Numer. Math. Theor. Meth. Appl.
doi: 10.4208/nmtma.2015.w04si

Vol. **8**, No. 2, pp. 168-198
May 2015

Algebraic Theory of Two-Grid Methods

Yvan Notay*

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

Received 02 November 2013; Accepted 28 October 2014

Abstract. About thirty years ago, Achi Brandt wrote a seminal paper providing a convergence theory for algebraic multigrid methods [Appl. Math. Comput., 19 (1986), pp. 23–56]. Since then, this theory has been improved and extended in a number of ways, and these results have been used in many works to analyze algebraic multigrid methods and guide their developments. This paper makes a concise exposition of the state of the art. Results for symmetric and nonsymmetric matrices are presented in a unified way, highlighting the influence of the smoothing scheme on the convergence estimates. Attention is also paid to sharp eigenvalue bounds for the case where one uses a single smoothing step, allowing straightforward application to deflation-based preconditioners and two-level domain decomposition methods. Some new results are introduced whenever needed to complete the picture, and the material is self-contained thanks to a collection of new proofs, often shorter than the original ones.

AMS subject classifications: 65F08, 65F10, 65F50

Key words: Multigrid, convergence analysis, algebraic multigrid, preconditioning, two-level method.

1. Introduction

Multigrid methods are among the most efficient iterative techniques to solve large sparse systems of linear equations. These methods combine two different iterations: a smoothing iteration, which is often a simple iterative method like the Gauss–Seidel method, and a coarse grid correction, which consists in computing an approximate solution to the residual equation on a coarser grid with fewer unknowns.

Perhaps because of this combination two different processes, multigrid methods are difficult to analyze. Abstract theories (e.g., [21,46]) are restricted to discretized partial differential equations on a regularly refined grid, and allow one to obtain only qualitative results. Fourier and local mode analyses (e.g., [55]) yield sharp quantitative estimates, but only when the system to solve stems from a constant or smoothly variable

*Corresponding author. *Email addresses:* ynotay@ulb.ac.be (Yvan Notay) Yvan Notay is Research Director of the Fonds de la Recherche Scientifique - FNRS.

grid stencil. In all other cases, the algebraic theory [6,10,17,18,42,48,51,57] appears as the right tool, despite its own limitations. In particular, it allows the assessment of "algebraic" multigrid (AMG) methods [8], in which the coarse grid correction is not related to the discretization on a coarser grid, but defined by applying proper algorithms to the system matrix. Examples of works where the theoretical results have helped to the design and/or the analysis of AMG methods include [2, 10, 14, 38, 39, 43, 48]; the algebraic theory also gave theoretical foundations to the coarsening by compatible relaxation [7,24], see [9,17,42].

Early analyses of multigrid methods using essentially algebraic arguments trace back to the eighties and include [3, 20, 26, 31–33]. They were quickly followed by Brandt's seminal paper [6], which provides the first convergence theory applicable to (and intended for) AMG methods. Since then, the theory has been improved and extended in a number of works; see, e.g., [10,17,18,42,48,51,57]. Now, each reference highlights its own improvements, and the reader searching for a clear summary or overview of the state of the art has to go through quite many specialized works. Moreover, many of these works are restricted to a specific smoothing scheme, making difficult the setting up of a clear picture of the available results.

Our main goal in this paper is to present such a clear picture, highlighting in particular the differences induced by the type of smoothing scheme.

The organization of the paper is a bit unusual. After the introduction of the general setting and the statement of the common assumptions (§2), we state in §3 the set of results as "Facts", which are given without proof nor comment. All comments are gathered in §4, whereas proofs are delayed until §5. Most of the facts are indeed not new, and hence have already been proved and commented in the original references. Regarding comments, we therefore focus on those which compare the facts that are similar but cover different situations, a viewpoint seldom taken in the literature. Logically, this can be done only after all facts have been stated.

Regarding proofs, some readers may feel that they are unnecessary. However, for most facts we are able to give a new proof, generally shorter than those in the original references, and we also often condense the proof of several facts in a single one. This allows us to be self-contained while keeping §5 to a fairly reasonable size. Moreover, some of the results are new and require in any case a proof. For instance, the necessary and sufficient conditions for the nonsingularity of the two-grid preconditioner are seemingly stated for the first time (Fact 1.1), and we are not aware of previous studies considering in detail the effect of the number of smoothing steps on the theoretical estimates (Fact 5.4).

The chosen format (inspired by [22]) does not allow us to include a thorough discussion of past contributions and how they influenced the progress in the field. For each fact stated in §3 we mention its origin in parenthesis, and we further make some connection with the literature in §4, especially in Remark 4 devoted to historical comments. However, this yet gives only a very partial account of previous developments. In particular, one should not forget that more general or sharper bounds are always indebted to their predecessors: even when the new proof is very different in nature,

it has in general been guided by the intuition of the result that one aims at. Besides the works cited elsewhere in this paper, it is thus also worth mentioning seminal papers like [5, 27, 62] that occupy a middle ground between algebraic and non algebraic convergence theories.

We could of course develop further the historical viewpoint initiated in Remark 4, but then we would overlap with the recent survey [25], which we definitely recommend to readers also interested by a more classical type of review. In fact, the present contribution and [25] are very complementary. Besides the differences in the presentation, [25] is focused on the symmetric and positive definite (SPD) case, whereas we aim at a unified presentation of the results available for symmetric and non symmetric matrices. We also pay attention to the sometimes subtle differences induced by the type of smoothing scheme, and we consider sharp eigenvalue estimates for the case where one uses only a single smoothing step, allowing straightforward application to deflation-based preconditioners and two-level domain decomposition methods (see the end of the next section for more comments on this topic). We also relate some non sharp but more tractable bounds to sharp estimates for a simplified two-grid scheme (see Remark 5 in §4). On the other hand, [25] contains an interesting further discussion of other non sharp but more easily computable bounds. We also refer to [25] for a review of theoretically motivated AMG algorithms, and more generally, for comments on the interplay between the theoretical developments and the design of AMG methods.

Now, before entering the core of this paper, let us make clear what is covered and what is not. We try to give a comprehensive overview of "algebraic" convergence analysis results for two-grid schemes with Galerkin coarse grid matrices and assuming that the restriction is the transpose of the prolongation. We also assume that the smoothing iteration is based on a regular preconditioner; i.e., all unknowns are affected by the relaxation. Hence the following situations are not considered.

Schemes where the restriction is not the transpose of the prolongation. In practice, many geometric and algebraic multigrid methods set the restriction equal to the transpose of the prolongation. For symmetric matrices, this is clearly the most sensible choice, as this allows (by choosing a symmetric smoothing scheme) to preserve the symmetry of the preconditioning operator. In the positive definite case, this further allows a variational formulation that provides a useful minimization principle. On the other hand, setting the restriction equal to the transpose of the prolongation is also widespread when the system matrix is not symmetric, although, from the theoretical viewpoint, there is in that case no analysis giving a clear advantage to this choice. A number of results even go in the opposite direction [13, 15, 49, 61]. It would therefore be desirable to include them in this survey. However, their form does not really allow a comparative discussion with the other facts mentioned in this work, hence we refer the interested readers to the original references.

Extension to multi-grid (analysis of multigrid cycles) and inexact coarse grid solves. We refer the reader to the literature for the available results; see the monographs [21, Chapter 7], [57, Chapter 5], or, e.g., the papers [33, 37, 41, 45, 62, 63] for analyses developing an algebraic viewpoint and forming therefore an ideal complement to the

results presented here. Note that the two-grid schemes as considered in the present work are also particular instances —actually the simplest ones— of the more involved schemes analyzed in these works. Hence the two-grid analysis is in any event a preliminary step, in the sense that there is little to hope from the study of the general case if, as often arises, the two-grid case is already too complicated to allow a fully rigorous approach.

Non Galerkin coarse grid matrices. This is a serious limitation regarding the application to geometric multigrid methods, which often use coarse grid matrices obtained from a rediscretization on the coarse mesh. To analyze these methods with the results given here, one has to consider such coarse grid matrices as perturbation of their Galerkin counterpart and apply, e.g., the results in [41], which is neither practical nor (likely) very accurate. However, this is not restrictive regarding algebraic multigrid methods, which most often use Galerkin coarse grid matrices.

Smoothing iterations based on a singular preconditioner. This includes schemes based on F–relaxation [51], in which coarse grid unknowns are not affected by the smoothing iteration and thus only updated via the coarse grid correction. A practical example is the hierarchical basis multigrid method [4]. Such schemes have a simpler algebraic structure that requires another type of analysis, for which we refer the reader to [4, 40]. Regarding AMG methods, excluding F–relaxation does not represent a strong limitation in the sense that this approach is rarely advocated in practice (The main motivation for the discussion of F–relaxation in [51] seems indeed to lie in the easier access to theoretical insight).

1.1. Notation

For any (possibly complex) vector $\mathbf{v}$, $\mathbf{v}^T$ is its transpose, $\mathbf{v}^*$ is its conjugate transpose, $\|\mathbf{v}\|$ is its 2-norm and, for arbitrary SPD matrix G, $\|\mathbf{v}\|_G$ is the associated energy norm: $\|\mathbf{v}\|_G = \sqrt{\mathbf{v}^* G\,\mathbf{v}}$. For any square matrix C, $\rho(C)$ is its spectral radius (i.e., its largest eigenvalue in modulus), $\sigma(C)$ is its spectrum, $\|C\| = \sqrt{\rho(C^T C)}$ is its usual 2-norm and, for arbitrary SPD matrix G, $\|C\|_G$ is the matrix norm induced by the energy norm associated to G: $\|C\|_G = \max_{\mathbf{v}\neq\mathbf{0}} \|C\,\mathbf{v}\|_G/\|\mathbf{v}\|_G = \|G^{1/2} C\, G^{-1/2}\|$. Eventually, if C has real eigenvalues (e.g., if it is similar to a symmetric matrix), $\lambda_{\min}(C)$ and $\lambda_{\max}(C)$ stand for its smallest and largest eigenvalue, respectively.

2. General setting

We assume that the system matrix A is a real $n \times n$ nonsingular matrix, and we consider the two-grid schemes that are described by the iteration matrix

$$T \;=\; \left(I - M_2^{-1}A\right)^{\nu_2} \left(I - P\,A_c^{-1}P^T A\right) \left(I - M_1^{-1}A\right)^{\nu_1} \;, \tag{2.1}$$

where M_1 (resp. M_2) is the preconditioner which defines the pre-smoother (resp. post-smoother) and where ν_1 (resp. ν_2) is the corresponding number of smoothing steps [†]; P is the prolongation matrix, of size $n \times n_c$ with $n_c < n$, and where A_c is the coarse grid matrix; we further restrict ourselves to Galerkin coarse grid matrices; i.e., we assume

$$A_c \;=\; P^T A\, P\;. \tag{2.2}$$

Of course, some basic assumptions are needed to make the above definition of T consistent: M_1, M_2 need to be nonsingular $n \times n$ matrices; A_c has to be nonsingular as well, which, given (2.2), also implies that P has full rank; i.e., has rank n_c. Note that when A is positive definite in $\mathbb{R}^n$ (e.g., SPD), it suffices to assume that P has full rank to guarantee the nonsingularity of A_c; this is, however, not true for a general (possibly indefinite) matrix A.

The following two matrices play a key role in the results to be presented. Firstly, the matrix Y which corresponds to the preconditioner that brings in one step the same effect as pre-smoothing followed by post-smoothing; that is, the matrix Y such that

$$I - Y^{-1}A \;=\; \left(I - M_2^{-1}A\right)^{\nu_2} \left(I - M_1^{-1}A\right)^{\nu_1}\;. \tag{2.3}$$

Next, the matrix X which corresponds to post-smoothing followed by pre-smoothing:

$$I - X^{-1}A \;=\; \left(I - M_1^{-1}A\right)^{\nu_1} \left(I - M_2^{-1}A\right)^{\nu_2}\;. \tag{2.4}$$

Stating this, we implicitly assume that the so defined X^{-1} and Y^{-1} are nonsingular; i.e., given that A is assumed non singular, we need that the products in the right hand sides of (2.3) and (2.4) define matrices that have no eigenvalue equal to 1. In general, this is ensured by assuming that the global action of the smoothing iterations defines a convergent iterative process; that is, by assuming

$$\rho\Big(\, \left(I - M_2^{-1}A\right)^{\nu_2} \left(I - M_1^{-1}A\right)^{\nu_1} \,\Big) \;<\; 1\;, \tag{2.5}$$

or, equivalently [‡]

$$\rho\Big(\, \left(I - M_1^{-1}A\right)^{\nu_1} \left(I - M_2^{-1}A\right)^{\nu_2} \,\Big) \;<\; 1\;.$$

Note, however, that when either $\nu_1 = 1$, $\nu_2 = 0$ or $\nu_1 = 0$, $\nu_2 = 1$, the nonsingularity of X and Y straightforwardly follows from that of M_1 or M_2; i.e., no additional assumption is needed in these cases. On the other hand, it is worth noting that, when $M_1 = M_2 = M$, (2.5) amounts to

$$\rho\left(I - M^{-1}A\right) \;<\; 1$$

(whatever the number of smoothing steps).

