

# Algebraic multigrid and algebraic multilevel methods: a theoretical comparison

Y. Notay<sup>\*,†</sup>

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

Dedicated to Owe Axelsson on the occasion of his 70th birthday

## SUMMARY

We consider algebraic methods of the two-level type for the iterative solution of large sparse linear systems. We assume that a fine/coarse partitioning and an algebraic interpolation have been defined in one way or another, and review different schemes that may be built with these ingredients. This includes algebraic multigrid (AMG) schemes, two-level approximate block factorizations, and several methods that exploit generalized hierarchical bases. We develop their theoretical analysis in a unified way, gathering some known results, rewriting some other and stating some new. This includes lower bounds, that is, we do not only investigate sufficient conditions of convergence, but also look at necessary conditions. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: algebraic multigrid; multilevel; C.B.S. constant; preconditioning

## 1. INTRODUCTION

For the iterative solution of (very) large sparse linear systems

$$A\mathbf{u} = \mathbf{b} \tag{1}$$

many recent works focus on the design of efficient algebraic multigrid or multilevel methods. The aim is to achieve multigrid-like speed of convergence [1, 2] with a robust algorithm that works essentially in a black box fashion, using no additional information besides the matrix structure and the matrix entries. Such techniques include algebraic multigrid (AMG) methods, multilevel approximate block factorizations, and several methods that exploit generalized hierarchical bases (see Section 2 for details and references).

---

\*Correspondence to: Y. Notay, Service de Métrologie Nucléaire, Université Libre de Bruxelles (C.P. 165/84), 50, Av. F.D. Roosevelt, B-1050 Brussels, Belgium.

†E-mail: ynotay@ulb.ac.be

Contract/grant sponsor: Fonds National de la Recherche Scientifique, Maître de recherches

Most of these methods share some common features. They are based on the recursive use of a two-level scheme, and basically the same two ingredients enter the definition of these two-level schemes.

The first of these ingredients is a partitioning of the unknowns in fine and coarse grid ones, according to which the system matrix is permuted (in general only implicitly) and written in the  $2 \times 2$  block form

$$A = \begin{pmatrix} A_{FF} & A_{FC} \\ A_{CF} & A_{CC} \end{pmatrix} \quad (2)$$

Here and in the following,  $F$  refers to the  $F$  set, i.e. the set of fine grid unknowns, and  $C$  to the  $C$  set, i.e. the set of coarse grid unknowns.

The second ingredient is then an ‘interpolation’ matrix  $J_{FC}$  that interpolates a vector defined on the  $C$  set onto the  $F$  set. To this matrix, one associates a prolongation matrix

$$p = \begin{pmatrix} J_{FC} \\ I \end{pmatrix} \quad (3)$$

and the Galerkin coarse grid matrix

$$\hat{A}_C = p^T A p = A_{CC} + A_{CF} J_{FC} + J_{FC}^T A_{FC} + J_{FC}^T A_{FF} J_{FC} \quad (4)$$

Many works discuss algorithms and strategies to perform the  $F/C$  partitioning and set up corresponding interpolation matrices, which is a subject of paramount importance. Here we consider a side problem: assuming these ingredients have been defined in one way or another, how to compare the different schemes that have been proposed, and, in particular, what can we say about their respective conditions of convergence? Note that we confine ourselves to the two-level schemes, assuming that the coarse grid matrix can be inverted exactly. We also restrict our attention to symmetric and positive definite matrices  $A$ . Although the algorithms presented below may in general be applied to non-symmetric matrices as well, their analyses indeed traditionally focus on the symmetric and positive definite case.

Several previous works address the algebraic analysis of schemes considered here [3–13]. In general, we do not improve them in the sense that we do not try to obtain better bounds. Our goal is indeed different. We primarily aim at developing the analysis of all the methods in a way that allows to better see their similarities and their differences. Doing so, we sometimes derive bounds that are slightly less accurate than those appeared in previous works, but offer a complementary information, being easier to interpret in the context of this study. We also try to develop more general analyses embracing at once a wider scope of applications.

These results inform us on sufficient conditions of convergence. In this study, we also complement this information by a discussion of *necessary* conditions of convergence, based on *lower* bounds. This novel approach allows to assess the relevance of the upper bounds and hence to develop the comparison on a firmer basis.

The paper is organized as follows. In Section 2, we present the different families of methods, stating their main properties and giving the corresponding algorithms. Their analysis is developed in Section 3. A few concluding remarks are given in Section 4.

### 1.1. Notation

All matrices and vectors are real. Their dimensions are conformed with dimensions used in the context. When a  $2 \times 2$  partitioning is referred for a matrix, it is assumed a partitioning of form (2), and the vectors are partitioned accordingly.

