SIAM J. NUMER. ANAL.
Vol. 45, No. 3, pp. 1035–1044

CONVERGENCE ANALYSIS OF PERTURBED TWO-GRID AND MULTIGRID METHODS*

YVAN NOTAY†

Abstract. We consider multigrid methods for symmetric positive definite linear systems. We present a new algebraic convergence analysis of two-grid schemes with inexact solution of the coarse grid system. This analysis allows us to bound the convergence factor of such perturbed two-grid schemes, assuming only a certain bound on the convergence factor for the unperturbed scheme (with exact solution of the coarse grid system). Applied to multigrid methods with the standard W-cycle, this analysis shows that if the convergence factor of the (unperturbed) two-grid method is uniformly bounded by $\sigma < 1/2$, then the convergence factor of the multigrid method is uniformly bounded by $\sigma/(1-\sigma)$. The analysis is purely algebraic and requires only that pre- and postsmoothing are applied in a symmetric way. It covers both geometric and algebraic multigrid methods, and the coarse grid matrix may be of any type (not necessarily Galerkin).

Key words. multigrid, convergence analysis, linear systems, W-cycle, condition number

AMS subject classifications. 65F10, 65N55

DOI. 10.1137/060652312

1. Introduction.

We consider multigrid methods for solving symmetric positive definite (SPD) $n \times n$ linear systems

$$A\,\mathbf{u} \;=\; \mathbf{b}. \tag{1.1}$$

We focus on symmetric multigrid schemes, more precisely on methods for which the basic two-grid cycle is defined as follows:

- Relax ν times on $A\,\mathbf{u} = \mathbf{b}$ using a smoother R;
  we assume that R is an $n \times n$ matrix such that $\rho(I - R\,A) < 1$,
  where $\rho(\cdot)$ stands for the spectral radius;
  ν (the number of pre- and postsmoothing steps) is a given positive integer.
- Perform the coarse grid correction: $\mathbf{u} \leftarrow \mathbf{u} + p\,A_C^{-1} p^T(\mathbf{b} - A\,\mathbf{u})$;
  we assume that A_C (the coarse grid matrix) is an $n_c \times n_c$ SPD matrix,
  where $n_c \le n$ is the number of coarse variables;
  p (the prolongation) is an $n \times n_c$ matrix.
- Relax ν times on $A\,\mathbf{u} = \mathbf{b}$ using R^T.

Note that we do not assume any specific form for the coarse grid matrix.

In this paper, we analyze the influence of the perturbations that arise when the coarse grid systems are solved approximately. More precisely, we consider schemes as above in which A_C^{-1} is exchanged for some $n_c \times n_c$ SPD matrix K_C that approximates it. Our main result relates the convergence of this perturbed two-grid method with that of the "ideal" two-grid method that requires the inversion of A_C.

To express this relation, consider the iteration matrices that govern the convergence of the schemes described above:

$$T_{\mathrm{TG}} \;=\; (I - R^T A)^{\nu}\,(I - p\,A_C^{-1}\,p^T A)\,(I - R\,A)^{\nu} \tag{1.2}$$

*Received by the editors February 17, 2006; accepted for publication (in revised form) December 19, 2006; published electronically May 7, 2007. This work was supported by the Belgian FNRS (Maître de Recherches).
http://www.siam.org/journals/sinum/45-3/65231.html

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

for the unperturbed two-grid cycle and

$$T_{\mathrm{PTG}} \;=\; (I - R^T A)^{\nu}\,(I - p\,K_C\,p^T A)\,(I - R\,A)^{\nu} \tag{1.3}$$

for the perturbed one; see, e.g., [19, p. 40]. As is well known, performing m iterations implies that

$$\hat{\mathbf{u}} - \mathbf{u}_{\mathrm{TG}}^{(m)} = \big(T_{\mathrm{TG}}\big)^m\big(\hat{\mathbf{u}} - \mathbf{u}_{\mathrm{TG}}^{(0)}\big) \qquad \text{or} \qquad \hat{\mathbf{u}} - \mathbf{u}_{\mathrm{PTG}}^{(m)} = \big(T_{\mathrm{PTG}}\big)^m\big(\hat{\mathbf{u}} - \mathbf{u}_{\mathrm{PTG}}^{(0)}\big)$$

(respectively), where $\hat{\mathbf{u}} = A^{-1}\mathbf{b}$ is the exact solution to (1.1). Hence, the asymptotic convergence rate is equal to the spectral radius of the iteration matrix, which is referred to as the *convergence factor*. On the other hand, both the perturbed and the unperturbed two-grid cycle implicitly define a preconditioner, which we denote B_{TG} and B_{PTG}, respectively. They are related to the iteration matrices by

$$I - B_{\mathrm{TG}}^{-1} A \;=\; T_{\mathrm{TG}} \qquad \text{and} \qquad I - B_{\mathrm{PTG}}^{-1} A \;=\; T_{\mathrm{PTG}}. \tag{1.4}$$