[†] ν_1, ν_2 are nonnegative integers, $\nu_1 = 0$ (resp. $\nu_2 = 0$) corresponding to no pre-smoothing (resp. no post-smoothing)

[‡] since permuting the factors of the product of two square matrices does not change its spectrum even when any of the factors may be singular, see [23, Theorem 1.3.22] or [42, Lemma A.1]

It is also interesting to look at the specific case where $\nu_1 = \nu_2 = 1$. Then,

$$\begin{aligned} Y &= M_1 \, (M_1 + M_2 - A)^{-1} \, M_2 \, , \\ X &= M_2 \, (M_1 + M_2 - A)^{-1} \, M_1 \, . \end{aligned}$$

If, more particularly, A is SPD and $M_1 = M_2^T = M$, it can further be shown that (2.5) holds if and only if $M + M^T - A$ is SPD; that is, if and only if the above defined Y and X are SPD [40, Lemma A.1].

Often convergence analysis results are expressed as bounds on the spectral radius of the iteration matrix T. However, more details are available if one expresses them as relations bounding the eigenvalues of the preconditioned matrix $\mathcal{B}A$, where the preconditioning matrix $\mathcal{B}$ is defined from

$$I - \mathcal{B}A \; = \; T \, . \tag{2.6}$$

We follow this approach, which is also more relevant when the two-grid method is used as a preconditioner for the conjugate gradient or another Krylov subspace method. Note that, with (2.3), (2.6) amounts to

$$\mathcal{B} \; = \; Y^{-1} + \left(I - M_2^{-1} A\right)^{\nu_2} \; P \, A_c^{-1} P^T \; \left(I - A \, M_1^{-1}\right)^{\nu_1} \; . \tag{2.7}$$

Observe that this explicit definition does no require the inverse of A.

To sum up, we thus use the following assumptions.

General Assumptions
A is a real nonsingular $n \times n$ matrix;
P is a real $n \times n_c$ matrix of rank $n_c < n$;
$A_c \; = \; P^T A \, P$ is non singular;
M_1, M_2 are real nonsingular $n \times n$ matrices and ν_1, ν_2 are nonnegative integers;
T is the $n \times n$ matrix defined by (2.1)*;*
$\mathcal{B}$ is the $n \times n$ matrix defined by (2.6)*, or, equivalently, by* (2.7)*;*
Either $\nu_1 + \nu_2 = 1$ or (2.5) *holds;*
Y and X are the $n \times n$ matrices defined by (2.3) *and* (2.4)*, respectively.*

In the next section, after the general case in §3.1 (i.e., no other assumption than those stated above), we consider successively the following four particular cases:

- A positive definite in $\mathbb{R}^n$ and either only a single pre-smoothing step with SPD M_1 or only a single post-smoothing step with SPD M_2 (§3.2, covering both non-symmetric and symmetric (i.e., SPD) A);
- A SPD and a general (possibly nonsymmetric) smoothing scheme (§3.3);
- A SPD and a symmetric smoothing scheme (i.e., $M_1 = M_2^T$ and $\nu_1 = \nu_2$, §3.4);
- A SPD and the same SPD $M_1 = M_2$ for both pre- and post-smoothing (§3.5).

Observe that a common feature of these four cases is the assumption that A is positive definite in $\mathbb{R}^n$; that is,

$$\mathbf{v}^T A\,\mathbf{v} \;>\; 0 \qquad \forall \mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}\}\ .$$

Note that this holds if and only if

$$A_S \;=\; \tfrac{1}{2}\left(A + A^T\right) \tag{2.8}$$

is SPD. As will be seen, a key role in the convergence analysis is then played by the quantity

$$\begin{aligned} K_G \;&=\; \max_{\mathbf{v}\in\mathbb{R}^n\backslash\{\mathbf{0}\}} \frac{\mathbf{v}^T G\big(I - P\left(P^T G\,P\right)^{-1} P^T G\big)\,\mathbf{v}}{\mathbf{v}^T A\,\mathbf{v}} \\ &= \left(\max_{\mathbf{v}\in\mathbb{R}^n\backslash\{\mathbf{0}\}} \frac{\|\big(I - P\left(P^T G\,P\right)^{-1} P^T G\big)\,\mathbf{v}\|_G}{\|\mathbf{v}\|_{A_S}} \right)^2 , \end{aligned} \tag{2.9}$$

where G is an SPD matrix related to the smoother and depending on the case at hand. Given their importance, we devote an entire section (§3.6) to results related to K_G ; when reading a fact where K_G is mentioned, it is thus a good idea to remember that it can be complemented with the facts in §3.6.

Now, as will be seen, not so much can be said in the general case (§3.1). Hence the above list of cases means that, for nonsymmetric matrices, the analysis is in practice restricted to a simplified scheme that uses only a single step of pre- or post-smoothing. Moreover, the preconditioner M_1 or M_2 has to be symmetric despite the matrix is not, a requirement satisfied with damped Jacobi smoothing but leaving little room for other possibilities. A sensible approach, followed in [43], is then to develop the analysis for this special case, and next check with numerical experiments that the convergence is indeed improved when using more common smoothing schemes.

The situation is more satisfactory regarding the SPD case, since the considered cases embrace virtually all possibilities. The last two are particular instances of the general general smoothing scheme considered in §3.3. They are analyzed separately because more insight can be gained. Similarly, if one uses the same smoother for pre- and post-smoothing, with the same number of pre- and post-smoothing steps, it is more instructive to look at the specific results in §3.5 although those in §3.4 apply as well.

That said, we also emphasize the situation with only a single smoothing step in the SPD case as well (§3.2). The reason is twofold. On the one hand, as discussed in §4 (see in particular Remark 5), in other cases, the sharp convergence estimates are often untractable, and their practical assessment requires some simplifications which in fact amount to analyze a basic scheme with a single smoothing step. This is also true for the estimate in the seminal paper [6], which the results in §3.2 allow to relate to a sharp bound for a simplified two grid scheme; see Remarks 4 and 5 in §4 for details.

On the other hand, two-level schemes with a single smoothing step play an important role in several non-multigrid applications, and it has perhaps so far not been

enough highlighted that they can also be analyzed with the help of the algebraic multigrid theory. This includes two-level domain decomposition methods (e.g., [47,50,54]) and deflation-based preconditioners (e.g., [19,58–60]). Indeed, these approaches often fit with the algebraic framework described above. They still combine a basic iterative method and a coarse grid correction, even though the prolongation that defines this latter is based on other principles and ideas than that commonly used in AMG methods. Going to the details, in these contexts, one rarely uses an implementation directly obtained from the above definition of $\mathcal{B}$, because, with a single smoothing step, such $\mathcal{B}$ is nonsymmetric even in the SPD case, and hence cannot be used as a preconditioner for the conjugate gradient method. However, see [35], one can use a deflation process to enforce that all iterates belong to a subspace in which $\mathcal{B}$ acts effectively as a symmetric operator. Alternatively, balancing Neumann-Neumann domain decomposition methods [28, 29, 50, 54] symmetrize the preconditioner by repeating the coarse grid correction twice. Nevertheless, the results in §3.2 remain applicable in both cases because the eigenvalue distribution of the preconditioned matrix is identical or closely related to that of $\mathcal{B}A$ with $\mathcal{B}$ as in (2.7) and, say, only a single post-smoothing step (i.e., $\nu_1 = 0$, $\nu_2 = 1$ and therefore $Y = M_2$); we refer to [53] for a comparative discussion of implementations and a detailed account on this matter.

Note a limitation of implementations based on a deflation process: exact inversion of A_c is strictly required; i.e., it is not possible to extend the two-grid scheme in a multilevel one and still consistently use the conjugate gradient method. Opposite to this, with a symmetric smoothing scheme, it is well known (and actually clear from (2.7) with Y SPD) that $\mathcal{B}$ remains SPD when A_c is exchanged for any SPD approximation. Of course, deflation-based preconditioners, like any scheme with a single smoothing step, can be extended in a multi-level scheme if one uses a standard implementation either as stand alone solver or as preconditioner for a general purpose Krylov method like GMRES; see [16] for an example.

Regarding these families of methods, observe also that, according to (2.1), we consider in this work only the *multiplicative* combination of a basic iteration and a coarse grid correction. However, *additive* variants are also popular in the contexts of domain decomposition methods and deflation-based preconditioners. We refer to [34, 36, 44] for a detailed comparative analysis of additive and multiplicative variants in the SPD case. We mention here only the main conclusion from [44]: with an appropriate scaling of the smoother, the multiplicative variant leads always to a smaller condition number, but the improvement is at most by a modest factor, hence both approaches can be seen as qualitatively equivalent.

3. Facts

3.1. General case

Fact 1.1. (Characterization of the eigenvalues from Theorem 2.1 of [42]; otherwise new)

Let the general assumptions hold. Then $\mathcal{B}$ is nonsingular if and only if $X_c = P^T X\, P$ is nonsingular.

In this case,

$$\mathcal{B}^{-1} \;=\; Y - Y\,\left(I - M_2^{-1}A\right)^{\nu_2}\, PX_c^{-1}P^T\,\left(I - A\,M_1^{-1}\right)^{\nu_1}\, Y\;, \tag{3.1}$$

and the eigenvalues of $\mathcal{B}A$ are 1 with multiplicity n_c plus the nonzero eigenvalues of

$$X^{-1}\,A\left(I - PA_c^{-1}P^TA\right)\;, \tag{3.2}$$

which are also the inverse of the nonzero eigenvalues of

$$A^{-1}\,X\left(I - PX_c^{-1}P^TX\right)\;. \tag{3.3}$$

3.2. Case: A positive definite in $\mathbb{R}^n$; either a single pre-smoothing step with SPD M_1 or a single post-smoothing step with SPD M_2

Thus:

$$\left\{\begin{array}{l} A \text{ is positive definite in } \mathbb{R}^n\,, \\ \left\{\begin{array}{rl} \text{either} & \nu_1 = 1\,,\, \nu_2 = 0 \text{ and } M_1 = M \text{ is SPD} \\ \text{or} & \nu_1 = 0\,,\, \nu_2 = 1 \text{ and } M_2 = M \text{ is SPD}\,. \end{array}\right. \end{array}\right. \tag{3.4}$$

As already noted, A is positive definite in $\mathbb{R}^n$ if and only if its symmetric part (2.8) is SPD. Then, the symmetric part of the inverse

$$[A^{-1}]_S \;=\; \tfrac{1}{2}\left(A^{-1} + A^{-T}\right) \;=\; A^{-1}\,A_S\,A^{-T} \tag{3.5}$$

is SPD as well; i.e., A^{-1} is also positive definite in $\mathbb{R}^n$. Of course, if A is symmetric, $A_S = A$ and $[A^{-1}]_S = A^{-1}$.

Fact 2.1. (New)
Let the general assumptions and (3.4) *hold. Then $X = M$ and $X_c = P^TM\,P$ are positive definite in $\mathbb{R}^n$ (thus nonsingular). Therefore, the additional condition in Fact 1.1 is satisfied; hence $\mathcal{B}$ is nonsingular and* (3.1) *holds as well as the characterization of the eigenvalues via* (3.2) *and* (3.3)*. Further,* (3.1) *amounts to*

$$\mathcal{B}^{-1} \;=\; \begin{cases} M - (M-A)\;PX_c^{-1}P^T\,M & \text{if } \nu_1 = 0,\; \nu_2 = 1, \\ M - M\;PX_c^{-1}P^T\;(M-A) & \text{if } \nu_1 = 1,\; \nu_2 = 0. \end{cases}$$

Fact 2.2. (Inspired by Corollary 2.1 of [42])
Let the general assumptions and (3.4) *hold. Any eigenvalue λ of $\mathcal{B}A$ that is not equal to 1 satisfies*

$$\left|\lambda - \tfrac{1}{2}\lambda_{\max}\left(M^{-1}\big([A^{-1}]_S\big)^{-1}\right)\right| \;\leq\; \tfrac{1}{2}\lambda_{\max}\left(M^{-1}\big([A^{-1}]_S\big)^{-1}\right)\;,$$

where $[A^{-1}]_S$ is defined by (3.5).

If, in addition, A is symmetric (i.e., SPD), the eigenvalues of $\mathcal{B}A$ are real and such that

$$\max_{\lambda\in\sigma(\mathcal{B}A)\setminus\{1\}} \;\leq\; \lambda_{\max}(M^{-1}A)\;.$$

Fact 2.3. (Corollary 2.2 of [42])
Let the general assumptions and (3.4) *hold. Any eigenvalue λ of $\mathcal{B}A$ that is not equal to 1 satisfies*

$$\Re e(\lambda) \;\geq\; \frac{1}{K_M}\;,$$

where K_M is defined by (2.9).