We note  $\rho(C)$  the spectral radius of any square matrix  $C$ , whereas  $\|C\|$  is its usual 2-norm, that is,  $\|C\|^2 = \rho(C^T C)$ ; if  $A$  is a symmetric positive definite matrix, we note  $\|\cdot\|_A$  the matrix norm induced by the corresponding energy norm, that is,  $\|C\|_A = \|A^{1/2} C A^{-1/2}\|$ . Eventually, if  $C$  has real eigenvalues (e.g. it is similar to a symmetric matrix),  $\lambda_{\min}(C)$  and  $\lambda_{\max}(C)$ , respectively, stand for the smallest and the largest eigenvalue of  $C$ .

## 2. TWO-LEVEL SCHEMES

### 2.1. Algebraic multigrid (AMG)

AMG methods are now quite popular. Pioneer works are due to Brandt [7], Ruge and Stüben [11, 14] and Stüben [15] giving rise to ‘classical’ AMG, which is fairly well presented in Reference [12]. Besides, AMG methods are developed in numerous recent works, see, e.g. References [9, 16–24].

All these schemes obey the same general template as classical multigrid algorithms, alternating smoothing steps and coarse grid corrections. In the framework stated above, the iteration matrix for the coarse grid correction is

$$I - p\hat{A}_C^{-1} p^T A \quad (5)$$

whereas the iteration matrix associated to one smoothing step is

$$I - M^{-1} A \quad (6)$$

where  $M$  stands for the smoothing operator.

In practice, one most often uses both pre- and post-smoothing, performed in a symmetric way, so as to preserve the symmetry of the global operator. This leads to iteration matrices of the form

$$(I - M^{-1} A)(I - p\hat{A}_C^{-1} p^T A)(I - M^{-T} A) \quad (7)$$

We give below an algorithm that implements such AMG schemes. We consider the case where AMG is used as preconditioner, that is, the given algorithm computes the action of  $B_{\text{AMG}}^{-1}$  as defined from

$$I - B_{\text{AMG}}^{-1} A = (I - M^{-1} A)(I - p\hat{A}_C^{-1} p^T A)(I - M^{-T} A) \quad (8)$$

This is unusual, but, in practice, for robustness and efficiency reasons, AMG is often used as a preconditioner for, e.g. the conjugate gradient method. Our motivation here is a bit different. The other schemes presented below all belong to the preconditioning family. Then, using the same format to present AMG allows to better highlight the practical differences between the methods.

**Algorithm 1** (*AMG as preconditioner*)

$\mathbf{v} = B_{\text{AMG}}^{-1} \mathbf{r}$  computed as

1.  $\mathbf{t} = M^{-\text{T}} \mathbf{r}; \mathbf{w} = \mathbf{r} - A\mathbf{t}$
2.  $\mathbf{y}_C = \mathbf{w}_C + J_{FC}^{\text{T}} \mathbf{w}_F$
3. Solve  $\hat{A}_C \mathbf{z}_C = \mathbf{y}_C$
4.  $\mathbf{z}_F = J_{FC} \mathbf{z}_C$
5.  $\mathbf{v} = \mathbf{t} + \mathbf{z} + M^{-1}(\mathbf{w} - A\mathbf{z})$

2.2. *Hierarchical basis block-diagonal preconditioning (HBBD)*

This scheme and the following two originate from the observation that using hierarchical bases improves the conditioning of finite element matrices [5, 25–27]. Most works in this direction consider standard hierarchical bases for regular meshes with geometric refinement. However, as observed in Reference [28], the matrix

$$J = \begin{pmatrix} I & J_{FC} \\ & I \end{pmatrix} \quad (9)$$

defines a generalized basis transformation that allows to exploit these ideas in the algebraic framework introduced in Section 1.

The system matrix in this generalized hierarchical basis is

$$\hat{A} = J^{\text{T}} A J = \begin{pmatrix} A_{FF} & A_{FC} + A_{FF} J_{FC} \\ A_{CF} + J_{FC}^{\text{T}} A_{FF} & \hat{A}_C \end{pmatrix} \quad (10)$$

Thus the basis transformation leaves unchanged the top left block, whereas the bottom right block is equal to the Galerkin coarse grid matrix  $\hat{A}_C$ .

A first approach to precondition this matrix consists in using a block-diagonal approximation of the form

$$\hat{B}_{\text{HBBD}} = \begin{pmatrix} Q_{FF} & \\ & \hat{A}_C \end{pmatrix} \quad (11)$$

where  $Q_{FF}$  stands for some approximation to  $A_{FF}$ , which is assumed symmetric and positive definite. This preconditioner is expressed in the transformed basis and, in practice, it is more convenient to bring it back to the original basis. This gives

$$B_{\text{HBBD}} = J^{-\text{T}} \hat{B}_{\text{HBBD}} J^{-1}$$

hence

$$\begin{aligned} B_{\text{HBBD}}^{-1} &= J \hat{B}_{\text{HBBD}}^{-1} J^{\text{T}} \\ &= \begin{pmatrix} I & J_{FC} \\ & I \end{pmatrix} \begin{pmatrix} Q_{FF}^{-1} & \\ & \hat{A}_C^{-1} \end{pmatrix} \begin{pmatrix} I & \\ J_{FC}^{\text{T}} & I \end{pmatrix} \\ &= q Q_{FF}^{-1} q^{\text{T}} + p \hat{A}_C^{-1} p^{\text{T}} \end{aligned}$$

where  $p$  is given by (3) and

$$q = \begin{pmatrix} I \\ 0 \end{pmatrix} \tag{12}$$

The corresponding algorithm is given below.

