varepsilon)^2+|(p_1)_2|^2\lambda_{1+N/2}^{(N)}(\varepsilon)^2}}{|(p_1)_1|^2\lambda_1^{(N)}(\varepsilon)+|(p_1)_2|^2\lambda_{1+N/2}^{(N)}(\varepsilon)}$$

$$= \sqrt{\frac{\sigma^{(2v)}(\lambda_1^{(N)}(\varepsilon))}{\sigma^{(2v)}(\lambda_{1+N/2}^{(N)}(\varepsilon))}} \frac{\sqrt{\cos^4(\mathrm{j}\pi N^{-1})\sin^2(\mathrm{j}\pi N^{-1})+\cos^2(\mathrm{j}\pi N^{-1})\sin^4(\mathrm{j}\pi N^{-1})+\mathcal{O}(\varepsilon)}}{2\cos^2(\mathrm{j}\pi N^{-1})\sin^2(\mathrm{j}\pi N^{-1})+\mathcal{O}(\varepsilon)}.$$

Further, using again the monotonicity of $\sigma^{(2v)}$, there holds

$$\frac{\sigma^{(2v)}(\lambda_1^{(N)}(\varepsilon))}{\sigma^{(2v)}(\lambda_{1+N/2}^{(N)}(\varepsilon))} \geqslant \frac{\lim_{\lambda\to 0}\sigma^{(2v)}(\lambda)}{\sigma^{(2v)}(4+\varepsilon N^{-1})} = \frac{(4+\varepsilon N^{-1})}{v\omega}\frac{1-(1-\omega)^{2v}}{(4+\varepsilon N^{-1})} = \frac{1-(1-\omega)^{2v}}{v\omega}$$

with $\omega\in(0,1)$. Hence, for $\varepsilon\to 0$, we have

$$\|\pi_A\|^2_{M^{(2v)}} \geqslant \frac{1-(1-\omega)^{2v}}{v\omega}\frac{1}{4\cos^2(\mathrm{j}\pi N^{-1})\sin^2(\mathrm{j}\pi N^{-1})} = \mathcal{O}(N^2).$$

Thus, $\|\pi_A\|^2_{M_k^{(2)}}$ increases with the problem size when ε is small enough, whereas, as shown in [11], the two-grid convergence factor remains bounded. Hence, we have an example of optimal two-grid method for which the V-cycle convergence estimate is poor. As seen in Table III, it turns out that the actual convergence factor also deteriorates with the number of levels, showing that the analysis based on $\|\pi_A\|^2_{M^{(2)}}$ is qualitatively correct.

4.3. Positive off-diagonal entries

We consider the $(2N_J-1)\times(2N_J-1)$ matrix

$$A = \begin{pmatrix} 2 & 1 & & & \\ 1 & 2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 2 & 1 \\ & & & 1 & 2 \end{pmatrix},$$

Table III. The values of main parameters for $\varepsilon=10^{-4}$ and for different problem sizes; the coarsest grid corresponds to $N_0=4$.

$J(N)$	1(8)	3(32)	5(128)	7(512)	9(2048)
$\|\pi_A\|^2_{M_J^{(2)}}$	1.471	13.58	208.0	3312	52 575
$\rho(E_{TG}^{(J)})$	0.375	0.490	0.499	0.5	0.5
$\rho(E_{MG}^{(J)})$	0.375	0.800	0.947	0.986	0.997

DOI: 10.1002/nla

Table IV. The values of main parameters for different problem sizes; the coarsest grid corresponds to $N_0 = 2$.

$J(N)$	1(4)	3(16)	5(64)	7(256)	9(1024)
$\|\pi_A\|^2_{M_J^{(2)}}$	1.235	1.479	1.498	1.5	1.5
$\rho(E_{TG}^{(J)})$	0.625	0.971	0.998	0.9999	0.99999
$\rho(E_{MG}^{(J)})$	0.625	0.971	0.998	0.9999	0.99999

with $N_k = N_0 \cdot 2^k$, corresponding to the one-dimensional stencil

$$[1 \quad 2 \quad 1]. \tag{26}$$

We also consider the $(2N_k - 1) \times (N_k - 1)$ prolongation matrix

$$P_k = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 1 & & & & & \\ & & 1 & 0 & 1 & & & \\ & & & & & \ddots & & \\ & & & & & 1 & 0 & 1 \end{pmatrix}^{\mathrm{T}} \tag{27}$$

and the damped Jacobi smoother $R_k = \frac{1}{2}\mathrm{diag}(A_k)$ with one pre-smoothing step and one post-smoothing step at each level. Note that the stencil (26) is preserved on all levels.