If, in addition, A is symmetric (i.e., SPD), the eigenvalues of $\mathcal{B}A$ are real and the above inequality is sharp:

$$\min_{\lambda\in\sigma(\mathcal{B}A)\setminus\{1\}} \lambda \;=\; \frac{1}{K_M}\;.$$

Fact 2.4. *(Straightforward corollary of Facts 2.1, 2.2 and 2.3)*
Let the general assumptions and (3.4) *hold, and assume further that A is symmetric (i.e., SPD). Let K_M be defined by (2.9). If $1 \in [K_M^{-1}\,,\,\lambda_{\max}\left(M^{-1}A\right)]$, there holds*

$$K_M\;\lambda_{\max}\big(M^{-1}\left(I-PA_c^{-1}P^TA\right)\big)\;\leq\;\frac{\lambda_{\max}(\mathcal{B}A)}{\lambda_{\min}(\mathcal{B}A)}\leq\;K_M\;\lambda_{\max}(M^{-1}A)\;. \tag{3.6}$$

3.3. Case: A SPD and unsymmetric smoothing scheme

Here we need some additional assumption on the smoother to obtain more insight than in the general case (§3.1) without entering the particular cases explored in the next two subsections. Thus we assume:

$$\begin{cases} A \text{ is SPD}\,, \\ \rho\big(\left(I-X^{-T}A\right)\left(I-X^{-1}A\right)\big) \;<\; 1\;. \end{cases} \tag{3.7}$$

Note that, as already mentioned in §2, the above condition on the spectral radius holds if and only if $X+X^T-A$ is SPD [40, Lemma A.1].

Fact 3.1. (Nonsingularity of $\mathcal{B}$ well known; otherwise new)
Let the general assumptions and (3.7) *hold. Then X and $X_c = P^TX\,P$ are positive definite in $\mathbb{R}^n$ (thus nonsingular). Therefore, the additional condition in Fact 1.1 is satisfied; hence $\mathcal{B}$ is nonsingular and* (3.1) *holds as well as the characterization of the eigenvalues via* (3.2) *and* (3.3).

Fact 3.2. (New)
Let the general assumptions and (3.7) *hold. Any eigenvalue λ of $\mathcal{B}A$ that is not equal to 1 satisfies*

$$\Re e(\lambda) \;\leq\; \lambda_{\max}\left([X^{-1}]_S\,A\right) \;\leq\; \lambda_{\max}\left((X_S)^{-1}A\right)\;,$$

where X_S is defined by (2.8) *and $[X^{-1}]_S$ by* (3.5).

Fact 3.3. (Theorem 2.2 of [17]; sharpness from Theorem 4.3 of [18])
Let the general assumptions and (3.7) *hold. Letting*

$$Z \;=\; X\,(X+X^T-A)^{-1}\,X^T\;,$$

the eigenvalues λ of $\mathcal{B}A$ satisfy

$$\rho(T) \;=\; \max_{\lambda\in\sigma(\mathcal{B}A)}|1-\lambda| \;\leq\; \|T\|_A \;\leq\; \left(1-\frac{1}{K_Z}\right)^{1/2}\;, \tag{3.8}$$

where K_Z is defined by (2.9). Moreover, the last inequality is sharp:

$$\|T\|_A^2 \;=\; 1-\frac{1}{K_Z}\;.$$

3.4. Case: A SPD and symmetric smoothing scheme

Thus:

$$\begin{cases} A \text{ is SPD}\,, \\ M_1 \;= M_2^T \;=\; M\;, \\ \nu_1 \;=\; \nu_2 \;=\; \nu\;. \end{cases} \tag{3.9}$$

Fact 4.1. (Positive definiteness of $\mathcal{B}$ well known (e.g., [57, Chapter 3]); otherwise new)
Let the general assumptions and (3.9) *hold. Then X and $X_c = P^T X\,P$ are SPD. Therefore, the additional condition in Fact 1.1 is satisfied; hence $\mathcal{B}$ is nonsingular and* (3.1) *holds as well as the characterization of the eigenvalues via* (3.2) *and* (3.3). *Moreover, the eigenvalues of $\mathcal{B}A$ are real and positive and $\mathcal{B}$ is SPD, whereas, when $\nu=1$,*

$$\begin{aligned}\mathcal{B}^{-1} =& M\,\left(M+M^T-A\right)^{-1}M^T \\ &- M\,\left(M+M^T-A\right)^{-1}(M^T-A)\,PX_c^{-1}P^T\,(M-A)\,\left(M+M^T-A\right)^{-1}M^T\;.\end{aligned}$$

Fact 4.2. *(Theorem 3.16 or 3.19 of [57])*
Let the general assumptions and (3.9) *hold. There holds*

$$\max_{\lambda\in\sigma(\mathcal{B}A)}\lambda \;=\; 1\;.$$

Fact 4.3. (Theorem 2.2 of [17]; sharpness from Theorem 4.3 of [18])
Let the general assumptions and (3.9) *hold. The eigenvalues λ of $\mathcal{B}A$ satisfy*

$$\lambda \;\geq\; \frac{1}{K_X}\;,$$

where K_X is defined by (2.9). Moreover, the inequality is sharp:

$$1-\rho\,(T) \;=\; 1-\rho\,(I-\mathcal{B}A) \;=\; \min_{\lambda\in\sigma(\mathcal{B}A)}\lambda \;=\; \frac{1}{K_X}\;.$$

3.5. Case: A SPD; same SPD $M_1 = M_2$ for both pre- and post-smoothing

Thus:

$$\begin{cases} A \text{ is SPD}\,, \\ M_1 \;= M_2 \;=\; M \quad \text{ with } M \text{ SPD}\,. \end{cases} \tag{3.10}$$

Observe that here that the condition (2.5) (required when $\overline{\nu} = \nu_1 + \nu_2 > 1$) amounts to

$$\lambda_{\max}(M^{-1}A) \;<\; 2\,.$$

Fact 5.1. (New)
Let the general assumptions and (3.10) *hold. Then X and $X_c = P^T X\, P$ are SPD. Therefore, the additional condition in Fact 1.1 is satisfied; hence $\mathcal{B}$ is nonsingular and* (3.1) *holds as well as the characterization of the eigenvalues via* (3.2) *and* (3.3). *Moreover, the eigenvalues of $\mathcal{B}A$ are real and positive.*

Fact 5.2. (see Fact 4.2 for $\nu_1 = \nu_2$; otherwise, corollary of [42, Corollary 2.1])
Let the general assumptions and (3.10) *hold. Letting $\overline{\nu} = \nu_1 + \nu_2$, the eigenvalues λ of $\mathcal{B}A$ that are not equal to 1 satisfy*

$$\lambda \;\le\; \lambda_{\max}\left(X^{-1}A\right) \;\le\; \begin{cases} 1 & \text{if } \overline{\nu} \text{ is even} \\ 1 + \left(\lambda_{\max}(M^{-1}A) - 1\right)^{\overline{\nu}} & \text{if } \overline{\nu} \text{ is odd}\,. \end{cases}$$

Fact 5.3. (see Fact 4.2 for $\nu_1 = \nu_2$; otherwise, corollary of [42, Theorem 2.1])
Let the general assumptions and (3.10) *hold. The eigenvalues λ of $\mathcal{B}A$ that are real not equal to 1 satisfy*

$$\lambda \;\ge\; \frac{1}{K_X}\,.$$

where K_X is defined by (2.9). Moreover, the inequality is sharp:

$$\min_{\lambda \in \sigma(\mathcal{B}A)\setminus\{1\}} \lambda \;=\; \frac{1}{K_X}\,.$$

Fact 5.4. (New)
Let the general assumptions and (3.10) *hold. Assume that $M = \omega^{-1} M_o$ for some SPD matrix M_o, and let K_{M_o} be defined by (2.9). Letting $\overline{\nu} = \nu_1 + \nu_2$, there holds*

$$\lambda_{\max}(\mathcal{B}A) \;\le\; \begin{cases} \max\left(1\,,\, \omega\, \lambda_{\max}(M_o^{-1}A)\right) & \text{if } \overline{\nu} = 1 \\ 1 & \text{if } \overline{\nu} \text{ is even or } \omega\, \lambda_{\max}(M_o^{-1}A) \le 1 \\ 1 + \left(\omega\, \lambda_{\max}(M_o^{-1}A) - 1\right)^{\overline{\nu}} & \text{otherwise} \end{cases}$$

(with equality when the upper bound is equal to 1),

$$\frac{\overline{\nu}\,\omega}{K_{M_o}} \;\ge\; \lambda_{\min}(\mathcal{B}A) \;\ge\; \begin{cases} \min\left(1\,,\, \frac{\omega}{K_{M_o}}\right) & \text{if } \overline{\nu} = 1 \\ \frac{1 - \left(1 - \omega\, \lambda_{\max}\left(M_o^{-1}A\right)\right)^{\overline{\nu}}}{\lambda_{\max}\left(M_o^{-1}A\right)\, K_{M_o}} & \text{if } \overline{\nu} \text{ is even or } \omega\, \lambda_{\max}(M_o^{-1}A) \le 1 \\ \frac{1}{\lambda_{\max}\left(M_o^{-1}A\right)\, K_{M_o}} & \text{otherwise} \end{cases}$$

and

$$\frac{\lambda_{\max}(\mathcal{B}A)}{\lambda_{\min}(\mathcal{B}A)} \;\leq\; \begin{cases} \max(K_{M_o}\,,\,\omega)\;\max\big(\lambda_{\max}(M_o^{-1}A)\,,\,\omega^{-1}\big) & \text{if } \overline{\nu}=1 \\ \dfrac{\lambda_{\max}\big(M_o^{-1}A\big)\;K_{M_o}}{1-\big(1-\omega\,\lambda_{\max}\big(M_o^{-1}A\big)\big)^{\overline{\nu}}} & \text{if } \overline{\nu} \text{ is even} \\ & \text{or } \omega\,\lambda_{\max}(M_o^{-1}A)\leq 1 \\ \lambda_{\max}\left(M_o^{-1}A\right)\Big(1+\left(\omega\,\lambda_{\max}\left(M_o^{-1}A\right)-1\right)^{\overline{\nu}}\Big)\;K_{M_o} & \text{otherwise} \end{cases}$$

3.6. Analysis of K_G

Fact 6.1. ((3.11): well known (straightforward); (3.12): partly from [48] and partly new)
Let A be an $n\times n$ matrix positive definite in $\mathbb{R}^n$, G an $n\times n$ SPD matrix and P an $n\times n_c$ matrix of rank $n_c<n$. K_G defined in (2.9) is the smallest constant K such that the following weak approximation condition holds:

$$\forall \mathbf{u}\in\mathbb{R}^n\;\exists \mathbf{v}\in\mathbb{R}^{n_c} \;\text{ such that }\; \|\mathbf{u}-P\,\mathbf{v}\|_G^2 \;\leq\; K\,\|\mathbf{u}\|_{A_S}^2\,, \tag{3.11}$$

where A_S is defined by (2.8)*. Further, if A is SPD, K_G is also the smallest constant K such that*

$$\|(I-P\,A_c^{-1}P^TA)\mathbf{u}\|_A^2 \;\leq\; K\,\|A(I-P\,A_c^{-1}P^TA)\mathbf{u}\|_{G^{-1}}^2 \qquad \forall \mathbf{u}\in\mathbb{R}^n\,. \tag{3.12}$$

Fact 6.2. (New)
Let A and G be $n\times n$ SPD matrices, and let P be an $n\times n_c$ matrix of rank $n_c<n$. Letting K_G be defined by (2.9), one has

$$K_G \;\geq\; \big(\lambda_{\max}(G^{-1}A)\big)^{-1}\,.$$

Fact 6.3. (Inspired by [17])
Let A be an $n\times n$ matrix positive definite in $\mathbb{R}^n$, let G_1 and G_2 be $n\times n$ SPD matrices, and let P be an $n\times n_c$ matrix of rank $n_c<n$. Let K_{G_1}, K_{G_2} be defined by (2.9). For any $n_c\times n$ matrix $\widetilde{R}$ such that $\widetilde{R}\,P=I_{n_c}$, there holds

$$\begin{aligned} K_{G_1} \;&\leq\; \sigma^{-1}\,K_{G_2} \\ &\leq\; \sigma^{-1}\max_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}}\frac{\mathbf{v}^T\left(I-\widetilde{R}^TP^T\right)G_2\left(I-P\,\widetilde{R}\right)\mathbf{v}}{\mathbf{v}^T\,A\,\mathbf{v}} \\ &=\; \sigma^{-1}\left(\max_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}}\frac{\left\|\left(I-P\,\widetilde{R}\right)\mathbf{v}\right\|_{G_2}}{\|\mathbf{v}\|_{A_S}}\right)^2 \end{aligned}$$

where

$$\sigma \;=\; \lambda_{\min}\left({G_1}^{-1}G_2\right)\,.$$

Moreover, if

$$P \;=\; \begin{pmatrix} J_{FC} \\ I_{n_c} \end{pmatrix}$$

for some $(n-n_c)\times n_c$ J_{FC}*, one has also*

$$K_{G_1} \;\leq\; \sigma^{-1} \left(\max_{\mathbf{v}=\begin{pmatrix}\mathbf{v}_F\\ \mathbf{v}_C\end{pmatrix}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\left\| \begin{pmatrix} \mathbf{v}_F - J_{FC}\mathbf{v}_C \\ 0 \end{pmatrix} \right\|_{G_2}}{\|\mathbf{v}\|_{A_S}} \right)^2 .$$

Fact 6.4. (Theorem 6.5 of [30])
Let A *be an* $n\times n$ *SPD matrix and* S *a nonsingular* $n\times n$ *matrix. The matrix* $S+S^T-A$ *is SPD if and only if* $\rho\big((I-S^{-T}A)(I-S^{-1}A)\big)<1$*, in which case, letting*

$$Z \;=\; S\,(S+S^T-A)^{-1}\,S^T\;,$$

for any $n\times n$ *positive definite matrix* G*, the constant*

$$\sigma \;=\; \lambda_{\min}\left(Z^{-1}G\right)$$

is also the smallest number such that

$$\left\|\left(I-S^{-1}A\right)\mathbf{v}\right\|_A^2 \;\leq\; \|\mathbf{v}\|_A^2 - \sigma\,\|A\,\mathbf{v}\|_{G^{-1}}^2 \tag{3.13}$$

holds for all $\mathbf{v}\in\mathbb{R}^n$*.*

4. Comments

Remark 1. The necessary and sufficient condition for having $\mathcal{B}$ non singular is a new result. In most previous works, the nonsingularity of $\mathcal{B}$ is not discussed explicitly, but implicitly addressed via lower bounds on (the real part of) the eigenvalues of $\mathcal{B}\,A$, or via less than one upper bounds on the spectral radius of T. In the SPD case with a symmetric smoothing scheme, the positive definiteness of $\mathcal{B}$ is shown in [57, Chapter 3] via spectral equivalence relations, and also explicitly addressed in [52, Theorem 3.5].