The recursive use of this approach leads to so-called parallel multilevel preconditioners [27, 29] and additive AMLI methods [30, 31].

**Algorithm 2** (*Preconditioning by HBBDD*)

$\mathbf{v} = B_{\text{HBBDD}}^{-1} \mathbf{r}$  computed as

1.  $\mathbf{y}_F = Q_{FF}^{-1} \mathbf{r}_F$
2.  $\mathbf{y}_C = \mathbf{r}_C + J_{FC}^T \mathbf{r}_F$
3. Solve  $\hat{A}_C \mathbf{v}_C = \mathbf{y}_C$
4.  $\mathbf{z}_F = J_{FC} \mathbf{v}_C$
5.  $\mathbf{v}_F = \mathbf{z}_F + \mathbf{y}_F$

2.3. *Hierarchical basis block-factorization preconditioning (HBBF)*

Here one also exploits the matrix in the generalized hierarchical form (10), but, following an idea tracing back to [5, 32], one preconditions this matrix with an approximate block factorization of the form

$$\hat{B}_{\text{HBBF}} = \begin{pmatrix} I & \\ (A_{CF} + J_{FC}^T A_{FF}) Q_{FF}^{-1} & I \end{pmatrix} \begin{pmatrix} Q_{FF} & \\ & \hat{A}_C \end{pmatrix} \begin{pmatrix} I & Q_{FF}^{-1} (A_{FC} + A_{FF} J_{FC}) \\ & I \end{pmatrix} \tag{13}$$

which is actually a kind of inexact block Gauss–Seidel preconditioning. Transforming back to the original basis gives

$$\begin{aligned} B_{\text{HBBF}}^{-1} &= J \hat{B}_{\text{HBBF}}^{-1} J^T \\ &= \begin{pmatrix} I & -Q_{FF}^{-1} A_{FC} + (I - Q_{FF}^{-1} A_{FF}) J_{FC} \\ & I \end{pmatrix} \begin{pmatrix} Q_{FF}^{-1} & \\ & \hat{A}_C^{-1} \end{pmatrix} \\ &\quad \times \begin{pmatrix} I & -A_{CF} Q_{FF}^{-1} + J_{FC}^T (I - A_{FF} Q_{FF}^{-1}) \\ & I \end{pmatrix} \\ &= q Q_{FF}^{-1} q^T + \tilde{p} \hat{A}_C^{-1} \tilde{p}^T \end{aligned}$$

where  $q$  is given by (12) and

$$\tilde{p} = \begin{pmatrix} -Q_{FF}^{-1} A_{FC} + (I - Q_{FF}^{-1} A_{FF}) J_{FC} \\ I \end{pmatrix} \tag{14}$$

The corresponding algorithm is given below. This approach is at the root of AMLI methods [6, 13, 33–37].

**Algorithm 3** (Preconditioning by HBBF)

$\mathbf{v} = B_{\text{HBBF}}^{-1} \mathbf{r}$  computed as

1.  $\mathbf{y}_F = Q_{FF}^{-1} \mathbf{r}_F$
2.  $\mathbf{y}_C = \mathbf{r}_C - A_{CF} \mathbf{y}_F + J_{FC}^T (\mathbf{r}_F - A_{FF} \mathbf{y}_F)$
3. Solve  $\hat{A}_C \mathbf{v}_C = \mathbf{y}_C$
4.  $\mathbf{z}_F = J_{FC} \mathbf{v}_C$
5.  $\mathbf{v}_F = \mathbf{z}_F + Q_{FF}^{-1} (\mathbf{r}_F - A_{FC} \mathbf{v}_C - A_{FF} \mathbf{z}_F)$

## 2.4. Hierarchical basis multigrid method (HBMG)

As its name indicates, the hierarchical basis multigrid method [38] bridges a gap between multigrid and the approaches based on (generalized) hierarchical bases. It is indeed close to HBBF, but at the same time corresponds to a standard multigrid scheme, except that only  $F$  unknowns are relaxed during the smoothing process. That is,

$$I - B_{\text{HBMG}}^{-1} A = (I - RA)(I - p\hat{A}_C^{-1} p^T A)(I - R^T A) \quad (15)$$

with

$$R = \begin{pmatrix} Q_{FF}^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

for some non-singular approximation  $Q_{FF}$  to  $A_{FF}$ . Here, non-symmetric approximations are allowed, but, as seen below, the resulting  $B_{\text{HBMG}}$  is positive definite if and only if

$$Q_{FF}^{-1} + Q_{FF}^{-T} - Q_{FF}^{-1} A_{FF} Q_{FF}^{-T} = Q_{FF}^{-1} (Q_{FF} + Q_{FF}^T - A) Q_{FF}^{-T}$$

is positive definite. This holds if and only if

$$\|I - Q_{FF}^{-1} A_{FF}\|_{A_{FF}} < 1$$

or, equivalently, if and only if

$$\rho(I - (\frac{1}{2}(Q_{FF} + Q_{FF}^T))^{-1} A_{FF}) < 1$$

(see Lemma A.1 of Appendix A). In other words, the symmetric part of  $Q_{FF}$  has to define a convergent iterative process for  $A_{FF}$ . This condition is naturally satisfied with Gauss–Seidel and symmetric Gauss–Seidel preconditioners originally considered in Reference [38].