Note that because A, B_{TG}, and B_{PTG} are SPD (see below), the eigenvalues of $B_{\mathrm{TG}}^{-1}A$ and $B_{\mathrm{PTG}}^{-1}A$ are real and the convergence factors satisfy

$$\rho\,(T_{\mathrm{TG}}) = \max\big(\lambda_{\max}(B_{\mathrm{TG}}^{-1}A) - 1,\ 1 - \lambda_{\min}(B_{\mathrm{TG}}^{-1}A)\big),$$
$$\rho\,(T_{\mathrm{PTG}}) = \max\big(\lambda_{\max}(B_{\mathrm{PTG}}^{-1}A) - 1,\ 1 - \lambda_{\min}(B_{\mathrm{PTG}}^{-1}A)\big),$$

where $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ stand for the largest and the smallest eigenvalue, respectively. In section 2, we prove

$$\lambda_{\max}(B_{\mathrm{PTG}}^{-1}\,A) \le \lambda_{\max}(B_{\mathrm{TG}}^{-1}\,A) \;\cdot\; \max(\lambda_{\max}(K_C\,A_C),\,1), \tag{1.5}$$
$$\lambda_{\min}(B_{\mathrm{PTG}}^{-1}\,A) \ge \lambda_{\min}(B_{\mathrm{TG}}^{-1}\,A) \;\cdot\; \min(\lambda_{\min}(K_C\,A_C),\,1). \tag{1.6}$$

Multigrid cycles are obtained when a two-grid method is used recursively, exchanging the solution of the coarse grid system for a given number γ of two-grid cycles on the coarser level, and so on, until the coarsest level on which an exact solve is performed. Multigrid cycles are thus particular cases of perturbed two-grid cycles, and the above results enable us to analyze them. For the so-called W-cycle (which corresponds to $\gamma = 2$), we show in section 3 that if $\sigma < 1/2$ is a uniform (i.e., holding at every level) bound on the convergence factor of the unperturbed two-grid method, then the convergence factor of the multigrid method is bounded by $\sigma/(1-\sigma)$.

This improves the state of the art in multigrid convergence theory. Analyzing the two-grid convergence factor is often sufficient to assess the convergence of a multigrid scheme; see, e.g., [19, p. 77]. However, this is not yet completely supported by theoretical results. The standard algebraic analysis considers the multigrid iteration matrix as a two-grid iteration matrix plus some perturbation term; see [11, sect. 4.2] or [19, Thm. 3.2.1]. It allows us to obtain a useful bound on the convergence factor of the multigrid method with the W-cycle if σ satisfies

$$\sigma \;\le\; \frac{1}{4\,C}, \tag{1.7}$$

where C is a constant whose exact value is difficult to predict, except that it is in general not smaller than 1 (see section 3 below for details). Because of this condition, this result on multigrid convergence is sometimes stated as follows: "if the two-grid method converges *sufficiently well*, then the multigrid method with W-cycle will have similar convergence properties" [19, p. 77]. However, condition (1.7) may be violated when textbook multigrid efficiency is difficult to achieve. It may also be difficult to check when using an algebraic multigrid (AMG) method. Below we also show that, when both our new analysis and the standard algebraic analysis apply, our bound on the convergence factor is generally sharper.

For the case of Galerkin coarse grid matrices (i.e., assuming $A_C = p^T A\, p$), an interesting analysis of the W-cycle multigrid has been developed by Braess in [6, pp. 226–228]. This analysis is based on two-grid schemes without postsmoothing, which also gives a worst case estimate for the general case. As will be seen in section 3, our bound on the convergence factor for the W-cycle is always equal to the square of Braess's bound. This suggests that, as proved for the two-grid case in [15, eq. (41)], the W-cycle multigrid scheme with symmetrized pre- and postsmoothing can converge twice as fast as the corresponding scheme without postsmoothing.

On the other hand, in the SPD case, it is also possible to prove optimal convergence properties (with respect to the number of levels) of multigrid methods via so-called smoothing and approximation properties or via the theory of subspace correction methods (using the multilevel splitting of finite element spaces); see, e.g., [5, 7, 11, 12, 13, 14, 16, 17, 22, 23]. However, bounds derived in this way do not, in general, give satisfactorily sharp predictions of actual multigrid convergence [19, p. 96]. Moreover, they require assumptions that are more restrictive than just the convergence of the two-grid method. To our knowledge, for instance, these assumptions have not yet been checked for AMG methods. These analyses, nevertheless, play an important role in the multigrid convergence theory, complementary to the algebraic approach developed here. Indeed, they cover the V-cycle, for which both the standard analysis mentioned above and our new analysis fail to deliver bounds independent of the number of levels.

Eventually, it should be noted that our bounds (1.5), (1.6) have a wider scope than just the analysis of standard multigrid cycles. First, these relations are similar to the ones holding for AMLI-type methods [2, 3]. Hence, when the two-grid method does not converge fast enough for the standard W-cycle, it is possible to use polynomially accelerated cycles based on Chebyshev polynomials, as considered in these references [21]. Another potential application lies in the simplification of coarse grid matrices: (1.5), (1.6) indeed show that one may replace a given coarse grid matrix by a spectrally equivalent approximation. For instance, the theory of algebraic two-grid methods heavily relies on the use of Galerkin coarse grid matrices, that is, on the assumption that $A_C = p^T A\, p$ [8, 9, 10, 15, 18]. However, in practice, such matrices may be costly to compute, and so it could be interesting to develop cheaper alternatives.

The remainder of this paper is organized as follows. In section 2, we prove the main inequalities (1.5), (1.6). Their application to multigrid cycles is discussed in section 3.