The values of $\|\pi_A\|^2_{M_J^{(2)}}$ and $\rho(E_{TG}^{(J)})$ on the finest grid, which are also the maximal values of these parameters over all grids, are given in Table IV together with the V-cycle convergence factor $\rho(E_{MG}^{(J)})$.

This example illustrates that $\|\pi_{A_k}\|^2_{M_k^{(2\nu)}}$ is a parameter essentially independent of $\rho(E_{TG}^{(k)})$, since it remains nicely bounded while both the two-grid and the V-cycle convergence factor deteriorate rapidly with the problem size.

APPENDIX A

In this appendix, we outline the proof of the following inequality:

$$\sup_{s_i \in (0,1)} g^{(\nu)}(s_i, 0) \leqslant \begin{cases} 2 - \dfrac{3\omega_{\mathrm{Jac}}}{4} & \text{if } \nu = 1 \\ 1 + \dfrac{1}{3\nu\omega_{\mathrm{Jac}}} & \text{if } \nu > 1, \end{cases} \tag{A1}$$

with $g^{(\nu)}$ defined by (23) and $\omega_{\mathrm{Jac}} \in [0, \frac{4}{3})$.

DOI: 10.1002/nla

Note that $g^{(v)}(s_i,0)=g^{(v)}(1-s_i,0)$ and it is sufficient to seek a supremum for $s_i\in(0,0.5)$. Next, exchanging $(\frac{3}{2})\omega_{\text{Jac}}$ for α (hence, $\alpha\in[0,2)$), one has

$$\begin{aligned}
g^{(v)}(s_i,0) &= 1+s_i(1-s_i)\left(\left[\frac{1-(1-\alpha s_i)^{2v}}{1-(1-\alpha(1-s_i))^{2v}}-1\right]+\left[\frac{1-(1-\alpha(1-s_i))^{2v}}{1-(1-\alpha s_i)^{2v}}-1\right]\right)\\
&= 1+s_i(1-s_i)[(1-\alpha s_i)^{2v}-(1-\alpha(1-s_i))^{2v}]\\
&\quad\times\left(\frac{1}{1-(1-\alpha s_i)^{2v}}-\frac{1}{1-(1-\alpha(1-s_i))^{2v}}\right)\\
&\leqslant 1+s_i(1-s_i)[(1-\alpha s_i)^{2v}-(1-\alpha(1-s_i))^{2v}]\left(\frac{1}{1-(1-\alpha s_i)^{2v}}\right)\\
&\leqslant 1+s_i(1-\alpha s_i)^{2v}\left(\frac{1}{1-(1-\alpha s_i)^{2v}}\right)\\
&= 1+\frac{(1-\alpha s_i)^{2v}}{\alpha\sum_{k=0}^{2v-1}(1-\alpha s_i)^k}\\
&\leqslant 1+\frac{1}{2v\alpha},
\end{aligned}\tag{A2}$$

the last inequality coming from the fact that $\alpha s_i\in[0,1)$. This proves (A1) for $v>1$.

On the other hand, if $v=1$, (A2) further gives

$$\begin{aligned}
g^{(1)}(s_i,0) &\leqslant 1+s_i(1-s_i)[(1-\alpha s_i)^2-(1-\alpha(1-s_i))^2]\left(\frac{1}{1-(1-\alpha s_i)^2}\right)\\
&= 1+s_i(1-s_i)[\alpha(2-\alpha)(1-2s_i)]\left(\frac{1}{\alpha s_i(2-\alpha s_i)}\right)\\
&= 1+(2-\alpha)(1-2s_i)\left(\frac{1}{\alpha}-\frac{2-\alpha}{\alpha(2-\alpha s_i)}\right) && \text{(A3)}\\
&\leqslant 1+(2-\alpha)\left(\frac{1}{\alpha}-\frac{2-\alpha}{2\alpha}\right) && \text{(A4)}\\
&= 2-\frac{\alpha}{2},
\end{aligned}$$

where the inequality (A4) comes from the fact that the expression (A3) is a decreasing function of s_i. This concludes the proof.

ACKNOWLEDGEMENTS

We thank an anonymous referee for the enlightened comments that form the basis of Section 3.4.