Now, some manipulations yield

$$B_{\text{HBMG}}^{-1} = R + R^T - RAR^T + (I - RA)p\hat{A}_C^{-1} p^T (I - AR^T)$$

Further,

$$(I - RA)p = \begin{pmatrix} -Q_{FF}^{-1} A_{FC} + (I - Q_{FF}^{-1} A_{FF}) J_{FC} \\ I \end{pmatrix} = \tilde{p}$$

( $\tilde{p}$  is thus the same as in HBBF, see (14), except that  $Q_{FF}$  here is allowed to be non-symmetric). One then obtains

$$B_{\text{HBMG}}^{-1} = q(Q_{FF}^{-1} + Q_{FF}^{-T} - Q_{FF}^{-1} A_{FF} Q_{FF}^{-T}) q^T + \tilde{p} \hat{A}_C^{-1} \tilde{p}^T$$

with  $q$  defined by (12). The corresponding algorithm is given below.

Eventually, in view of later theoretical developments, it is also interesting to express this preconditioner in the generalized hierarchical basis. This gives

$$\hat{B}_{\text{HBMG}} = \begin{pmatrix} I & & \\ (A_{CF} + J_{FC}^T A_{FF}) Q_{FF}^{-T} & I & \end{pmatrix} \begin{pmatrix} (Q_{FF}^{-1} + Q_{FF}^{-T} - Q_{FF}^{-1} A_{FF} Q_{FF}^{-T})^{-1} & \\ & \hat{A}_C \end{pmatrix} \\ \times \begin{pmatrix} I & Q_{FF}^{-1} (A_{FC} + A_{FF} J_{FC}) \\ & I \end{pmatrix} \tag{16}$$

Comparing with (13), the similarity is striking.

**Algorithm 4** (*Preconditioning by HBMG*)

$\mathbf{v} = B_{\text{HBMG}}^{-1} \mathbf{r}$  computed as

1.  $\mathbf{y}_F = Q_{FF}^{-T} \mathbf{r}_F$
2.  $\mathbf{y}_C = \mathbf{r}_C - A_{CF} \mathbf{y}_F + J_{FC}^T (\mathbf{r}_F - A_{FF} \mathbf{y}_F)$
3. Solve  $\hat{A}_C \mathbf{v}_C = \mathbf{y}_C$
4.  $\mathbf{z}_F = \mathbf{y}_F + J_{FC} \mathbf{v}_C$
5.  $\mathbf{v}_F = \mathbf{z}_F + Q_{FF}^{-1} (\mathbf{r}_F - A_{FC} \mathbf{v}_C - A_{FF} \mathbf{z}_F)$

2.5. *Multilevel block factorization (MBF)*

Here we consider the methods that bypass the hierarchical form (10) and perform directly a block incomplete factorization of the matrix  $A$  in its original form. Since

$$A = \begin{pmatrix} I & \\ A_{CF} A_{FF}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{FF} & \\ & S_A \end{pmatrix} \begin{pmatrix} I & A_{FF}^{-1} A_{FC} \\ & I \end{pmatrix} \tag{17}$$

where

$$S_A = A_{CC} - A_{CF} A_{FF}^{-1} A_{FC} \tag{18}$$

is the Schur complement, this leads to preconditioners of the form

$$B_{\text{MBF}} = \begin{pmatrix} I & \\ A_{CF} P_{FF}^{-1} & I \end{pmatrix} \begin{pmatrix} P_{FF} & \\ & S \end{pmatrix} \begin{pmatrix} I & P_{FF}^{-1} A_{FC} \\ & I \end{pmatrix} \tag{19}$$

where  $P_{FF}$  and  $S$  are approximations to  $A_{FF}$  and  $S_A$ , respectively. Note that we could use  $Q_{FF}$  as above to note the approximation to  $A_{FF}$ , but we deliberately avoid this because, as will be seen later, this approximation plays here a wider role than in previous methods.

There are numerous such approaches (e.g. References [39–60]). Some of them do not fit exactly in the above framework and a few incorporate multigrid ideas such as the combined use with a smoother, but we do not want to enter these details here. Most of these methods construct the approximate Schur complement  $S$  algebraically, using ILU-like techniques justified either heuristically or theoretically. Here again, we do not want to enter these considerations, our purpose being much less a comparison of these methods between themselves than a general comparison of this approach with AMG, HBBD, HBBF and HBMG.