2. Perturbed two-grid methods.

We first show that B_{TG} is the Schur complement of an extended matrix $\widehat{B}_{\mathrm{TG}}$ given in factored form. Note that this holds for any SPD coarse grid matrix A_C; hence, similarly, B_{PTG} is the Schur complement of an extended matrix $\widehat{B}_{\mathrm{PTG}}$. This factored form has been inspired by the factored form existing for the preconditioner defined by the so-called hierarchical basis multigrid method[1] [4]. Our derivation is also related to equation (15) in [10].

[1] This latter method does not fit into our framework because only fine grid unknowns are relaxed during smoothing steps, and hence R is singular.

Let us introduce some notation. Let M be the matrix such that

$$I - M^{-1}A \;=\; (I - R\,A)^{\nu}; \tag{2.1}$$

from the assumption, $\rho(I - R\,A) < 1$, and M exists, is invertible, and is such that $\rho(I - M^{-1}A) < 1$. This latter relation implies that

$$Q \;= M^{-1} + M^{-T} - M^{-T}\,A\,M^{-1} \;=\; M^{-T}\left(M + M^{T} - A\right)M^{-1} \tag{2.2}$$

is positive definite; see [9, 15].

Define

(2.3)
$$\widehat{B}_{\mathrm{TG}} \;=\; \begin{pmatrix} I_{n\times n} & 0 \\ -p^{T}(I - A\,M^{-1}) & I_{n_c\times n_c} \end{pmatrix} \begin{pmatrix} Q^{-1} & 0 \\ 0 & A_C \end{pmatrix} \begin{pmatrix} I_{n\times n} & -(I - M^{-T}A)p \\ 0 & I_{n_c\times n_c} \end{pmatrix}.$$

Straightforward calculation shows that

$$\begin{aligned}
\widehat{B}_{\mathrm{TG}}^{\;-1} &= \begin{pmatrix} I_{n\times n} & (I - M^{-T}A)p \\ 0 & I_{n_c\times n_c} \end{pmatrix} \begin{pmatrix} Q & 0 \\ 0 & A_C^{-1} \end{pmatrix} \begin{pmatrix} I_{n\times n} & 0 \\ p^{T}(I - A\,M^{-1}) & I_{n_c\times n_c} \end{pmatrix} \\
&= \begin{pmatrix} Q + (I - M^{-T}A)p\,A_C^{-1}\,p^{T}(I - A\,M^{-1}) & (I - M^{-T}A)p\,A_C^{-1} \\ A_C^{-1}\,p^{T}(I - A\,M^{-1}) & A_C^{-1} \end{pmatrix} \\
&= \begin{pmatrix} B_{\mathrm{TG}}^{-1} & (I - M^{-T}A)p\,A_C^{-1} \\ A_C^{-1}\,p^{T}(I - A\,M^{-1}) & A_C^{-1} \end{pmatrix}.
\end{aligned}$$

Because $\widehat{B}_{\mathrm{TG}}$ is SPD, this first shows that B_{TG} is SPD too. Further, the inverse of B_{TG} is a principal submatrix of the inverse of $\widehat{B}_{\mathrm{TG}}$ if and only if B_{TG} is equal to the corresponding Schur complement in $\widehat{B}_{\mathrm{TG}}$; see, e.g., [1, eq. (3.4), p. 93]. That is, considering the 2×2 block form

$$\widehat{B}_{\mathrm{TG}} \;=\; \begin{pmatrix} (\widehat{B}_{\mathrm{TG}})_{FF} & (\widehat{B}_{\mathrm{TG}})_{FC} \\ (\widehat{B}_{\mathrm{TG}})_{CF} & (\widehat{B}_{\mathrm{TG}})_{CC} \end{pmatrix}$$

(where $(\widehat{B}_{\mathrm{TG}})_{FF}$ is $n\times n$ and $(\widehat{B}_{\mathrm{TG}})_{CC}$ is $n_c\times n_c$), B_{TG} is the Schur complement of $\widehat{B}_{\mathrm{TG}}$ with respect to the bottom right block:

$$B_{\mathrm{TG}} \;=\; (\widehat{B}_{\mathrm{TG}})_{FF} - (\widehat{B}_{\mathrm{TG}})_{FC}\,(\widehat{B}_{\mathrm{TG}})_{CC}^{-1}\,(\widehat{B}_{\mathrm{TG}})_{CF}.$$

This, together with Theorem 3.8 in [1], proves the following lemma.

LEMMA 2.1. *Let B_{TG} be defined by* (1.2), (1.4) *with A, R, p, A_C satisfying the assumptions stated in section* 1. *Let $\widehat{B}_{TG}$ be defined by* (2.3) *with M defined by* (2.1) *and Q defined by* (2.2). *B_{TG} is SPD, and one has, for all $\mathbf{z}\in\Re^{n}$,*

$$\mathbf{z}^{T}\,B_{TG}\,\mathbf{z} \;=\; \min_{\mathbf{w}_C\in\Re^{n_c}} \begin{pmatrix} \mathbf{z}^{T} & \mathbf{w}_C^{T} \end{pmatrix} \widehat{B}_{TG} \begin{pmatrix} \mathbf{z} \\ \mathbf{w}_C \end{pmatrix}. \tag{2.4}$$