Remark 2. It is striking that the results for the case A SPD and a single smoothing step with an SPD M just particularize the results for A positive definite in $\mathbb{R}^n$ (see §3.2), except that the lower bound in Fact 2.3 is sharp in the former case and not in the latter. Thus, in practice, the gap between zero and the closest eigenvalue of $\mathcal{B}\,A$ may be larger for a two-grid scheme applied to an unsymmetric matrix than for the same two-grid scheme applied to its symmetric part (see [42] for an example). This does not mean, however, that the eigenvalues are necessarily better clustered in the nonsymmetric case. Indeed, not only they may have a nonzero imaginary part, but also, as shown in [1, Lemma 3.5],

$$\mathbf{v}^T\,[A^{-1}]_S\,\mathbf{v} \;\leq\; \mathbf{v}^T\;(A_S)^{-1}\;\mathbf{v} \qquad \forall \mathbf{v}\in\mathbb{R}^n\;,$$

and hence $\lambda_{\max}\left(M^{-1}\left([A^{-1}]_S\right)^{-1}\right) \geq \lambda_{\max}\left(M^{-1}A_S\right)$. Therefore, going from a symmetric problem to an unsymmetric one without changing A_S may entail an increase of the magnitude of the largest eigenvalues.

Remark 3. Consider A SPD, and let $T_{\rm post}$ and $T_{\rm sym}$ denote the iterations matrices associated with, respectively, only post-smoothing (i.e., $\nu_1 = 0$) and the corresponding symmetric smoothing scheme (i.e., the same ν_2 and M_2 with now $\nu_1 = \nu_2$ and $M_1 = M_2^T$). Another known fact, which traces back to Lemma 4.2 of [32], is

$$\left(\rho\left(T_{\rm post}\right)\right)^2 \;\leq\; \|T_{\rm post}\|_A^2 \;=\; \|T_{\rm sym}\|_A \;=\; \rho\left(T_{\rm sym}\right)\;. \tag{4.1}$$

Our proof below of Fact 3.3 is based on the combination of Fact 4.3 with this identity, and one may observe that, conversely, (4.1) can be recovered as a corollary of the sharpness results in Facts 3.3 and 4.3.

Most previous works on the SPD case focus on the derivation of bounds on $\rho\left(T_{\rm post}\right)$ via bounds on $\|T_{\rm post}\|_A$, while implicitly, via (4.1), addressing as well the analysis of $\rho(T_{\rm sym})$.

Remark 4. Fact 3.3 or (equivalently, given (4.1)) Fact 4.3 may be seen as the final result of many successive improvements of the convergence analysis in the seminal paper [6].

Restated in our notation, the main convergence result in [6] is based on the assumptions of Fact 3.3 and proves

$$\rho(T) \;\leq\; \left(1-\frac{\sigma}{C}\right)^{1/2}$$

where C is such that

$$\forall \mathbf{u}\in\mathbb{R}^n\;\exists \mathbf{v}\in\mathbb{R}^{n_c} \;\text{ such that }\; \|\mathbf{u}-P\,\mathbf{v}\|^2_{\mathsf{diag}(A)} \;\leq\; C\,\|\mathbf{u}\|_A^2\;,$$

and where σ is such that

$$\|\left(I-X^{-1}A\right)\mathbf{v}\|_A^2 \;\leq\; \|\mathbf{v}\|_A^2 - \sigma\,\|A\,\mathbf{v}\|^2_{(\mathsf{diag}(A))^{-1}}\;.$$

This result may thus be recovered by combining Fact 3.3 with the upper bound on K_Z that is obtained by using $K_Z \leq \sigma^{-1}K_{\mathsf{diag}(A)}$ (Fact 6.3) with $K_{\mathsf{diag}(A)}$ defined via (3.11) (Fact 6.1) and σ as given in (3.13) with $S = X$ (Fact 6.4); the paper [6] also discusses an extension where the block diagonal part of A is used instead of $\mathrm{diag}(A)$. Estimates for σ are further given for different smoothers by bounding below the decrease of the error in energy norm after one smoothing step (i.e., $\|\mathbf{e}\|_A^2 - \|\left(I-S^{-1}A\right)\mathbf{e}\|_A^2$) in function of the residual norm (i.e., $\|\mathbf{r}\|^2_{(\mathsf{diag}(A))^{-1}}$ with $\mathbf{r} = A\,\mathbf{e}$). On the other hand, it is shown how $K_{\mathsf{diag}(A)}$ can be bounded in several cases using the last inequality in Fact 6.3 and assumed properties of the matrix A and the interpolation J_{FC}.

Regarding the two-grid convergence theory, the main new contribution in [48] is the definition of $K_{\mathsf{diag}(A)}$ via (3.12) instead of (3.11), together with a proof that (3.11)

provides a sufficient condition for (3.12). This latter property is referred to as "approximation property", whereas (3.13) is given the name "smoothing property". The introduction of (3.12) allows an enlightening comparison with the condition for V-cycle convergence as stated in [33], which is satisfied when both the smoothing property (3.13) and

$$\|(I - P\,A_c^{-1}P^T A)\mathbf{u}\|_A^2 \;\leq\; K\;\|A\,\mathbf{u}\|_{G^{-1}}^2 \qquad \forall \mathbf{u} \in \mathbb{R}^n\;.$$

hold. Indeed, this latter property boils down to (3.12) whenever restricted to vectors $\mathbf{u}$ in the range of the projector $I - P\,A_c^{-1}P^T A$. Reference [48] also discusses the equivalence between the smoothing property (3.13) and $\lambda_{\min}(Z^{-1}G)$ (see Fact 6.4), a result first proved in [30].

The reference [10] improves the theory for Richardson smoothing iterations. Under the assumptions of Fact 3.3 with, in addition, $\nu_2 = 1$ and $M_2 = \omega^{-1}\|A\|\,I$, it is proven

$$\rho(T) \;\leq\; \left(1 - \frac{\omega(2-\omega)}{C\;\|A\|}\right)^{1/2}$$

where

$$C \;=\; \max_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{v}^T\,\left(I - \widetilde{R}^T P^T\right)\,\left(I - P\,\widetilde{R}\right)\,\mathbf{v}}{\mathbf{v}^T\,A\,\mathbf{v}}\;.$$

for an arbitrary $\widetilde{R}$ such that $\widetilde{R}\,P = I_{n_c}$. This result can also be recovered from Fact 3.3, using the upper bound on K_Z obtained by applying Fact 6.3 with $G_2 = I$,

$$G_1 \;=\; Z \;=\; M_2(M_2 + M_2^T - A)^{-1}M_2^T \;=\; \omega^{-2}\|A\|^2\left(2\,\omega^{-1}\|A\| - A\right)^{-1}$$

and

$$\sigma \;=\; \lambda_{\min}(Z^{-1}) \;=\; \lambda_{\min}\big(\omega\,\|A\|^{-1}(2 - \omega\,\|A\|^{-1}A)\big) \;\geq\; \omega(2-\omega)\,\|A\|^{-1}$$

This approach is extended to arbitrary smoother in [17], where

$$\rho(T) \;\leq\; \left(1 - \frac{1}{K_Z^{(\widetilde{R})}}\right)^{1/2}$$

is shown with

$$K_Z^{(\widetilde{R})} \;=\; \max_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{v}^T\,\left(I - \widetilde{R}^T P^T\right)\,Z\,\left(I - P\,\widetilde{R}\right)\,\mathbf{v}}{\mathbf{v}^T\,A\,\mathbf{v}}\;.$$

As noted in [17], $K_Z^{(\widetilde{R})}$ is minimal for $\widetilde{R} = (P^T Z\,P)^{-1}P^T Z$, which yields the "final" form of Fact 3.3 (except the sharpness statement). Conversely, the above bound is recovered by combining Fact 3.3 and Fact 6.3. Note that in [17] it is also shown how $K_Z^{(\widetilde{R})}$ can be further bounded below and above as a function of $K_{\frac{1}{2}(M_2+M_2^T)}^{(\widetilde{R})}$.

The sharpness of (3.8) is finally first proved in [18]; more precisely, this latter reference analyzes symmetric smoothing schemes and proves the sharpness of the bound in Fact 4.3, which, combined with the identity (4.1) from [32], further shows the sharpness of the bound in Fact 3.3.

The extension of this approach to nonsymmetric and/or indefinite matrices is considered in [42], which introduces most results stated in §3.1 and 3.2, and some more. These results allow, in turn, to also gain some insight on the SPD case: characterization of all the eigenvalues (not only the extremal ones) and sharp result as soon as X is SPD (without requiring a symmetric smoothing scheme).

Remark 5. The above developments show that, from a theoretical viewpoint, the results stated in Facts 3.3 and 4.3 are an improvement over the earlier results, in particular the seminal results in [6]. From a practical viewpoint, however, the advantage is less neat. The constants K_Z or K_X can rarely be analyzed directly for the matrices Z or X corresponding to realistic smoothing schemes. A more viable approach consists then in analyzing K_M for some M "simple enough", and deduce a bound on K_Z or K_X via Fact 6.3 and an analysis of $\sigma^{-1} = \lambda_{\max}(M^{-1}Z)$ or $\sigma^{-1} = \lambda_{\max}(M^{-1}X)$. Of course, among the "simple enough" matrices, $M = \mathrm{diag}(A)$ is a good candidate, which brings us back to essentially the result in [6].

The analysis in §3.2 brings however a new light on the approaches based on the analysis of K_M for "simple" (SPD) M. Indeed, see Fact 2.4, K_M determines the convergence of a basic but feasible two-grid method that uses only a single smoothing step based on M. This weakens the role of $\lambda_{\max}(M^{-1}Z)$ or $\lambda_{\max}(M^{-1}X)$, or equivalently, of the smoothing property constant in (3.13). Ideally, the analysis of these quantities should show that improving the smoothing scheme yields better convergence estimates. However, in practice, as seen on two typical examples below (see Remarks 7 and 8), one often hardly shows that the convergence does not deteriorate with a (supposed) better smoothing scheme. This is clearly a shortcoming of the analysis based on "simple" M: what is gained thanks to a better smoothing scheme is hidden by the loss of sharpness in the analysis.

From an heuristic viewpoint, the analysis of $\lambda_{\max}(M^{-1}Z)$ or $\lambda_{\max}(M^{-1}X)$ (or the related smoothing property constants) appears then less necessary than it was in the past: if a scheme with only one smoothing step is validated via the analysis of K_M, it seems sensible to rely on numerical evidences to check that the convergence is effectively improved by using more steps or a (supposed) better smoother.

It is worth noting that this mixed answer about convergence improvement via smoother enhancement is well in agreement with the AMG philosophy and the related research trends. Most often, with AMG methods, the smoother is kept fixed and not discussed in detail, whereas numerous work address the improvement of the prolongation; to mention just a few, this includes the improvement of the prolongation in classical AMG schemes (see [51] for an overview), the development of "adaptive" or "bootstrap" variants (e.g., [11, 12]), and the smoothing of a tentative prolongation in methods based on "smoothed aggregation" [56].