In this view, note that if an interpolation matrix  $J_{FC}$  has been defined, it is possible to use the corresponding Galerkin coarse grid matrix  $\hat{A}_C$  as approximate Schur complement [28, 55]. In the following section, we particularize the analysis to the case  $S = \hat{A}_C$ , as this gives a closer comparison with other approaches. Concerning the general case, we state the results in terms of function of the extremal eigenvalues of  $S^{-1}S_A$ , without investigating further how these quantities may be bounded for the different methods.

Now, it is interesting to derive an expression of  $B_{\text{MBF}}^{-1}$ , as we did for other approaches. One obtains

$$B_{\text{MBF}}^{-1} = \begin{pmatrix} I & -P_{FF}^{-1}A_{FC} \\ & I \end{pmatrix} \begin{pmatrix} P_{FF}^{-1} & \\ & S \end{pmatrix} \begin{pmatrix} I & \\ -A_{CF}P_{FF}^{-1} & I \end{pmatrix} = qP_{FF}^{-1}q^T + \bar{p}S^{-1}\bar{p}^T$$

where

$$\bar{p} = \begin{pmatrix} -P_{FF}^{-1}A_{FC} \\ I \end{pmatrix} \quad (20)$$

The corresponding algorithm is given below.

**Algorithm 5** (*Preconditioning by MBF*)

$\mathbf{v} = B_{\text{MBF}}^{-1}\mathbf{r}$  computed as

1.  $\mathbf{y}_F = P_{FF}^{-1}\mathbf{r}_F$
2.  $\mathbf{y}_C = \mathbf{r}_C - A_{CF}\mathbf{y}_F$
3. Solve  $S\mathbf{v}_C = \mathbf{y}_C$
4.  $\mathbf{v}_F = P_{FF}^{-1}(\mathbf{r}_F - A_{FC}\mathbf{v}_C)$

### 3. THEORETICAL ANALYSIS

#### 3.1. The strengthened Cauchy–Bunyakowski–Schwarz (C.B.S.) inequality

The so-called strengthened C.B.S. inequality and the associated constant play a key role in the analysis of multilevel methods [3, 4, 8]. We first recall the definition of the C.B.S. constant.

*Definition 1*

Let

$$A = \begin{pmatrix} A_{FF} & A_{FC} \\ A_{CF} & A_{CC} \end{pmatrix}$$

be a symmetric positive definite matrix partitioned in  $2 \times 2$  block form. The C.B.S. constant associated with this matrix and this partitioning is

$$\gamma = \max_{\substack{\mathbf{v}_F \neq 0 \\ \mathbf{v}_C \neq 0}} \frac{\mathbf{v}_F^T A_{FC} \mathbf{v}_C}{(\mathbf{v}_F^T A_{FF} \mathbf{v}_F)^{1/2} (\mathbf{v}_C^T A_{CC} \mathbf{v}_C)^{1/2}} \quad (21)$$

In the following lemma we gather the most important properties of this constant. These results are well known (e.g. Reference [3, Lemma 9.2 and Corollary 9.4]).

*Lemma 1*

Let  $A$  be a symmetric positive definite matrix partitioned in  $2 \times 2$  block form, and let  $\gamma$  be the associated C.B.S. constant.

Then:

(1)

$$\begin{aligned}\gamma^2 &= \max_{\mathbf{v}_C \neq 0} \frac{\mathbf{v}_C^T A_{CF} A_{FF}^{-1} A_{FC} \mathbf{v}_C}{\mathbf{v}_C^T A_{CC} \mathbf{v}_C} \\ &= \max_{\mathbf{v}_F \neq 0} \frac{\mathbf{v}_F^T A_{FC} A_{CC}^{-1} A_{CF} \mathbf{v}_F}{\mathbf{v}_F^T A_{FF} \mathbf{v}_F}\end{aligned}$$

(2)

$$\lambda_{\min}(A_{CC}^{-1} S_A) = \lambda_{\min}(A_{FF}^{-1} S_A^{(F)}) = 1 - \gamma^2$$

and

$$\lambda_{\max}(A_{CC}^{-1} S_A) \leq 1, \quad \lambda_{\max}(A_{FF}^{-1} S_A^{(F)}) \leq 1$$

where

$$\begin{aligned}S_A &= A_{CC} - A_{CF} A_{FF}^{-1} A_{FC} \\ S_A^{(F)} &= A_{FF} - A_{FC} A_{CC}^{-1} A_{CF}\end{aligned}$$