Moreover,

$$\lambda_{\max}(B_{TG}^{-1}A) = \max_{\mathbf{z}\in\Re^{n}\setminus\{0\}} \; \max_{\mathbf{w}_C\in\Re^{n_c}} \; \frac{\mathbf{z}^{T}\,A\,\mathbf{z}}{\begin{pmatrix} \mathbf{z}^{T} & \mathbf{w}_C^{T} \end{pmatrix} \widehat{B}_{TG} \begin{pmatrix} \mathbf{z} \\ \mathbf{w}_C \end{pmatrix}}, \tag{2.5}$$

$$\lambda_{\min}(B_{TG}^{-1}A) = \min_{\mathbf{z}\in\Re^{n}\setminus\{0\}} \; \max_{\mathbf{w}_C\in\Re^{n_c}} \; \frac{\mathbf{z}^{T}\,A\,\mathbf{z}}{\begin{pmatrix} \mathbf{z}^{T} & \mathbf{w}_C^{T} \end{pmatrix} \widehat{B}_{TG} \begin{pmatrix} \mathbf{z} \\ \mathbf{w}_C \end{pmatrix}}. \tag{2.6}$$

We now prove (1.5), (1.6). Note that these inequalities may also be proved from the factored form of two-grid preconditioners as obtained by Vassilevski in [20, 21]. Their use is also implicit in the building of AMLI-cycle multigrid developed independently by the same author [21].

THEOREM 2.2. *Let B_{TG}, B_{PTG} be defined by* (1.2), (1.3), (1.4) *with A, R, p, A_C, K_C satisfying the assumptions stated in section* 1. *Inequalities* (1.5) *and* (1.6) *hold.*

Proof. Let M, Q, $\widehat{B}_{\mathrm{TG}}$ be defined by (2.1), (2.2), (2.3), and define $\widehat{B}_{\mathrm{PTG}}$ similarly to $\widehat{B}_{\mathrm{TG}}$, exchanging A_C for K_C^{-1} in (2.3). Lemma 2.1 yields

$$\begin{aligned}
\lambda_{\max}(B_{\mathrm{PTG}}^{-1} A) &= \max_{\mathbf{z}\in\Re^n\setminus\{0\}} \max_{\mathbf{w}_C\in\Re^{n_c}} \frac{\mathbf{z}^T A\,\mathbf{z}}{\begin{pmatrix}\mathbf{z}^T & \mathbf{w}_C^T\end{pmatrix} \widehat{B}_{\mathrm{PTG}} \begin{pmatrix}\mathbf{z}\\ \mathbf{w}_C\end{pmatrix}} \\
&\le \max_{\mathbf{z}\in\Re^n\setminus\{0\}} \max_{\mathbf{w}_C\in\Re^{n_c}} \frac{\mathbf{z}^T A\,\mathbf{z}}{\begin{pmatrix}\mathbf{z}^T & \mathbf{w}_C^T\end{pmatrix} \widehat{B}_{\mathrm{TG}} \begin{pmatrix}\mathbf{z}\\ \mathbf{w}_C\end{pmatrix}} \cdot \max_{\widehat{\mathbf{z}}\in\Re^{n+n_c}\setminus\{0\}} \frac{\widehat{\mathbf{z}}^T \widehat{B}_{\mathrm{TG}}\,\widehat{\mathbf{z}}}{\widehat{\mathbf{z}}^T \widehat{B}_{\mathrm{PTG}}\,\widehat{\mathbf{z}}} \\
&= \lambda_{\max}(B_{\mathrm{TG}}^{-1}A) \cdot \max_{\widehat{\mathbf{w}}\in\Re^{n+n_c}\setminus\{0\}} \frac{\widehat{\mathbf{w}}^T \begin{pmatrix} Q^{-1} & 0\\ 0 & A_C\end{pmatrix}\widehat{\mathbf{w}}}{\widehat{\mathbf{w}}^T \begin{pmatrix} Q^{-1} & 0\\ 0 & K_C^{-1}\end{pmatrix}\widehat{\mathbf{w}}} \\
&= \lambda_{\max}(B_{\mathrm{TG}}^{-1}A) \cdot \max(\lambda_{\max}(K_C\,A_C),\,1).
\end{aligned}$$

Similarly, one finds

$$\begin{aligned}
\lambda_{\min}(B_{\mathrm{PTG}}^{-1} A) &= \min_{\mathbf{z}\in\Re^n\setminus\{0\}} \max_{\mathbf{w}_C\in\Re^{n_c}} \frac{\mathbf{z}^T A\,\mathbf{z}}{\begin{pmatrix}\mathbf{z}^T & \mathbf{w}_C^T\end{pmatrix} \widehat{B}_{\mathrm{PTG}} \begin{pmatrix}\mathbf{z}\\ \mathbf{w}_C\end{pmatrix}} \\
&\ge \min_{\mathbf{z}\in\Re^n\setminus\{0\}} \max_{\mathbf{w}_C\in\Re^{n_c}} \frac{\mathbf{z}^T A\,\mathbf{z}}{\begin{pmatrix}\mathbf{z}^T & \mathbf{w}_C^T\end{pmatrix} \widehat{B}_{\mathrm{TG}} \begin{pmatrix}\mathbf{z}\\ \mathbf{w}_C\end{pmatrix}} \cdot \min_{\widehat{\mathbf{z}}\in\Re^{n+n_c}\setminus\{0\}} \frac{\widehat{\mathbf{z}}^T \widehat{B}_{\mathrm{TG}}\,\widehat{\mathbf{z}}}{\widehat{\mathbf{z}}^T \widehat{B}_{\mathrm{PTG}}\,\widehat{\mathbf{z}}} \\
&= \lambda_{\min}(B_{\mathrm{TG}}^{-1}A) \cdot \min_{\widehat{\mathbf{w}}\in\Re^{n+n_c}\setminus\{0\}} \frac{\widehat{\mathbf{w}}^T \begin{pmatrix} Q^{-1} & 0\\ 0 & A_C\end{pmatrix}\widehat{\mathbf{w}}}{\widehat{\mathbf{w}}^T \begin{pmatrix} Q^{-1} & 0\\ 0 & K_C^{-1}\end{pmatrix}\widehat{\mathbf{w}}} \\
&= \lambda_{\min}(B_{\mathrm{TG}}^{-1}A) \cdot \min(\lambda_{\min}(K_C\,A_C),\,1). \qquad \square
\end{aligned}$$