Remark 6. In view of the preceding remark, it is worth discussing further how K_M determines the convergence of a basic two-grid scheme with a single smoothing step based on M. Clues are given in Fact 2.4. The condition $1 \in [K_M^{-1}\,,\,\lambda_{\max}\left(M^{-1}A\right)]$ is actually a condition on the scaling of M that makes sure that 1 is inside the interval containing the other eigenvalues. When the eigenvalue 1 is outside this interval, the condition number is indeed necessarily larger. The scaling condition is however easy to meet, hence (3.6) reflects well the potentialities of such a grid-scheme.

Further, observe that

$$\lambda_{\max}\big(M^{-1}A\left(I-PA_c^{-1}P^TA\right)\big) = \max_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{v}^T\,A\left(I-PA_c^{-1}P^TA\right)\,\mathbf{v}}{\mathbf{v}^T\,M\,\mathbf{v}}\,,$$

$$\lambda_{\max}\left(M^{-1}A\right) = \max_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{v}^T\,A\,\mathbf{v}}{\mathbf{v}^T\,M\,\mathbf{v}}\,,$$

whereas, if $\mathbf{v}$ is a "rough" vector for which the coarse grid correction is ineffective, one has

$$\frac{\mathbf{v}^T\,A\left(I-PA_c^{-1}P^TA\right)\,\mathbf{v}}{\mathbf{v}^T\,M\,\mathbf{v}} \;\approx\; \frac{\mathbf{v}^T\,A\,\mathbf{v}}{\mathbf{v}^T\,M\,\mathbf{v}}\,.$$

Hence, if there is such a vector $\mathbf{v}$ for which the ratio in the right hand side is close to its maximum, this maximum (i.e., $\lambda_{\max}(M^{-1}A)$) will be approximately equal to the maximum of the left hand side (i.e., to $\lambda_{\max}\big(M^{-1}A\left(I-PA_c^{-1}P^TA\right)\big)$). Then, the upper and lower bounds (3.6) are close to each other; i.e., the condition number associated with the two-grid scheme is practically equal to its upper bound $K_M\,\lambda_{\max}(M^{-1}A)$.

It is worth noting that both this upper bound and the lower bound in (3.6) are invariant under the scaling of M. This is actually a consequence of Fact 1.1 when $X = M$: if M is multiplied by α, all eigenvalues of (3.2), and hence all eigenvalues of $\mathcal{B}A$ but the eigenvalue 1, are multiplied by α^{-1}. Thus the ratio between the largest and the smallest eigenvalues of $\mathcal{B}A$ is unaffected, as long as 1 remains inside the interval containing the other eigenvalues.

This scaling invariance is counter intuitive. For instance, it seems contradict the well known fact that Jacobi smoothing requires damping. The exception to this rule comes from the uncommon simultaneous occurrence of two features. Firstly, only a single smoothing step is considered. When more steps are performed, the scaling of M is crucial, as can be seen with the role of the parameter ω in Fact 5.4; see also the example in Section 4 of [52] for a clear illustration. Secondly, by discussing the condition number, we implicitly assume that the two-grid method is used as a preconditioner. If it is used as a solver, what matters is the spectral radius of $T = I - \mathcal{B}A$, for which an appropriate scaling of M is crucial as well, even in the case of a single smoothing step.

Remark 7. At the light of Remark 5, it is interesting to compare, in the SPD case, a single step of damped Jacobi smoothing with the symmetric Gauss–Seidel smoothing scheme.

In this latter case, $M_1 = \text{low}(A)$ corresponds to forward Gauss–Seidel, $M_2 = \text{upp}(A)$ to backward Gauss–Seidel, and, assuming $\nu_1 = \nu_2 = 1$ for the sake of simplicity, one has $X + X^T - A = M_1 + M_2 - A = \text{diag}(A)$, which is always SPD (the

diagonal entries of an SPD matrix have to be positive). This ensures that (2.5) holds, and hence the results in §3.4 apply, proving

$$\kappa_{\text{SGS}} \;=\; \frac{\lambda_{\max}(\mathcal{B}A)}{\lambda_{\min}(\mathcal{B}A)} \;=\; K_X \;,$$

where, setting $D = \text{diag}(A)$

$$X \;=\; \text{upp}(A)\, D^{-1}\, \text{low}(A) \;.$$

On the other hand, Facts 2.4 shows that, for a single step of damped Jacobi smoothing,

$$\kappa_{\text{Jac}} \;=\; \frac{\lambda_{\max}(\mathcal{B}A)}{\lambda_{\min}(\mathcal{B}A)} \;\leq\; \lambda_{\max}\left(D^{-1}A\right)\; K_D$$

for appropriately chosen damping factor ω (i.e., in the interval $[\left(\lambda_{\max}(D^{-1}A)\right)^{-1}\,,\, K_D]$, to ensure that $1 \in [\omega\, K_D^{-1}\,,\, \lambda_{\max}\left(\omega\, D^{-1}A\right)]$). Moreover, see Remark 6, κ_{Jac} is often either very close or equal to its upper bound. Noting this latter $\overline{\kappa}_{\text{Jac}}$ (i.e., $\overline{\kappa}_{\text{Jac}} = \lambda_{\max}\left(D^{-1}A\right)\; K_D$), one obtains, with Fact 6.3,

$$\frac{\lambda_{\min}(D^{-1}X)\;\overline{\kappa}_{\text{Jac}}}{\lambda_{\max}(D^{-1}A)} \;\leq\; \kappa_{\text{SGS}} \;\leq\; \frac{\overline{\kappa}_{\text{Jac}}}{\lambda_{\max}(D^{-1}A)\;\lambda_{\min}(X^{-1}D)} \;,$$

where $\lambda_{\min}(X^{-1}D)$ can be analyzed at the light of Fact 6.4.

Note that, since $\lambda_{\max}(X^{-1}A) \leq 1$ [1, Theorem 7.16], one has

$$\begin{aligned}
\lambda_{\max}(D^{-1}A)\;\lambda_{\min}(X^{-1}D) &= \max_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{v}^T A\,\mathbf{v}}{\mathbf{v}^T D\,\mathbf{v}} \quad \min_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{v}^T D\,\mathbf{v}}{\mathbf{v}^T X\,\mathbf{v}} \\
&\leq \max_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{v}^T X\,\mathbf{v}}{\mathbf{v}^T D\,\mathbf{v}} \quad \min_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{v}^T D\,\mathbf{v}}{\mathbf{v}^T X\,\mathbf{v}} \\
&= \qquad\qquad\qquad\qquad\qquad\qquad 1 \;.
\end{aligned}$$

Hence, if one has just an analysis of $\overline{\kappa}_{\text{Jac}}$ (K_D is indeed a lot easier to analyze than K_X), the algebraic theory cannot prove that the convergence is indeed better with the symmetric Gauss–Seidel smoothing scheme. But $(\lambda_{\min}(D^{-1}X))^{-1} = \lambda_{\max}(X^{-1}D)$ is by definition larger than $\lambda_{\min}(X^{-1}D)$. Thus the *lower* bound on κ_{SGS} suggests that there is room for an effective improvement.

Remark 8. The results in Fact 5.4 allow to discuss the influence of the number of smoothing steps. However, according the discussion in Remark 5, the insight is limited by the loss of sharpness of the estimates when using more than one smoothing step. Indeed, for $\overline{\nu} > 1$, the upper bound on the condition number $\lambda_{\max}(\mathcal{B}A)/\lambda_{\min}(\mathcal{B}A)$ is not sharp, being obtained via Fact 6.3 applied with $G_1 = X$ and $G_2 = M$ (see the proof in the next section). It is minimal for $\omega = \left(\lambda_{\max}\left(M_o^{-1}A\right)\right)^{-1}$, and is then equal to $\lambda_{\max}\left(M_o^{-1}A\right)\; K_{M_o}$; that is, the value achieved with just one smoothing step ($\overline{\nu} = 1$) for any ω in the interval $[\left(\lambda_{\max}(M_o^{-1}A)\right)^{-1}\,,\, K_{M_o}]$. The observations made in Remark 6 apply however to the case $\overline{\nu} = 1$ (also covered by Fact 2.4), and tell us that the upper

bound is then often practically equal to the exact condition number. On the contrary, for $\overline{\nu} > 1$, the difference between the upper and lower estimates for $\lambda_{\min}(\mathcal{B}\,A)$ shows that there is room for an effective improvement as the number of smoothing steps is increased.

More insight can of course be gained when using additional assumptions. In the Section 4 of [52], one can find an analysis of the particular case where the columns of P coincide with the eigenvectors of $M_o^{-1}A$ associated with the smallest eigenvalues; in particular, Figure 4.1 in [52] clearly shows that, with proper scaling (e.g., ω such that $\lambda_{\max}(M^{-1}A) = 1$), the condition number with two smoothing steps (noted $\kappa_{\rm MG}$ in [52]) is always strictly smaller than with just one smoothing step (noted $\kappa_{\rm DEF}$).

5. Proofs

We start with a technical lemma which is needed to prove Fact 1.1. We next proceed with its proof, which is at the heart of the technique presented here. Indeed, all other Facts *.1 mainly ensure that the additional condition in Fact 1.1 (i.e., the nonsingularity of X_c) is met. Further, all Facts *.2 exploit the first characterization of the eigenvalues obtained in Facts 1.1 (related to $X^{-1}\,A\,\big(I - PA_c^{-1}P^TA\big)$), whereas all Facts *.3 exploit their second characterization (related to $A^{-1}\,X\,\big(I - PX_c^{-1}P^TX\big)$). Often this exploitation is relatively straightforward, hence we condensate the proof of several facts in a single one for the sake of brevity. The proof of Fact 3.3 is presented separately because it relies on the combination of Fact 4.3 with the identity (4.1).

Finally, we present the proofs of Facts 6.* in §3.6, and conclude with the proof of Fact 5.4 (which is delayed because it needs Fact 6.3). We do not give an explicit proof for Fact 2.4 because it is a straightforward corollary of Facts 2.1, 2.2 and 2.3.

Lemma 1. *Let A, M_1, M_2 be nonsingular $n \times n$ matrices, and let ν_1, ν_2 be nonnegative integers. Assume that* (2.3), (2.4) *consistently define non singular matrices X, Y. There holds*

$$\big(I - A\,M_1^{-1}\big)^{\nu_1}\;Y\;\big(I - M_2^{-1}A\big)^{\nu_2} = X - A$$

Proof. Multiply first both sides of (2.3) to the left by $\big(I - M_1^{-1}A\big)^{\nu_1}$, and next both sides of (2.4) to the right by the same matrix. After this transformation, the right hand sides of both equations are equal, showing the equality of the left hand sides; i.e.,

$$\big(I - M_1^{-1}A\big)^{\nu_1}\;\;\big(I - Y^{-1}A\big) = \big(I - X^{-1}A\big)\;\;\big(I - M_1^{-1}A\big)^{\nu_1}\;\;.$$

Canceling equal terms, one obtains

$$-\big(I - M_1^{-1}A\big)^{\nu_1}\;Y^{-1}A = -X^{-1}A\;\big(I - M_1^{-1}A\big)^{\nu_1} = -X^{-1}\;\big(I - A\,M_1^{-1}\big)^{\nu_1}\;A\,.$$

Hence, multiplying both sides to the left by X and to the right by $A^{-1}Y$,

$$X\;\big(I - M_1^{-1}\,A\big)^{\nu_1} = \big(I - A\,M_1^{-1}\big)^{\nu_1}\;Y$$

and therefore

$$\begin{aligned} & \left(I - A\,M_1^{-1}\right)^{\nu_1}\; Y\; \left(I - M_2^{-1}A\right)^{\nu_2} \\ = \;& X\; \left(I - M_1^{-1}\,A\right)^{\nu_1} \left(I - M_2^{-1}\,A\right)^{\nu_2} = X\; \left(I - X^{-1}A\right)\;; \end{aligned}$$

i.e., the required result ∎

Proof of Fact 1.1. We first analyze the eigenvalues of $\mathcal{B}A$ without assuming anything on X_c (to prove the condition on X_c, we shall use the fact that $\mathcal{B}$ is nonsingular if and only if $\mathcal{B}A$ has no eigenvalue equal to 0). Observe that permuting the factors in the definition (2.1) of T does not change its eigenvalues, and therefore that of $\mathcal{B}A$ [23, Theorem 1.3.22]. Hence these eigenvalues in the general case are the same as those in the case where there is no pre-smoothing and only one post-smoothing step is performed with X, for which we have

$$\mathcal{B} = X^{-1} + (I - X^{-1}A)P\,A_c^{-1}P^T\;.$$

Then the eigenvalues of $\mathcal{B}A$ are also the eigenvalues of the generalized eigenvalue problem $X\mathcal{B}A\mathbf{z} = \lambda X\mathbf{z}$, i.e.,

$$\left(A\;\left(I - P\,A_c^{-1}P^TA\right) + X\,P\,A_c^{-1}P^TA\right)\mathbf{z} = \lambda\,X\,\mathbf{z}\;. \tag{5.1}$$

One sees that any vector in the range of P is eigenvector associated with the eigenvalue 1, whose multiplicity is therefore (at least) equal to the rank of P; that is, n_c. For any other eigenvalue (including further occurrence of the eigenvalue 1), there should be a *left* eigenvector $\mathbf{w}^T$ such that $\mathbf{w}^T X\,P\,\mathbf{z}_c = 0$ for any $\mathbf{z}_c$; that is, the other eigenvalues are obtained by solving the problem

$$\begin{cases} \mathbf{w}^T\left(A\;\left(I - P\,A_c^{-1}P^TA\right) + X\,P\,A_c^{-1}P^TA\right) = \lambda\,\mathbf{w}^T\,X \\ \mathbf{w}^T X\,P\;=\;0\;, \end{cases}$$

which may be rewritten

$$\begin{cases} \mathbf{w}^T\,A\;\left(I - P\,A_c^{-1}P^TA\right) = \lambda\,\mathbf{w}^T\,X \\ \mathbf{w}^T X\,P\;=\;0\;. \end{cases} \tag{5.2}$$

Observe here that the left kernel of $A\;\left(I - P\,A_c^{-1}P^TA\right)$ is equal to the range of P. Hence $\mathcal{B}A$ has a zero eigenvalue if and only if there is a non trivial vector $\mathbf{w} = P\,\mathbf{w}_c$ in the range of P such that $\mathbf{w}^T X\,P = \mathbf{w}_c^T P^T X\,P = 0$; that is if and only if $X_c = P^TXP$ has a non trivial vector in its left kernel; i.e., is singular. Thus we have proved the first statement of Fact 1.1.