DOI: 10.1002/nla

REFERENCES

1. Trottenberg U, Oosterlee CW, Schüller A. *Multigrid*. Academic Press: London, 2001.
2. Stüben K, Trottenberg KU. Multigrid methods: fundamental algorithms, model problem analysis and applications. In *Multigrid Methods*, Hackbusch W, Trottenberg U (eds). Lectures Notes in Mathematics, vol. 960. Springer: Berlin, Heidelberg, New York, 1982; 1–176.
3. Wienands R, Joppich W. *Practical Fourier Analysis for Multigrid Methods*. Chapman & Hall/CRC Press: Boca Raton, FL, 2005.
4. Brandt A. Algebraic multigrid theory: the symmetric case. *Applied Mathematics and Computation* 1986; **19**:23–56.
5. Falgout RD, Vassilevski PS. On generalizing the algebraic multigrid framework. *SIAM Journal on Numerical Analysis* 2005; **42**:1669–1693.
6. Falgout RD, Vassilevski PS, Zikatanov LT. On two-grid convergence estimates. *Numerical Linear Algebra with Applications* 2005; **12**:471–494.
7. Notay Y. Algebraic multigrid and algebraic multilevel methods: a theoretical comparison. *Numerical Linear Algebra with Applications* 2005; **12**:419–451.
8. Stüben K. An introduction to algebraic multigrid. In *Multigrid*, Trottenberg U, Oosterlee CW, Schiiller A (eds). Academic Press: London, 2001; 413–532. Appendix A.
9. Braess D. *Finite Elements*: *Theory*, *Fast Solvers*, *and Applications in Solid Mechanics*. Cambridge University Press: Cambridge, 1997.
10. Notay Y. Convergence analysis of perturbed two-grid and multigrid methods. *SIAM Journal on Numerical Analysis* 2007; **45**:1035–1044.
11. Muresan AC, Notay Y. Analysis of aggregation-based multigrid. *SIAM Journal on Scientific Computing* 2008; **30**:1082–1103.
12. Bramble JH, Pasciak JE, Wang J, Xu J. Convergence estimates for multigrid algorithms without regularity assumptions. *Mathematics of Computation* 1991; **57**:23–45.
13. Oswald P. *Multilevel Finite Element Approximation*: *Theory and Applications*. Teubner Skripte zur Numerik Teubner: Stuttgart, 1994.
14. Oswald P. Subspace correction methods and multigrid theory. In *Multigrid*, Trottenberg U, Oosterlee CW, Schiiller A (eds). Academic Press: London, 2001; 533–572. Appendix A.
15. Griebel M, Oswald P. On the abstract theory of additive and multiplicative Schwarz algorithms. *Numerische Mathematik* 1995; **70**:163–180.
16. Xu J. Iterative methods by space decomposition and subspace correction. *SIAM Review* 1992; **34**:581–613.
17. Yserentant H. Old and new convergence proofs for multigrid methods. *Acta Numerica* 1993; **2**:285–326.
18. Xu J, Zikatanov LT. The method of alternating projections and the method of subspace corrections in Hilbert space. *Journal of the American Mathematical Society* 2002; **15**:573–597.
19. Braess D. The convergence rate of a multigrid method with Gauss–Seidel relaxation for the Poisson equation. In *Multigrid Methods*, Hackbusch W, Trottenberg U (eds). Lectures Notes in Mathematics, vol. 960. Springer: Berlin, Heidelberg, New York, 1982; 368–386.
20. Braess D, Hackbusch W. A new convergence proof for the multigrid method including the V-cycle. *SIAM Journal on Numerical Analysis* 1983; **20**:967–975.
21. Hackbusch W. Multi-grid convergence theory. In *Multigrid Methods*, Hackbusch W, Trottenberg U (eds). Lectures Notes in Mathematics, vol. 960. Springer: Berlin, Heidelberg, New York, 1982; 177–219.
22. Hackbusch W. *Multi-grid Methods and Applications*. Springer: Berlin, 1985.
23. Mandel J, McCormick SF, Ruge JW. An algebraic theory for multigrid methods for variational problems. *SIAM Journal on Numerical Analysis* 1988; **25**:91–110.
24. McCormick SF. Multigrid methods for variational problems: general theory for the V-cycle. *SIAM Journal on Numerical Analysis* 1985; **22**:634–643.
25. Ruge JW, Stüben K. Algebraic multigrid (AMG). In *Multigrid Methods*. McCormick SF (ed.). Frontiers in Applied Mathematics, vol. 3. SIAM: Philadelphia, PA, 1987; 73–130.
26. Napov A, Notay Y. Comparison of bounds for V-cycle multigrid. *Applied Numerical Mathematics* 2009. Published online on ScienceDirect. DOI: 10.1016/j.apnum.2009.11.003.
27. Vassilevski PS. *Multilevel Block Factorization Preconditioners*. Springer: New York, 2008.
28. Xu J. Theory of multilevel methods. *Ph.D. Thesis*, Cornell University, 1989.