3. Multigrid cycles. Multigrid methods are recursively defined. In the SPD case considered here, the iteration matrix $T_{\mathrm{MG}}^{(\ell)}$ at level ℓ depends on the iteration matrix $T_{\mathrm{MG}}^{(\ell-1)}$ at level $\ell-1$ (the next coarser level) according to

$$T_{\mathrm{MG}}^{(\ell)} \;=\; \left(I - R_\ell^T\,A_\ell\right)^{\nu_\ell} \left(I - p_\ell\left(I - \left(T_{\mathrm{MG}}^{(\ell-1)}\right)^{\gamma}\right) A_{\ell-1}^{-1}\,p_\ell^T\,A_\ell\right) \left(I - R_\ell\,A_\ell\right)^{\nu_\ell} \tag{3.1}$$

(see, e.g., [19, pp. 48–49]). In this equation, γ is the *cycle index*; $\gamma=1$ corresponds to the V-cycle and $\gamma=2$ to the W-cycle; larger values of γ are seldom considered in practice.

Now $T_{\rm MG}^{(\ell)}$ is a perturbed two-grid iteration matrix (1.3) with $A = A_\ell$, $R = R_\ell$, $\nu = \nu_\ell$, $p = p_\ell$ and K_C given by

$$K_C \;=\; \left(I - \big(T_{\rm MG}^{(\ell-1)}\big)^\gamma\right) A_{\ell-1}^{-1}.$$

Defining the iteration matrix $T_{\rm TG}^{(\ell)}$ of the (unperturbed) two-grid method by (1.2) with $A_C = A_{\ell-1}$ and letting $B_{\rm TG}^{(\ell)}$, $B_{\rm MG}^{(\ell)}$ be such that

$$I - {B_{\rm TG}^{(\ell)}}^{-1} A_\ell \;=\; T_{\rm TG}^{(\ell)}, \qquad I - {B_{\rm MG}^{(\ell)}}^{-1} A_\ell \;=\; T_{\rm MG}^{(\ell)},$$

inequalities (1.5), (1.6) imply

$$\begin{aligned}
\lambda_{\max}\left({B_{\rm MG}^{(\ell)}}^{-1} A_\ell\right) &\le \lambda_{\max}\left({B_{\rm TG}^{(\ell)}}^{-1} A_\ell\right) \;\cdot\; \max\left(\lambda_{\max}\big(I - \big(T_{\rm MG}^{(\ell-1)}\big)^\gamma\big),\, 1\right),\\
\lambda_{\min}\left({B_{\rm MG}^{(\ell)}}^{-1} A_\ell\right) &\ge \lambda_{\min}\left({B_{\rm TG}^{(\ell)}}^{-1} A_\ell\right) \;\cdot\; \min\left(\lambda_{\min}\big(I - \big(T_{\rm MG}^{(\ell-1)}\big)^\gamma\big),\, 1\right).
\end{aligned}$$

Let

$$\sigma_{\rm MG}^{(\ell)} \;=\; \rho\left(T_{\rm MG}^{(\ell)}\right) \;=\; \max\left(\lambda_{\max}\big({B_{\rm MG}^{(\ell)}}^{-1} A_\ell\big) - 1,\; 1 - \lambda_{\min}\big({B_{\rm MG}^{(\ell)}}^{-1} A_\ell\big)\right)$$

be the convergence factor of the multigrid method at level ℓ. One has, assuming $\sigma_{\rm MG}^{(\ell-1)} \le 1$,

$$\lambda_{\max}\left({B_{\rm MG}^{(\ell)}}^{-1} A_\ell\right) - 1 \;\le\; \begin{cases} \lambda_{\max}\left({B_{\rm TG}^{(\ell)}}^{-1} A_\ell\right)\left(1 + \big(\sigma_{\rm MG}^{(\ell-1)}\big)^\gamma\right) - 1 & \text{if } \gamma \text{ is odd and } T_{\rm MG}^{(\ell-1)} \text{ has some negative eig.,} \\ \lambda_{\max}\left({B_{\rm TG}^{(\ell)}}^{-1} A_\ell\right) - 1 & \text{otherwise,} \end{cases}$$

whereas

$$1 - \lambda_{\min}\left({B_{\rm MG}^{(\ell)}}^{-1} A_\ell\right) \;\le\; 1 - \lambda_{\min}\left({B_{\rm TG}^{(\ell)}}^{-1} A_\ell\right)\left(1 - \big(\sigma_{\rm MG}^{(\ell-1)}\big)^\gamma\right).$$