We now use (5.2) to prove the first characterization of the eigenvalues. Consider the generalized eigenvalue problem

$$\left(A\;\left(I - P\,A_c^{-1}P^TA\right)\right)\mathbf{z} = \lambda\,X\,\mathbf{z}\;,$$

whose eigenvalues are also the eigenvalues of $X^{-1}A\,\left(I - P\,A_c^{-1}P^T A\right)$. The eigenspace associated with the eigenvalue 0 is the null space of the left hand side matrix, which is equal to the range of P. Reasoning as above on (5.1), it follows that the nonzero eigenvalues are also obtained by solving the same problem (5.2), proving the statement related to (3.2).

We next check (3.1) (assuming X_c nonsingular):

$$\begin{aligned}
&\mathcal{B}\left(Y - Y\,\left(I - M_2^{-1}A\right)^{\nu_2}\,PX_c^{-1}P^T\,\left(I - A\,M_1^{-1}\right)^{\nu_1}\,Y\right)\\
&\quad= I\;+\;\left(I - M_2^{-1}A\right)^{\nu_2}\,P\,A_c^{-1}P^T\,\left(I - A\,M_1^{-1}\right)^{\nu_1}\,Y\\
&\qquad\quad-\;\left(I - M_2^{-1}A\right)^{\nu_2}\,PX_c^{-1}P^T\,\left(I - A\,M_1^{-1}\right)^{\nu_1}\,Y\\
&\qquad\quad-\;\left(I - M_2^{-1}A\right)^{\nu_2}\,P\,A_c^{-1}\\
&\qquad\qquad P^T\left(\left(I - A\,M_1^{-1}\right)^{\nu_1}\,Y\,\left(I - M_2^{-1}A\right)^{\nu_2}\right)P\,X_c^{-1}P^T\,\left(I - A\,M_1^{-1}\right)^{\nu_1}\,Y\\
&\quad= I\;+\;\left(I - M_2^{-1}A\right)^{\nu_2}\,P\,A_c^{-1}P^T\,\left(I - A\,M_1^{-1}\right)^{\nu_1}\,Y\\
&\qquad\quad-\;\left(I - M_2^{-1}A\right)^{\nu_2}\,PX_c^{-1}P^T\,\left(I - A\,M_1^{-1}\right)^{\nu_1}\,Y\\
&\qquad\quad-\;\left(I - M_2^{-1}A\right)^{\nu_2}\,P\,A_c^{-1}P^T\,(X - A)\,P\,X_c^{-1}P^T\,\left(I - A\,M_1^{-1}\right)^{\nu_1}\,Y\\
&\quad= I\;+\;\left(I - M_2^{-1}A\right)^{\nu_2}\,P\,A_c^{-1}P^T\,\left(I - A\,M_1^{-1}\right)^{\nu_1}\,Y\\
&\qquad\quad-\;\left(I - M_2^{-1}A\right)^{\nu_2}\,PX_c^{-1}P^T\,\left(I - A\,M_1^{-1}\right)^{\nu_1}\,Y\\
&\qquad\quad-\;\left(I - M_2^{-1}A\right)^{\nu_2}\,P\,\left(A_c^{-1} - X_c^{-1}\right)P^T\,\left(I - A\,M_1^{-1}\right)^{\nu_1}\,Y\\
&\quad= I\,.
\end{aligned}$$

Finally, to prove the second characterization of the eigenvalues, we use the same trick as for the first one: the eigenvalues of $\mathcal{B}A$ in the general case are the same as those in the case where there is no pre-smoothing and only one post-smoothing step is performed with X, for which we have, as just shown,

$$\mathcal{B}^{-1}\;=\;X - (X - A)P\,X_c^{-1}P^T X\,.$$

Then these eigenvalues are also the inverse of the eigenvalues of the generalized eigenvalue problem

$$\left(X\,\left(I - P\,X_c^{-1}P^T X\right) + A\,P\,X_c^{-1}P^T X\right)\mathbf{z}\;=\;\mu\,A\,\mathbf{z}\,.$$

One may observe that this is the same problem as (5.1) in which the roles of A and X have been permuted, whereas the expression (3.3) is also (3.2) after permuting the roles of A and X. Repeating the above reasoning on (5.1) then yields the required result. ∎

Proof of Fact 2.1. Since P has full rank, the positive definiteness in $\mathbb{R}^n$ of $X = M$ entails that of $X_c = P^T X\,P$. ∎

Proof of Fact 3.1. The assumptions (3.7) imply that the matrix

$$\left(X^{-T} + X^{-1} - X^{-T}A\,X^{-1}\right)A\;=\;X^{-T}(X + X^T - A)X^{-1}A$$

has all its eigenvalues in the interval $(0,2)$. Hence $X + X^T - A$ is SPD, showing that X is positive definite in $\mathbb{R}^n$. Since P has full rank, $X_c = P^T X\, P$ is then positive definite in $\mathbb{R}^n$ as well. ∎

Proof of Facts 4.1 and 5.1. If $\nu_1 + \nu_2 = 1$ (Fact 5.1 only), the positive definiteness of $X = M$ is obvious. If not, because the assumptions (3.9) or (3.10) imply that Y and X are symmetric, the condition (2.5) amounts to state the the eigenvalues of $Y^{-1}A$ are in the interval $(0,2)$. Since $X^{-1}A$ has the same eigenvalues as $Y^{-1}A$ [23, Theorem 1.3.22], and since A is SPD, X has then to be SPD as well. Since P has full rank, this further implies the positive definiteness of $X_c = P^T X\, P$. Moreover, when A and X are SPD, it is clear that the nonzero eigenvalues of (3.2) and (3.3) (and hence the eigenvalues of $\mathcal{B}A$) are real and positive. When the smoothing scheme is symmetric (Fact 4.1), $\mathcal{B}$ is in addition symmetric, and thus SPD as soon as the eigenvalues of $\mathcal{B}A$ are positive. ∎

Proof of Fact 2.2. By Fact 1.1, the eigenvalues of $\mathcal{B}A$ different from 1 are the nonzero eigenvalues of (3.2) with $X = M$, which are themselves the same as those of $\left(I - PA_c^{-1}P^TA\right) M^{-1}\, A$. That is, they are the nonzero solutions of the generalized eigenvalue problem

$$\left(I - PA_c^{-1}P^TA\right) M^{-1}\, \mathbf{z} \;=\; \lambda\, A^{-1}\, \mathbf{z}\,.$$

Multiplying both sides to the left by P^TA shows that $P^T\mathbf{z} = \mathbf{0}$. Hence:

$$\lambda^{-1} \;=\; \frac{\mathbf{z}^*\, A^{-1}\, \mathbf{z}}{\mathbf{z}^*\, \left(I - PA_c^{-1}P^TA\right) M^{-1}\, \mathbf{z}} \;=\; \frac{\mathbf{z}^*\, A^{-1}\, \mathbf{z}}{\mathbf{z}^*\, M^{-1}\, \mathbf{z}}\,.$$

Thus, for any eigenvalue λ of $\mathcal{B}A$ not equal to 1,

$$\Re\mathrm{e}\left(\lambda^{-1}\right) \;\geq\; \lambda_{\min}\left(M\,[A^{-1}]_S\right) \;=\; \left(\lambda_{\max}\left(M^{-1}\left([A^{-1}]_S\right)^{-1}\right)\right)^{-1}\,.$$

The proof for general A is completed by observing that, for any complex number λ and positive real ζ, $\Re\mathrm{e}\left(\lambda^{-1}\right) \geq \zeta$ if and only if $|\lambda - (2\,\zeta)^{-1}| \leq (2\,\zeta)^{-1}$.

When A is in addition SPD, it is clear that the eigenvalues of (3.2) are real; knowing this, the general result just proven straightforwardly yields the given upper bound. ∎

Proof of Facts 3.2, 4.2 and 5.2. By Fact 1.1, the eigenvalues λ of $\mathcal{B}A$ different from 1 coincide with the nonzero eigenvalue of (3.2). Letting $Q = I - A^{1/2}P\, A_c^{-1}P^TA^{1/2}$, these eigenvalues are also that of $A^{1/2}X^{-1}A^{1/2}Q$, which are themselves the same as those of $A^{1/2}X^{-1}A^{1/2}Q^2$, or, equivalently [23, Theorem 1.3.22], those of $Q\, A^{1/2}X^{-1}A^{1/2}Q$. Hence, there holds (since $Q = Q^T$, and since X is positive definite in $\mathbb{R}^n$, as shown in

Fact *.1 of the same subsection):

$$\begin{aligned}\Re\mathrm{e}(\lambda) &\leq \max_{\mathbf{z}\in\mathbb{C}^n\,,\,\|\mathbf{z}\|=1} \Re\mathrm{e}\left(\mathbf{z}^*\, Q\, A^{1/2} X^{-1} A^{1/2} Q\,\mathbf{z}\right)\\ &\leq \max_{\mathbf{z}\in\mathbb{C}^n\,,\,\|\mathbf{z}\|=1} \Re\mathrm{e}\left(\mathbf{z}^*\, A^{1/2} X^{-1} A^{1/2}\,\mathbf{z}\right)\\ &= \max_{\mathbf{z}\in\mathbb{C}^n\,,\,\|\mathbf{z}\|=1} \Re\mathrm{e}\left(\mathbf{z}^*\, A^{1/2} [X^{-1}]_S A^{1/2}\,\mathbf{z}\right)\\ &= \lambda_{\max}\left([X^{-1}]_S\, A\right)\ . \end{aligned} \tag{5.3}$$

This is the left inequality of Fact 3.2, whereas the right one follows because, as shown in [1, Lemma 3.5], for any matrix X positive definite in $\mathbb{R}^n$,

$$\mathbf{v}^T\,[X^{-1}]_S\,\mathbf{v}\ \leq\ \mathbf{v}^T\ (X_S)^{-1}\ \mathbf{v} \qquad \forall \mathbf{v}\in\mathbb{R}^n\ .$$

On the other hand, Fact 4.2 also follows because X is symmetric (hence $[X^{-1}]_S = X^{-1}$), whereas (2.4) and (3.4) imply

$$I - A^{1/2} X^{-1} A^{1/2} = \left(I - A^{1/2} M_1^{-1} A^{1/2}\right)^{\nu} \left(I - A^{1/2} M_1^{-T} A^{1/2}\right)^{\nu}\ ,$$

showing that $\lambda_{\max}(X^{-1}A) \leq 1$, since the matrix in the right hand side is symmetric and nonnegative definite. Then (5.3) implies $\lambda_{\max}(\mathcal{B}A) = 1$ since 1 is in any case an eigenvalue of $\mathcal{B}A$.