(3)

$$\lambda_{\min}(D^{-1}A) = 1 - \gamma, \quad \lambda_{\max}(D^{-1}A) = 1 + \gamma$$

where

$$D = \begin{pmatrix} A_{FF} & \\ & A_{CC} \end{pmatrix}$$

*Proof*

See References [3, Lemma 9.2] for (1) and (2) and Reference [3, Corollary 9.4] for a proof that  $1 - \gamma$  and  $1 + \gamma$  are lower and upper bounds on the eigenvalues of  $D^{-1}A$ . We give a proof of the sharpness of these bounds for the sake of completeness. Let  $\mathbf{v}_C$  be an eigenvector of  $A_{CC}^{-1} A_{CF} A_{FF}^{-1} A_{FC}$  with eigenvalue  $\gamma^2$  (such an eigenvector exists by (1)). One checks that

$$\begin{pmatrix} A_{FF} & A_{FC} \\ A_{CF} & A_{CC} \end{pmatrix} \begin{pmatrix} \gamma^{-1} A_{FF}^{-1} A_{FC} \mathbf{v}_C \\ \mathbf{v}_C \end{pmatrix} = (1 + \gamma) \begin{pmatrix} A_{FF} & \\ & A_{CC} \end{pmatrix} \begin{pmatrix} \gamma^{-1} A_{FF}^{-1} A_{FC} \mathbf{v}_C \\ \mathbf{v}_C \end{pmatrix}$$

and that

$$\begin{pmatrix} A_{FF} & A_{FC} \\ A_{CF} & A_{CC} \end{pmatrix} \begin{pmatrix} -\gamma^{-1} A_{FF}^{-1} A_{FC} \mathbf{v}_C \\ \mathbf{v}_C \end{pmatrix} = (1 - \gamma) \begin{pmatrix} A_{FF} & \\ & A_{CC} \end{pmatrix} \begin{pmatrix} -\gamma^{-1} A_{FF}^{-1} A_{FC} \mathbf{v}_C \\ \mathbf{v}_C \end{pmatrix}$$

showing that the extremal eigenvalues are effectively equal to their respective bounds.  $\square$

In general, for finite element matrices expressed in the usual nodal basis,  $\gamma$  is very close to 1 and therefore these results are of little help. But the above framework applies as well

to matrices in hierarchical form, and several works show that the C.B.S. constant is nicely bounded away from 1 in such cases [61–66]. How this may be extended to generalized algebraic basis transformations is considered in Section 3.5. In the following, we note  $\hat{\gamma}$  the C.B.S. constant associated with matrices  $\hat{A}$  in the generalized hierarchical form (10).

Here, it is important to observe that both matrices  $A$  and  $\hat{A}$  have the same Schur complement, that is,

$$A_{CC} - A_{CF}A_{FF}^{-1}A_{FC} = \hat{A}_C - (A_{CF} + J_{FC}^T A_{FF})A_{FF}^{-1}(A_{FC} + A_{FF}J_{FC})$$

(see References [4, 13]). Hence, when  $\hat{\gamma}$  is bounded away from 1, Lemma 1 also shows that  $\hat{A}_C$  is spectrally equivalent to the Schur complement of the original matrix  $A$ .

In (3) of Lemma 1, the focus is on exact block Jacobi preconditioning. We conclude this subsection with a theorem showing that the C.B.S. constant characterizes more widely the condition number associated with any block-diagonal preconditioning.

*Theorem 2*

Let  $A$  be a symmetric positive definite matrix partitioned in  $2 \times 2$  block form, and let  $\gamma$  be the associated C.B.S. constant. Let

$$B = \begin{pmatrix} B_{FF} & \\ & B_{CC} \end{pmatrix}$$

be a block-diagonal matrix with symmetric positive definite diagonal blocks  $B_{FF}$ ,  $B_{CC}$ . Let

$$\begin{aligned} \mu_M &= \max(\lambda_{\max}(B_{FF}^{-1}A_{FF}), \lambda_{\max}(B_{CC}^{-1}A_{CC})) \\ \mu_m &= \min(\lambda_{\min}(B_{FF}^{-1}A_{FF}), \lambda_{\min}(B_{CC}^{-1}A_{CC})) \end{aligned}$$

Then:

$$\begin{aligned} \max(\mu_M, (1 + \gamma)\mu_m) &\leq \lambda_{\max}(B^{-1}A) \leq (1 + \gamma)\mu_M \\ \min(\mu_m, (1 - \gamma)\mu_M) &\geq \lambda_{\min}(B^{-1}A) \geq (1 - \gamma)\mu_m \end{aligned}$$

and

$$\begin{aligned} \kappa(B^{-1}A) &\leq \frac{1 + \gamma}{1 - \gamma} \frac{\mu_M}{\mu_m} \\ \kappa(B^{-1}A) &\geq \max\left(\frac{\mu_M}{\mu_m}, \frac{1}{1 - \gamma}\right) \end{aligned}$$