Then let

$$\sigma_{\rm TG}^{(\ell)} \;=\; \rho\left(T_{\rm TG}^{(\ell)}\right) \;=\; \max\left(\lambda_{\max}\big({B_{\rm TG}^{(\ell)}}^{-1} A_\ell\big) - 1,\; 1 - \lambda_{\min}\big({B_{\rm TG}^{(\ell)}}^{-1} A_\ell\big)\right)$$

be the convergence factor of the (unperturbed) two-grid method at level ℓ. If $\sigma_{\rm TG}^{(\ell)} \le 1$ and if either $\lambda_{\max}\big({B_{\rm TG}^{(\ell)}}^{-1} A_\ell\big) \le 1$ for all ℓ (as occurs when using Galerkin coarse grid matrices) or γ is even (or both), there holds

$$\sigma_{\rm MG}^{(\ell)} \;\le\; 1 - \left(1 - \sigma_{\rm TG}^{(\ell)}\right)\left(1 - \big(\sigma_{\rm MG}^{(\ell-1)}\big)^\gamma\right) \;\le\; 1. \tag{3.2}$$

If the matrix A_0 on the coarsest level is inverted exactly, one has $\sigma_{\rm MG}^{(1)} = \sigma_{\rm TG}^{(1)}$, and (3.2) defines a recursion which may be followed from $\ell = 2, 3, \ldots$ until the finest level.

For $\gamma = 1$ (the V-cycle), this does not yield bounds independent of the number of levels, although the estimates may be practically relevant for few levels if the $\sigma_{\rm TG}^{(\ell)}$ are small. For instance, $\sigma_{\rm TG}^{(\ell)} \le 0.1$ yields $\sigma_{\rm MG}^{(\ell)} \le 0.41$ for $\ell = 5$, which is not that bad.

The most interesting application is $\gamma = 2$ (the W-cycle). No additional assumption on $\lambda_{\max}({B_{\rm TG}^{(\ell)}}^{-1} A)$ is needed (γ is even), and one may check that if

$$\sigma_{\rm TG}^{(\ell)} \;\leq\; \sigma$$

and

$$\sigma_{\rm MG}^{(\ell-1)} \;\leq\; \frac{\sigma}{1-\sigma}$$

hold for some $\sigma < 1/2$, then

$$\sigma_{\rm MG}^{(\ell)} \;\leq\; \frac{\sigma}{1-\sigma}.$$

This proves the following theorem.

THEOREM 3.1. *Consider a multigrid method recursively defined by the iteration matrix* (3.1) *with* $\gamma = 2$ *for* $\ell = 1, 2, \ldots$ *and* $T_{MG}^{(0)} = 0$ *(exact inversion on the coarsest level). Assume that* A_ℓ, $\ell = 0, 1, \ldots,$ *is SPD and that* R_ℓ, $\ell = 1, 2, \ldots,$ *is such that* $\rho(I - R_\ell A_\ell) < 1$. *If the spectral radius of the two-grid iteration matrix* (1.2) *with* $A = A_\ell$, $R = R_\ell$, $\nu = \nu_\ell$, $p = p_\ell$, *and* $A_C = A_{\ell-l}$ *is bounded by some* $\sigma < 1/2$ *independently of* ℓ, *then the spectral radius of the multigrid iteration matrix* (3.1) *is bounded by* $\sigma/(1-\sigma)$, *independently of* ℓ.

Comparison with the standard algebraic analysis. This analysis, see, e.g., [11, 19], is based on matrix norms, instead of spectral radii, and applies to the nonsymmetric case as well. In the framework considered here, this analysis proves that if

$$\|T_{\rm TG}^{(\ell)}\| \;\leq\; \sigma^*$$

and

$$\left\|\left(I - R_\ell^T\, A_\ell\right)^{\nu_\ell}\, p_\ell\right\| \;\cdot\; \left\|A_{\ell-1}^{-1}\, p_\ell^T\, A_\ell\, \left(I - R_\ell\, A_\ell\right)^{\nu_\ell}\right\| \;\leq\; C$$

hold for some σ^*, C such that

$$4\, C\, \sigma^* \;\leq\; 1, \tag{3.3}$$

then the iteration matrix for the W-cycle ($\gamma = 2$) satisfies

$$\|T_{\rm TG}^{(\ell)}\| \;\leq\; \frac{1 - \sqrt{1 - 4\, C\, \sigma^*}}{2\, C} \;\leq\; 2\, \sigma^*. \tag{3.4}$$

Note that this result holds for any matrix norm $\|\cdot\|$.

To comment on it, first observe that

$$\begin{aligned}
&\left\|\left(I - R_\ell^T\, A_\ell\right)^{\nu_\ell}\, p_\ell\right\| \;\cdot\; \left\|A_{\ell-1}^{-1}\, p_\ell^T\, A_\ell\, \left(I - R_\ell\, A_\ell\right)^{\nu_\ell}\right\| \\
&\geq \left\|\left(I - R_\ell^T\, A_\ell\right)^{\nu_\ell}\, p_\ell\, A_{\ell-1}^{-1}\, p_\ell^T\, A_\ell\, \left(I - R_\ell\, A_\ell\right)^{\nu_\ell}\right\| &(3.5)\\
&= \left\|\left(I - R_\ell^T\, A_\ell\right)^{\nu_\ell}\, \left(I - R_\ell\, A_\ell\right)^{\nu_\ell} - T_{\rm TG}^{(\ell)}\right\| \\
&\geq \left\|\left(I - R_\ell^T\, A_\ell\right)^{\nu_\ell}\, \left(I - R_\ell\, A_\ell\right)^{\nu_\ell}\right\| - \left\|T_{\rm TG}^{(\ell)}\right\| \\
&\geq \left(\rho\, (I - R_\ell\, A_\ell)\right)^{2\,\nu_\ell} - \sigma^*. &(3.6)
\end{aligned}$$