Finally, Fact 5.2 is also a straightforward corollary of (5.3), X being again symmetric, and using

$$\lambda_{\max}(X^{-1}A) = \max_{\lambda\in\sigma(M^{-1}A)} 1-(1-\lambda)^{\overline{\nu}}$$

to derive the right inequality. ∎

Proof of Fact 2.3. By Fact 1.1, any eigenvalue λ of $\mathcal{B}A$ different from 1 is the inverse of an eigenvalue μ of (3.3). For any such eigenvalue, there exists some $\mathbf{z}\in\mathbb{C}^n$ such that (using $X = M$ because we are under the assumptions of Fact 2.3)

$$\mu\ =\ \frac{\mathbf{z}^T\, M\big(I - P\left(P^T M\, P\right)^{-1} P^T M\big)\,\mathbf{z}}{\mathbf{z}^T\, A\,\mathbf{z}}\ ,$$

implying, with $\lambda = \mu^{-1}$

$$\begin{aligned}\Re\mathrm{e}(\lambda) =& \frac{\Re\mathrm{e}\big(\mathbf{z}^*\, A\,\mathbf{z}\big)}{\mathbf{z}^*\, M\big(I - P\left(P^T M\, P\right)^{-1} P^T M\big)\,\mathbf{z}}\\ =& \frac{\mathbf{z}^*\, A_S\,\mathbf{z}}{\mathbf{z}^*\, M\big(I - P\left(P^T M\, P\right)^{-1} P^T M\big)\,\mathbf{z}}\\ \geq& \left(\max_{\mathbf{v}\in\mathbb{C}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{v}^T\, M\big(I - P\left(P^T M\, P\right)^{-1} P^T M\big)\,\mathbf{v}}{\mathbf{v}^T\, A_S\,\mathbf{v}}\right)^{-1}\\ =& K_M^{-1}\ ,\end{aligned}$$

the last equality following because the matrices in the numerator and the denominator of the previous expression are symmetric, hence the maximum over $\mathbf{z} \in \mathbb{C}^n$ coincides with the maximum over $\mathbf{z} \in \mathbb{R}^n$. ∎

Proof of Facts 4.3 and 5.3. These facts are straightforward consequences of Facts 4.1 and 5.1 (which state the positive definiteness of X) and of Fact 1.1: the second characterization of the eigenvalues implies that $\lambda_{\min}(\mathcal{B}A)$ is the inverse of the largest eigenvalue of$A^{-1}\,X\left(I - PX_c^{-1}P^TX\right)$,which coincide with K_X when X and A are SPD. ∎

Proof of Fact 3.3. One has:

$$\begin{aligned}
\big(\rho(T)\big)^2 &= \big(\rho\big((I-X^{-1}A)\left(I-P\,A_c^{-1}P^TA\right)\big)\big)^2 \\
&\leq \|(I-X^{-1}A)\left(I-P\,A_c^{-1}P^TA\right)\|_A^2 \\
&= \rho\big(A^{-1/2}\left(I-A\,P\,A_c^{-1}P^T\right)(I-AX^{-T})A(I-X^{-1}A)\left(I-P\,A_c^{-1}P^TA\right)A^{-1/2}\big) \\
&= \rho\big(\left(I-P\,A_c^{-1}P^TA\right)(I-X^{-T}A)(I-X^{-1}A)\left(I-P\,A_c^{-1}P^TA\right)\big) \\
&= \rho\big((I-X^{-T}A)(I-X^{-1}A)\left(I-P\,A_c^{-1}P^TA\right)^2\big) \\
&= \rho\big((I-X^{-1}A)\left(I-P\,A_c^{-1}P^TA\right)(I-X^{-T}A)\big) \\
&= 1-K_Z^{-1}\,,
\end{aligned}$$

where the last equality follows from Fact 4.3 applied to the case $\nu_1 = \nu_2 = 1$ and $M_2 = M_1^T = X$, entailing that the matrix $\mathcal{B}A$ analyzed in Fact 4.3 is such that $I - \mathcal{B}A = (I-X^{-1}A)\left(I-P\,A_c^{-1}P^TA\right)(I-X^{-T}A)$; further, with these settings the relevant constant is $K_{X(X+X^T-A)^{-1}X^T}$; that is, K_Z (remembering the definition of Z in Fact 3.3). ∎

Proof of Fact 6.1. Letting $G_c = P^TG\,P$, any $\mathbf{u} \in \mathbb{R}^n$ can be decomposed as $\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2$ with $\mathbf{u}_1 = P\,G_c^{-1}P^TG\,\mathbf{u}$ and $\mathbf{u}_2 = \left(I - P\,G_c^{-1}P^TG\right)\mathbf{u}$. Because $\mathbf{u}_2^TG\,P = \mathbf{u}^T\left(I - GP\,G_c^{-1}P^T\right)G\,P = 0$, one has for any $\mathbf{v} \in \mathbb{R}^{n_c}$,

$$\|\mathbf{u} - P\,\mathbf{v}\|_G^2 \;=\; \|\mathbf{u}_1 - P\,\mathbf{v}\|_G^2 + \|\mathbf{u}_2\|_G^2\,.$$

Hence the norm in the left hand side is minimal for $\mathbf{v} = G_c^{-1}P^TG\,\mathbf{u}$, and, further,

$$\min_{\mathbf{v}\in\mathbb{R}^{n_c}} \|\mathbf{u} - P\,\mathbf{v}\|_G \;=\; \|\mathbf{u}_2\|_G \;=\; \|\left(I - P\,G_c^{-1}P^TG\right)\mathbf{u}\|_G\,.$$

The equivalence between the approximation property (3.11) and (2.9) straightforwardly follows.

On the other hand, letting $Q = I - P\,A_c^{-1}P^TA$, and noting that $Q^2 = Q$, the inverse

of the smallest K satisfying (3.12) is:

$$\min_{\substack{\mathbf{u}\in\mathbb{R}^n\\ Q\mathbf{u}\neq\mathbf{0}}} \frac{\mathbf{u}^T Q^T A\, G^{-1} A\, Q\,\mathbf{u}}{\mathbf{u}^T Q^T A\, Q\,\mathbf{u}} = \min_{\substack{\mathbf{u}\in\mathbb{R}^n\\ Q\mathbf{u}\neq\mathbf{0}}} \frac{\mathbf{u}^T Q^T (Q^T A\, G^{-1} A\, Q)\, Q\,\mathbf{u}}{\mathbf{u}^T Q^T A\, Q\,\mathbf{u}}$$
$$= \min_{\substack{\mathbf{u}\in\mathbb{R}^n\\ Q\mathbf{u}\neq\mathbf{0}}} \frac{\mathbf{u}^T Q^T A\, G^{-1} A\, Q\,\mathbf{u}}{\mathbf{u}^T A\,\mathbf{u}} = \min_{\substack{\lambda\,\in\,\sigma(A^{-1}Q^T A\, G^{-1} A\, Q)\\ \lambda\neq 0}} \lambda\,.$$

But $A^{-1}Q^T = Q\,A^{-1}$ and $Q\,G^{-1}A\,Q$ has the same nonzero eigenvalues as $G^{-1}A\,Q^2 = G^{-1}A\,Q$ [23, Theorem 1.3.22], which itself coincides with the matrix (3.2) for the case $X = G$. These nonzero eigenvalues are therefore the inverse of the nonzero eigenvalues of $A^{-1}G(I - P\,(P^T G\,P)^{-1}G)$, whose largest one is precisely K_G. ∎

Proof of Fact 6.2. Because, for all $\mathbf{v}\in\mathbb{R}^n$,

$$\mathbf{v}^T A\,\mathbf{v} \;\leq\; \lambda_{\max}(G^{-1}A)\;\mathbf{v}^T G\,\mathbf{v}\;,$$

one has

$$K_G \geq \big(\lambda_{\max}(G^{-1}A)\big)^{-1} \max_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{v}^T G\big(I - P\,\big(P^T G\,P\big)^{-1} P^T G\big)\,\mathbf{v}}{\mathbf{v}^T G\,\mathbf{v}}$$
$$= \big(\lambda_{\max}(G^{-1}A)\big)^{-1}\,,$$

the last equality following from the fact that $I - G^{1/2}P\,\big(P^T G\,P\big)^{-1} P^T G^{1/2}$ is a orthogonal projector which cannot be identically zero because $n_c < n$, hence its norm is equal to 1. ∎

Proof of Fact 6.3. For any H SPD and $\mathbf{v}\in\mathbb{R}^n$,

$$\begin{aligned}
&\mathbf{v}^T H\big(I - P\,\big(P^T H\,P\big)^{-1} P^T H\big)\,\mathbf{v}\\
&= \mathbf{v}^T \left(I - \widetilde{R}^T P^T\right) H\big(I - P\,\big(P^T H\,P\big)^{-1} P^T H\big) \left(I - P\,\widetilde{R}\right)\mathbf{v}\\
&\leq \mathbf{v}^T \left(I - \widetilde{R}^T P^T\right) H \left(I - P\,\widetilde{R}\right)\mathbf{v}\\
&\leq \lambda_{\max}(G_2^{-1}H)\;\mathbf{v}^T \left(I - \widetilde{R}^T P^T\right) G_2 \left(I - P\,\widetilde{R}\right)\mathbf{v}\,.
\end{aligned}$$

The inequality $K_{G_1} \leq \sigma^{-1} K_{G_2}$ then follows by taking $H = G_1$ and $\widetilde{R} = (P^T G_2\,P)^{-1}P^T G_2$. On the other hand, considering $H = G_2$ shows that

$$K_{G_2} \;\leq\; \max_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{v}^T \left(I - \widetilde{R}^T P^T\right) G_2 \left(I - P\,\widetilde{R}\right)\mathbf{v}}{\mathbf{v}^T A\,\mathbf{v}}\,.$$

Finally, the last statement is obtained by applying the previous one with $\widetilde{R} = \begin{pmatrix} 0 & I_{n_c} \end{pmatrix}$, which indeed satisfies $\widetilde{R}\,P = I_{n_c}$ for the given P. ∎

Proof of Fact 6.4. The equivalence between the positive definiteness of $S + S^T - A$ and condition $\rho\big((I - S^{-T}A)(I - S^{-1}A)\big) < 1$ has already been mentioned [40, Lemma A.1]. Assuming this, one has

$$\begin{aligned}
&\min_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\|\mathbf{v}\|_A^2 - \|\left(I - S^{-1}A\right)\mathbf{v}\|_A^2}{\|A\,\mathbf{v}\|_{G^{-1}}^2} \\
&= \min_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{v}^T A\left(S^{-T} + S^{-1}\right)A\,\mathbf{v} - \mathbf{v}^T A\,S^{-T}A\,S^{-1}A\,\mathbf{v}}{\|A\,\mathbf{v}\|_{G^{-1}}^2} \\
&= \min_{\mathbf{w}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{w}^T\left(S^{-T} + S^{-1} - S^{-T}AS^{-1}\right)\mathbf{w}}{\mathbf{w}^T G^{-1}\,\mathbf{w}} \\
&= \min_{\mathbf{w}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{w}^T S^{-T}\left(S + S^T - A\right)S^{-1}\,\mathbf{w}}{\mathbf{w}^T G^{-1}\,\mathbf{w}} \\
&= \lambda_{\max}\left(Z^{-1}G\right)\ .
\end{aligned}$$

∎

Proof of Fact 5.4. The results about $\lambda_{\max}(\mathcal{B}A)$ straightforwardly follow from Fact 5.2, whereas the results about $\lambda_{\min}(\mathcal{B}A)$ are based on Facts 5.3 and 6.3, which yield

$$\frac{\lambda_{\max}(M^{-1}X)}{K_M} = \frac{1}{\lambda_{\min}(M^{-1}X)\;K_M} \;\geq\; \lambda_{\min}(\mathcal{B}A) = \frac{1}{K_X} \;\geq\; \frac{\lambda_{\min}(X^{-1}M)}{K_M}\ .$$

Further

$$X^{-1}M \;=\; \big(I - (I - M^{-1}A)^{\overline{\nu}}\big)A^{-1}M$$

and hence

$$\sigma(X^{-1}M) \;\subset\; \left[\inf_{\lambda\in(0\,,\,\lambda_{\max}(M^{-1}A)]} \frac{1-(1-\lambda)^{\overline{\nu}}}{\lambda}\,,\; \sup_{\lambda\in(0\,,\,\lambda_{\max}(M^{-1}A)]} \frac{1-(1-\lambda)^{\overline{\nu}}}{\lambda}\right]\ .$$

Consider then the function

$$f(\lambda) \;=\; \frac{1-(1-\lambda)^{\overline{\nu}}}{\lambda} \;=\; \sum_{k=0}^{\overline{\nu}-1}(1-\lambda)^k\ .$$

It is clear from the right hand side that it is bounded above by $\overline{\nu}$ over the interval $(0, 2]$. Hence, $\lambda_{\max}(M^{-1}X) \leq \overline{\nu}$, yielding the upper bound on $\lambda_{\min}(\mathcal{B}A)$ together with $K_M = \omega^{-1}K_{M_o}$.

On the other hand, $f(\lambda)$ is decreasing in the interval $(0\,,\,2]$ when $\overline{\nu}$ is even: this is clear from the middle expression regarding the interval $[1\,,\,2]$ (the numerator is decreasing and the denominator is increasing) and from the right expression regarding the interval $(0\,,\,1]$ (all terms of the sum are decreasing). Hence we have then

$$\lambda_{\min}(X^{-1}M) \geq \frac{1-(1-\lambda_{\max}(M^{-1}A))^{\overline{\nu}}}{\lambda_{\max}(M^{-1}A)}\,,$$

yielding the lower bound on $\lambda_{\min}(\mathcal{B}A)$ for $\overline{\nu}$ even. Moreover, this lower bound also holds for $\overline{\nu}$ odd when $\lambda_{\max}(M^{-1}A) \leq 1$, because the above argument still applies to show that $f(\lambda)$ is decreasing in the interval $(0\,,\,1]$.