*Proof*

The lower bound on  $\lambda_{\min}(B^{-1}A)$  and the upper bounds on  $\lambda_{\max}(B^{-1}A)$  and  $\kappa(B^{-1}A)$  are straightforward corollaries of Lemma 1. Further,

$$\begin{aligned} \lambda_{\max}(B^{-1}A) &= \max_{\mathbf{z} \neq 0} \frac{\mathbf{z}^T A \mathbf{z}}{\mathbf{z}^T B \mathbf{z}} \\ &\geq \max\left(\max_{\mathbf{z}_F \neq 0} \frac{\mathbf{z}_F^T A_{FF} \mathbf{z}_F}{\mathbf{z}_F^T B_{FF} \mathbf{z}_F}, \max_{\mathbf{z}_C \neq 0} \frac{\mathbf{z}_C^T A_{CC} \mathbf{z}_C}{\mathbf{z}_C^T B_{CC} \mathbf{z}_C}\right) \\ &= \mu_M \end{aligned}$$

whereas, letting  $D$  be the block-diagonal part of  $A$ ,

$$\begin{aligned}\lambda_{\max}(B^{-1}A) &= \max_{\mathbf{z} \neq 0} \frac{\mathbf{z}^T A \mathbf{z}}{\mathbf{z}^T D \mathbf{z}} \frac{\mathbf{z}^T D \mathbf{z}}{\mathbf{z}^T B \mathbf{z}} \\ &\geq \left( \max_{\mathbf{z} \neq 0} \frac{\mathbf{z}^T A \mathbf{z}}{\mathbf{z}^T D \mathbf{z}} \right) \left( \min_{\mathbf{z} \neq 0} \frac{\mathbf{z}^T D \mathbf{z}}{\mathbf{z}^T B \mathbf{z}} \right) \\ &= (1 + \gamma) \mu_m\end{aligned}$$

This proves the lower bound on  $\lambda_{\max}(B^{-1}A)$ . A similar reasoning gives the upper bound on  $\lambda_{\min}(B^{-1}A)$ . The lower bound on  $\kappa(B^{-1}A)$  is then obtained by taking the ratio of  $\mu_M$  and the smallest of the upper bounds on  $\lambda_{\min}(B^{-1}A)$ .  $\square$

The upper bound on the condition number is not new and may even be improved as in Reference [3, Theorem 9.3] (We do not display the latter expression because it is not easy to interpret). The lower bound, however, tells us something more: there is really no way to obtain an efficient block-diagonal preconditioning if the C.B.S. constant is not bounded away from 1.

### 3.2. Analysis of HBBD

Here, it suffices to apply Theorem 2 to the system matrix and its preconditioner, both expressed in the generalized hierarchical basis. We omit the proof because it is straightforward.

#### Theorem 3 (analysis of HBBD)

Let  $A$  be a symmetric positive definite matrix partitioned in  $2 \times 2$  block form, and let  $J_{FC}$  be some interpolation matrix. Let  $\hat{A}$  be the matrix resulting from the application of the generalized basis transformation defined by (9), (10), and let  $\hat{\gamma}$  be the C.B.S. constant associated with  $\hat{A}$ . Let  $B_{\text{HBBD}}$  be the HBBD preconditioning matrix, as defined in Section 2.2 (with symmetric positive definite approximation  $Q_{FF}$  to  $A_{FF}$ ). Then:

$$\begin{aligned}\kappa(B_{\text{HBBD}}^{-1}A) &\leq \frac{1 + \hat{\gamma} \lambda_M}{1 - \hat{\gamma} \lambda_m} \\ \kappa(B_{\text{HBBD}}^{-1}A) &\geq \max \left( \frac{\lambda_M}{\lambda_m}, \frac{1}{1 - \hat{\gamma}} \right)\end{aligned}$$

where

$$\lambda_M = \lambda_{\max}(Q_{FF}^{-1}A_{FF}), \quad \lambda_m = \lambda_{\min}(Q_{FF}^{-1}A_{FF})$$

Thus, this approach can be efficient if and only if  $\hat{\gamma}$  is bounded away from 1 and  $\kappa(Q_{FF}^{-1}A_{FF})$  reasonably small. This does not mean that  $A_{FF}$  has to be well conditioned. In some cases, its structure is such that nice approximations may be found even when it is near singular. This occurs for instance with 2D anisotropic problems on regular grids with geometric refinement [31, 54].

However, a good *purely* algebraic method should work independently of the matrix structure, and should not rely on specialized approximations  $Q_{FF}$ . From that point of view, HBBD

requires that  $A_{FF}$  is well conditioned. As seen below, the other methods investigated in this paper share this requirement.

### 3.3. Analysis of HBBF and HBMG

Here we start from the observation that  $B_{\text{HBBF}}^{-1}A$  and  $B_{\text{HBMG}}^{-1}A$  (or, equivalently,  $\hat{B}_{\text{HBBF}}^{-1}\hat{A}$  and  $\hat{B}_{\text{HBMG}}^{-1}\hat{A}$ ) are similar to, respectively,  $\tilde{B}_{\text{HBBF}}^{-1}\tilde{A}$  and  $\tilde{B}_{\text{HBMG}}^{-1}\tilde{A}$ , where

$$\tilde{B}_{\text{HBBF}} = \begin{pmatrix} Q_{FF} & \\ & \hat{A}_C \end{pmatrix}$$

$$\tilde{B}_{\text{HBMG}} = \begin{pmatrix} (Q_{FF}^{-1} + Q_{FF}^{-T} - Q_{FF}^{-1}A_{FF}Q_{FF}^{-T})^{-1} & \\ & \hat{A}_C \end{pmatrix}$$

and

$$\tilde{A} = \begin{pmatrix} I & \\ -\hat{A}_{CF}Q_{FF}^{-T} & I \end{pmatrix} \begin{pmatrix} A_{FF} & \hat{A}_{FC} \\ \hat{A}_{CF} & \hat{A}_C \end{pmatrix} \begin{pmatrix} I & -Q_{FF}^{-1}\hat{A}_{FC} \\ & I \end{pmatrix} \quad (22)$$

with

$$\hat{A}_{FC} = A_{FC} + A_{FF}J_{FC}, \quad \hat{A}_{CF} = A_{CF} + J_{FC}^T A_{FF}$$