In practical situations, $\rho\,(I - R_\ell\, A_\ell) \approx 1$ (otherwise, the coarse grid correction would be unnecessary for fast convergence). Hence, $C \gtrsim 1-\sigma^*$. Moreover, if $A_{\ell-1} = p_\ell^T A_\ell\, p_\ell$ (i.e., with a Galerkin coarse grid matrix), the middle term in the right-hand side of (3.5) is a projector that leaves the vectors in the range of p_ℓ unchanged, leading to expect $C \gtrsim 1$ in such cases. Since $\sigma^* \geq \sigma$ (the spectral radius is a lower bound on the matrix norm for any norm), the condition (3.3) is thus generally more restrictive than our condition $\sigma < 1/2$, even if one uses the energy norm for which $\sigma^* = \sigma$. For instance, $C \gtrsim 1$ then means that (3.3) requires $\sigma \leq \sigma^* \lesssim 1/4$.

When both bounds apply, if, in addition,

$$C \;\geq\; 1 - \sigma^*$$

(as one expects according to (3.6) and the discussion above), then our bound $\sigma/(1-\sigma)$ is always better than the bound (3.4). Indeed, taking $\sigma^* = \sigma$ (which is the most favorable for (3.4)), one has (since $\sigma(1+2\,C) = \sigma + \frac{1}{2}4\,C\,\sigma < 1$)

$$\begin{aligned} \frac{\sigma}{1-\sigma} \;\leq\; \frac{1-\sqrt{1-4\,C\,\sigma}}{2\,C} &\iff \sqrt{1-4\,C\,\sigma^*} \;\leq\; \frac{1-\sigma(1+2\,C)}{1-\sigma} \\ &\iff (1-\sigma)^2(1-4\,C\,\sigma) - (1-\sigma(1+2\,C))^2 \;\leq\; 0 \\ &\iff -4\,C\,\sigma^3 + (4\,C - 4\,C^2)\sigma^2 \;\leq\; 0 \\ &\iff -\sigma + 1 - C \;\leq\; 0. \end{aligned}$$

By way of illustration, consider the case $\sigma = \sigma^* = 1/4$ and $C = 1$. Then our bound for the W-cycle is $1/3$, whereas (3.4) gives $1/2$. Note, however, that, for σ going to 0, both bounds converge to σ.

Comparison with Braess's analysis. Braess's analysis [6, pp. 226–228] assumes Galerkin coarse grid matrices and is based on two-grid and multigrid schemes without postsmoothing. To make things clear, let

$$\begin{aligned} \widetilde{T}_{\rm TG}^{(\ell)} &= \left(I - p_\ell\, A_{\ell-1}^{-1}\, p_\ell^T\, A_\ell\right)\,(I - R_\ell\, A_\ell)^{\nu_\ell}\,, \\ \widetilde{T}_{\rm MG}^{(\ell)} &= \Big(I - p_\ell\left(I - \big(\widetilde{T}_{\rm MG}^{(\ell-1)}\big)^\gamma\right) A_{\ell-1}^{-1}\, p_\ell^T\, A_\ell\Big)\,(I - R_\ell\, A_\ell)^{\nu_\ell} \end{aligned}$$

be the corresponding iteration matrices, and denote by $\widetilde{\rho}_{\rm TG}^{(\ell)}$, $\widetilde{\rho}_{\rm MG}^{(\ell)}$ their energy norm:

$$\widetilde{\rho}_{\rm TG}^{(\ell)} \;=\; \|\widetilde{T}_{\rm TG}^{(\ell)}\|_{A_\ell}, \qquad \widetilde{\rho}_{\rm MG}^{(\ell)} \;=\; \|\widetilde{T}_{\rm MG}^{(\ell)}\|_{A_\ell}.$$

The main convergence result for multigrid cycles in [6] is inequality (3.9) from Chapter V. With the above notation, this inequality amounts to

$$\widetilde{\rho}_{\rm MG}^{(\ell)}{}^{\,2} \;\leq\; 1 \;-\; \left(1 - \left(\widetilde{\rho}_{\rm TG}^{(\ell)}\right)^2\right)\left(1 - \left(\widetilde{\rho}_{\rm MG}^{(\ell-1)}\right)^{2\gamma}\right). \tag{3.7}$$

Comparing with (3.2), the requirement on $\widetilde{\rho}_{\rm TG}^{(\ell)}$ to have an optimal method is less restrictive than the requirement we have on $\sigma_{\rm TG}^{(\ell)}$. However, as seen in [15, eq. (41)], when $A_{\ell-1} = p_\ell^T A_\ell\, p_\ell$ (as needed to prove (3.7)), there holds

$$A_\ell\, T_{\rm TG}^{(\ell)} \;=\; \left(\widetilde{T}_{\rm TG}^{(\ell)}\right)^T\, A_\ell\, \widetilde{T}_{\rm TG}^{(\ell)}$$

entailing

$$\sigma_{\mathrm{TG}}^{(\ell)} = \left(\widetilde{\rho}_{\mathrm{TG}}^{(\ell)}\right)^2 . \tag{3.8}$$

That is, if one uses the same smoother and prolongation, the two-grid method with symmetrized pre- and postsmoothing converges twice as fast as the method without postsmoothing.