For $\overline{\nu}$ odd and $\lambda_{\max}(M^{-1}A) \geq 1$, we use the fact that $f(\lambda) \geq 1$ for $\lambda \in (0\,,\,1]$ (as shown above, the minimum is attained for $\lambda = 1$), whereas, for $\lambda \in (1\,,\,2)$, we anyway have $f(\lambda) \geq \lambda^{-1}$ since the numerator of the middle expression is larger than 1 for $\lambda > 1$ and $\overline{\nu}$ odd. Hence, $\lambda_{\min}(X^{-1}M) \geq (\lambda_{\max}(M^{-1}A))^{-1}$, which allows to conclude the proof of the lower bound on $\lambda_{\min}(\mathcal{B}A)$, whereas the upper bound on the ratio $\lambda_{\max}(\mathcal{B}A)/\lambda_{\min}(\mathcal{B}A)$ straightforwardly follows from the just proved upper and lower bounds. ■

Acknowledgments I thank Artem Napov, Michiel Hochstenbach and the referees for useful comments and suggestions.

References

[1] O. AXELSSON, *Iterative Solution Methods*, Cambridge University Press, Cambridge, UK, 1994.

[2] A. H. BAKER, R. D. FALGOUT, T. V. KOLEV, AND U. M. YANG, *Multigrid smoothers for ultraparallel computing*, SIAM J. Sci. Comput., 33 (2011), pp. 2864–2887.

[3] R. BANK AND C. DOUGLAS, *Sharp estimates for multigrid rates of convergence with general smoothing and acceleration*, SIAM J. Numer. Anal., 22 (1985), pp. 617–633.

[4] R. E. BANK, T. F. DUPONT, AND H. YSERENTANT, *The hierarchical basis multigrid method*, Numer. Math., 52 (1988), pp. 427–458.

[5] J. H. BRAMBLE, J. E. PASCIAK, J. WANG, AND J. XU, *Convergence estimates for product iterative methods with applications to domain decomposition*, Math. Comp., 57 (1991), pp. 1–21.

[6] A. BRANDT, *Algebraic multigrid theory: The symmetric case*, Appl. Math. Comput., 19 (1986), pp. 23–56.

[7] A. BRANDT, *General highly accurate algebraic coarsening*, Electron. Trans. Numer. Anal., 10 (2000), pp. 1–20.

[8] A. BRANDT, S. F. MCCORMICK, AND J. W. RUGE, *Algebraic multigrid (AMG) for sparse matrix equations*, in Sparsity and its Application, D. J. Evans, ed., Cambridge University Press, Cambridge, 1984, pp. 257–284.

[9] J. J. BRANNICK AND R. D. FALGOUT, *Compatible relaxation and coarsening in algebraic multigrid*, SIAM J. Sci. Comput., 32 (2010), pp. 1393–1416.

[10] M. BREZINA, A. J. CLEARY, R. D. FALGOUT, V. E. HENSON, J. E. JONES, T. A. MANTEUFFEL, S. F. MCCORMICK, AND J. W. RUGE, *Algebraic multigrid based on element interpolation (AMGe)*, SIAM J. Sci. Comput., 22 (2000), pp. 1570–1592.
[11] M. BREZINA, R. FALGOUT, S. MACLACHLAN, T. MANTEUFFEL, S. MCCORMICK, AND J. RUGE, *Adaptive smoothed aggregation (αSA)*, SIAM Review, 47 (2005), pp. 317–346.
[12] M. BREZINA, C. KETELSEN, T. MANTEUFFEL, S. MCCORMICK, M. PARK, AND J. RUGE, *Relaxation-corrected bootstrap algebraic multigrid (rBAMG)*, Numer. Linear Algebra Appl., 19 (2012), pp. 178–193.
[13] M. BREZINA, T. MANTEUFFEL, S. MCCORMICK, J. RUGE, AND G. SANDERS, *Towards adaptive smoothed aggregation (αSA) for nonsymmetric problems*, SIAM J. Sci. Comput., 32 (2010), pp. 14–39.
[14] T. CHARTIER, R. D. FALGOUT, V. E. HENSON, J. JONES, T. MANTEUFFEL, S. MCCORMICK, J. RUGE, AND P. S. VASSILEVSKI, *Spectral AMGe (ρAMGe)*, SIAM J. Sci. Comput., 25 (2004), pp. 1–26.
[15] J. E. DENDY, *Black box multigrid for nonsymmetric problems*, Appl. Math. Comput., 13 (1983), pp. 261–283.
[16] Y. A. ERLANGGA AND R. NABBEN, *Multilevel projection-based nested Krylov iteration for boundary value problems*, SIAM J. Sci. Comput., 30 (2008), pp. 1572–1595.
[17] R. D. FALGOUT AND P. S. VASSILEVSKI, *On generalizing the algebraic multigrid framework*, SIAM J. Numer. Anal., 42 (2004), pp. 1669–1693.
[18] R. D. FALGOUT, P. S. VASSILEVSKI, AND L. T. ZIKATANOV, *On two-grid convergence estimates*, Numer. Linear Algebra Appl., 12 (2005), pp. 471–494.
[19] J. FRANK AND C. VUIK, *On the construction of deflation-based preconditioners*, SIAM J. Sci. Comput., 23 (2001), pp. 442–462.
[20] W. HACKBUSCH, *Convergence of multi-grid iterations applied to difference equations*, Math. Comp., 34 (1980), pp. 425–440.
[21] W. HACKBUSCH, *Multi-grid Methods and Applications*, Springer, Berlin, 1985.
[22] L. HOGBEN, ed., *Handbook of Linear Algebra*, CRC Press, Boca Raton, 2007.
[23] R. HORN AND C. JOHNSON, *Matrix Analysis, 2nd Ed.*, Cambridge University Press, New York, 2013.
[24] O. E. LIVNE, *Coarsening by compatible relaxation*, Numer. Linear Algebra Appl., 11 (2004), pp. 205–227.
[25] S. P. MACLACHLAN AND L. N. OLSON, *Theoretical bounds for algebraic multigrid performance: review and analysis*, Numer. Linear Algebra Appl., 21 (2014), pp. 194–220.
[26] J. MAITRE AND F. MUSY, *Algebraic formalization of the multigrid method in the symmetric and positive definite case – a convergence estimation for the V-cycle*, in Multigrid Methods for Integral and Differential Equations (Bristol, 1983), D. Paddon and H. Holsein, eds., vol. 3 of Institute of Mathematics and Its Applications Conference Series, Oxford, 1985, Clarendon Press, p. 213223.
[27] J. MANDEL, *Algebraic study of multigrid methods for symmetric, definite problems*, Appl. Math. Comput., 25 (1988), pp. 39–56.
[28] J. MANDEL, *Balancing domain decomposition*, Comm. Numer. Methods Engrg., 9 (1993), pp. 233–241.
[29] J. MANDEL AND M. BREZINA, *Balancing domain decomposition for problems with large jumps in coefficients*, Math. Comp., 65 (1996), pp. 1387–1401.
[30] J. MANDEL, S. F. MCCORMICK, AND J. W. RUGE, *An algebraic theory for multigrid methods for variational problems*, SIAM J. Numer. Anal., 25 (1988), pp. 91–110.
[31] S. MCCORMICK, *An algebraic interpretation of multigrid methods*, SIAM J. Numer. Anal.,

19 (1982), pp. 548–560.
[32] S. McCormick, *Multigrid methods for variational problems: Further results*, SIAM J. Numer. Anal., 21 (1984), pp. 255–263.
[33] S. McCormick, *Multigrid methods for variational problems: general theory for the V-cycle*, SIAM J. Numer. Anal., 22 (1985), pp. 634–643.
[34] R. Nabben and C. Vuik, *A comparison of deflation and coarse grid correction applied to porous media flow*, SIAM J. Numer. Anal., 42 (2004), pp. 1631–1647.
[35] R. Nabben and C. Vuik, *A comparison of deflation and the balancing preconditioner*, SIAM J. Sci. Comput., 27 (2006), pp. 1742–1759.
[36] R. Nabben and C. Vuik, *A comparison of abstract versions of deflation, balancing and additive coarse grid correction preconditioners*, Numer. Linear Algebra Appl., 15 (2008), pp. 355–372.
[37] A. Napov and Y. Notay, *Comparison of bounds for V-cycle multigrid*, Appl. Numer. Math., 60 (2010), pp. 176–192.
[38] A. Napov and Y. Notay, *Algebraic analysis of aggregation-based multigrid*, Numer. Linear Algebra Appl., 18 (2011), pp. 539–564.
[39] A. Napov and Y. Notay, *An algebraic multigrid method with guaranteed convergence rate*, SIAM J. Sci. Comput., 34 (2012), pp. A1079–A1109.
[40] Y. Notay, *Algebraic multigrid and algebraic multilevel methods: A theoretical comparison*, Numer. Linear Algebra Appl., 12 (2005), pp. 419–451.
[41] Y. Notay, *Convergence analysis of perturbed two-grid and multigrid methods*, SIAM J. Numer. Anal., 45 (2007), pp. 1035–1044.
[42] Y. Notay, *Algebraic analysis of two-grid methods: The nonsymmetric case*, Numer. Linear Algebra Appl., 17 (2010), pp. 73–96.
[43] Y. Notay, *Aggregation-based algebraic multigrid for convection-diffusion equations*, SIAM J. Sci. Comput., 34 (2012), pp. A2288–A2316.
[44] Y. Notay and A. Napov, *Further comparison of additive and multiplicative coarse grid correction*, Appl. Numer. Math., 65 (2013), pp. 53–62.
[45] Y. Notay and P. S. Vassilevski, *Recursive Krylov-based multigrid cycles*, Numer. Linear Algebra Appl., 15 (2008), pp. 473–487.
[46] P. Oswald, *Multilevel Finite Element Approximation: Theory and Applications*, Teubner Skripte zur Numerik, Teubner, Stuttgart, 1994.
[47] A. Quarteroni and A. Valli, *Domain Decomposition Methods for Partial Differential Equations*, Oxford Science Publications, Oxford, 1999.
[48] J. W. Ruge and K. Stüben, *Algebraic multigrid (AMG)*, in Multigrid Methods, S. F. McCormick, ed., vol. 3 of Frontiers in Applied Mathematics, SIAM, Philadelphia, PA, 1987, pp. 73–130.
[49] M. Sala and R. S. Tuminaro, *A new Petrov–Galerkin smoothed aggregation preconditioner for nonsymmetric linear systems*, SIAM J. Sci. Comput., 31 (2008), pp. 143–166.
[50] B. Smith, P. Bjørstad, and W. Gropp, *Domain Decomposition*, Cambridge University Press, Cambridge, 1996.
[51] K. Stüben, *An introduction to algebraic multigrid*, in Trottenberg et al. [55], 2001, pp. 413–532. Appendix A.
[52] J. Tang, S. MacLachlan, R. Nabben, and C. Vuik, *A comparison of two-level preconditioners based on multigrid and deflation*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 1715–1739.
[53] J. Tang, R. Nabben, C. Vuik, and Y. Erlangga, *Comparison of two-level preconditioners derived fromdeflation, domain decomposition and multigrid methods*, J. Sci. Comput., 39

(2009), pp. 340–370.
[54] A. TOSELLI AND W. WIDLUND, *Domain Decomposition*, Springer Ser. Comput. Math. 34, Springer-Verlag, Berlin, 2005.
[55] U. TROTTENBERG, C. W. OOSTERLEE, AND A. SCHÜLLER, *Multigrid*, Academic Press, London, 2001.
[56] P. VANĚK, M. BREZINA, AND J. MANDEL, *Convergence of algebraic multigrid based on smoothed aggregation*, Numer. Math., 88 (2001), pp. 559–579.
[57] P. S. VASSILEVSKI, *Multilevel Block Factorization Preconditioners*, Springer, New York, 2008.
[58] C. VUIK, A. SEGAL, L. EL YAAKOUBI, AND E. DUFOUR, *A comparison of various deflation vectors applied to elliptic problems with discontinuous coefficients*, Appl. Numer. Math., 41 (2002), pp. 219–233.
[59] C. VUIK, A. SEGAL, AND J. A. MEIJERINK, *An efficient preconditioned cg method for the solution of a class of layered problems with extreme contrasts in the coefficients*, J. Comput. Physics, 152 (1999), pp. 385–403.
[60] C. VUIK, A. SEGAL, J. A. MEIJERINK, AND G. T. WIJMA, *The construction of projection vectors for a deflated ICCG method applied to problems with extreme contrasts in the coefficients*, J. Comput. Physics, 172 (2001), pp. 426–450.
[61] R. WIENANDS AND C. W. OOSTERLEE, *On three-grid Fourier analysis for multigrid*, SIAM J. Sci. Comput., 23 (2001), pp. 651–671.
[62] J. XU, *Iterative methods by space decomposition and subspace correction*, SIAM Review, 34 (1992), pp. 581–613.
[63] H. YSERENTANT, *Old and new convergence proofs for multigrid methods*, Acta Numer., 2 (1993), pp. 285–326.