Acknowledging this fact, one sees that (3.7) defines recursively a bound on $\widetilde{\rho}_{\mathrm{MG}}^{(\ell)}$ which is equal to the square root of the bound on $\sigma_{\mathrm{MG}}^{(\ell)}$ obtained from our result (3.2). This suggests that, as shown by (3.8) in the two-grid case, the W-cycle multigrid with symmetrized smoothing can converge twice as fast as the corresponding algorithm without postsmoothing.

Sometimes the bound for the scheme without postsmoothing is used as worst case estimate for the general case. From that point of view, our analysis gives sharper bounds. For instance, Theorem 3.4 in [6, Chapter V] states that the convergence factor for the W-cycle is not larger that 3/5 when $\widetilde{\rho}_{\mathrm{TG}}^{(\ell)} \leq 1/2$ for all ℓ, whereas, with $\sigma_{\mathrm{TG}}^{(\ell)} \leq 1/4$, our theorem, Theorem 3.1, then proves that the convergence factor for the W-cycle does not exceed 1/3.

Acknowledgments. I thank M. Hochstenbach for a careful reading of the manuscript. An anonymous referee suggested numerous improvements which are deeply appreciated. Another referee drew our attention to [6].

REFERENCES

[1] O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, Cambridge, UK, 1994.
[2] O. Axelsson and P. S. Vassilevski, *Algebraic multilevel preconditioning methods*, I, Numer. Math., 56 (1989), pp. 157–177.
[3] O. Axelsson and P. S. Vassilevski, *Algebraic multilevel preconditioning methods*, II, SIAM J. Numer. Anal., 27 (1990), pp. 1569–1590.
[4] R. E. Bank, T. F. Dupont, and H. Yserentant, *The hierarchical basis multigrid method*, Numer. Math., 52 (1988), pp. 427–458.
[5] D. Braess, *The convergence rate of a multigrid method with Gauss-Seidel relaxation for the Poisson equation*, in Multigrid Methods, W. Hackbusch and U. Trottenberg, eds., Lecture Notes in Math. 960, Springer-Verlag, Berlin, Heidelberg, New York, 1982, pp. 368–386.
[6] D. Braess, *Finite Elements*, Cambridge University Press, Cambridge, UK, 1997.
[7] D. Braess and W. Hackbusch, *A new convergence proof for the multigrid method including the V-cycle*, SIAM J. Numer. Anal., 20 (1983), pp. 967–975.
[8] A. Brandt, *Algebraic multigrid theory: The symmetric case*, Appl. Math. Comput., 19 (1986), pp. 23–56.
[9] R. D. Falgout and P. S. Vassilevski, *On generalizing the algebraic multigrid framework*, SIAM J. Numer. Anal., 42 (2004), pp. 1669–1693.
[10] R. D. Falgout, P. S. Vassilevski, and L. T. Zikatanov, *On two-grid convergence estimates*, Numer. Linear Algebra Appl., 12 (2005), pp. 471–494.
[11] W. Hackbusch, *Multi-grid convergence theory*, in Multigrid Methods, W. Hackbusch and U. Trottenberg, eds., Lecture Notes in Math. 960, Springer-Verlag, Berlin, Heidelberg, New York, 1982, pp. 177–219.
[12] W. Hackbusch, *Multi-grid Methods and Applications*, Springer-Verlag, Berlin, 1985.
[13] J. Mandel, S. McCormick, and J. Ruge, *An algebraic theory for multigrid methods for variational problems*, SIAM J. Numer. Anal., 25 (1988), pp. 91–110.
[14] S. F. McCormick, *Multigrid methods for variational problems: General theory for the V-cycle*, SIAM J. Numer. Anal., 22 (1985), pp. 634–643.
[15] Y. Notay, *Algebraic multigrid and algebraic multilevel methods: A theoretical comparison*, Numer. Linear Algebra Appl., 12 (2005), pp. 419–451.
[16] P. Oswald, *Multilevel Finite Element Approximation: Theory and Applications*, Teubner Skr. Numer., Teubner, Stuttgart, 1994.

[17] P. Oswald, *Subspace correction methods and multigrid theory*, in Multigrid, Academic Press, London, 2001, pp. 533–572.
[18] K. Stüben, *An introduction to algebraic multigrid*, in Multigrid, Academic Press, London, 2001, pp. 413–532.
[19] U. Trottenberg, C. W. Oosterlee, and A. Schüller, *Multigrid*, Academic Press, London, 2001.
[20] P. S. Vassilevski, *A block–factorization (algebraic) formulation of multigrid and Schwarz methods*, East-West J. Numer. Math., 6 (1998), pp. 65–79.
[21] P. S. Vassilevski, *Multilevel Block Factorization Preconditioners*, Springer-Verlag, New York, to appear.
[22] J. Xu, *Iterative methods by space decomposition and subspace correction*, SIAM Rev., 34 (1992), pp. 581–613.
[23] H. Yserentant, *Old and new convergence proofs for multigrid methods*, Acta Numer., 2 (1993), pp. 285–326.