Observe here that (22) defines a transformation similar to (10), except that in (22) the (pseudo) interpolation matrix has a prescribed form, deduced from the matrix entries. Thus, both HBBF and HBMG are block-diagonal preconditioning after a further algebraic basis transformation. Hence, according to Theorem 2, some significant improvement over HBBF is possible if and only if this further transformation results in a smaller C.B.S. constant. This is investigated in Reference [4, Theorem 5.2], whose main result is extended to non-symmetric  $Q_{FF}$  in the following theorem.

#### Theorem 4

Let

$$\hat{A} = \begin{pmatrix} A_{FF} & \hat{A}_{FC} \\ \hat{A}_{CF} & \hat{A}_C \end{pmatrix}$$

be a symmetric positive definite matrix partitioned in  $2 \times 2$  block form, and let  $\hat{\gamma}$  be the associated C.B.S. constant. Let  $Q_{FF}$  be some non-singular approximation to  $A_{FF}$ , and let

$$\tilde{A} = \begin{pmatrix} I & \\ -\hat{A}_{CF}Q_{FF}^{-T} & I \end{pmatrix} \begin{pmatrix} A_{FF} & \hat{A}_{FC} \\ \hat{A}_{CF} & \hat{A}_C \end{pmatrix} \begin{pmatrix} I & -Q_{FF}^{-1}\hat{A}_{FC} \\ & I \end{pmatrix}$$

If

$$\eta = \|I - Q_{FF}^{-1}A_{FF}\|_{A_{FF}} \leq 1$$

then, the C.B.S. constant  $\tilde{\gamma}$  associated with  $\tilde{A}$  satisfies

$$\tilde{\gamma} \leq \frac{\eta\hat{\gamma}}{\sqrt{1 - \hat{\gamma}^2(1 - \eta^2)}} \tag{23}$$

*Proof*

By Lemma 1,

$$1 - \tilde{\gamma}^2 = \lambda_{\min}(\tilde{A}_{CC}^{-1}S_{\tilde{A}})$$

with

$$S_{\tilde{A}} = S_{\hat{A}} = \hat{A}_C - \hat{A}_{CF}A_{FF}^{-1}\hat{A}_{FC}$$

and

$$\tilde{A}_{CC} = \hat{A}_C + \hat{A}_{CF}(Q_{FF}^{-T}A_{FF}Q_{FF}^{-1} - Q_{FF} - Q_{FF}^T)\hat{A}_{FC}$$

Observe that

$$\tilde{A}_{CC} - S_{\tilde{A}} = \hat{A}_{CF}(I - Q_{FF}^{-T}A_{FF})(A_{FF}^{-1}(I - A_{FF}Q_{FF}^{-1})\hat{A}_{FC}$$

Hence, using Lemma 1 again, and noting that  $\eta = \|I - A_{FF}^{1/2}Q_{FF}^{-1}A_{FF}^{1/2}\|$ ,

$$\begin{aligned} 1 - \tilde{\gamma}^2 &= \min_{z_C \neq 0} \frac{z_C^T S_{\tilde{A}} z_C}{z_C^T \tilde{A}_{CC} z_C} \\ &= \min_{z_C \neq 0} \frac{1}{1 + \frac{z_C^T (\tilde{A}_{CC} - S_{\tilde{A}}) z_C}{z_C^T \hat{A}_{CF} A_{FF}^{-1} \hat{A}_{FC} z_C} \frac{z_C^T \hat{A}_{CF} A_{FF}^{-1} \hat{A}_{FC} z_C}{z_C^T (\hat{A}_C - \hat{A}_{CF} A_{FF}^{-1} \hat{A}_{FC}) z_C}} \\ &\geq \min_{z_C \neq 0} \frac{1}{1 + \frac{(A_{FF}^{-1/2} \hat{A}_{FC} z_C)^T (I - A_{FF}^{1/2} Q_{FF}^{-T} A_{FF}^{1/2}) (I - A_{FF}^{1/2} Q_{FF}^{-1} A_{FF}^{1/2}) (A_{FF}^{-1/2} \hat{A}_{FC} z_C)}{(A_{FF}^{-1/2} \hat{A}_{FC} z_C)^T (A_{FF}^{-1/2} \hat{A}_{FC} z_C)} \frac{1}{\hat{\gamma}^{-2} - 1}} \\ &\geq \frac{1}{1 + \eta^2(\hat{\gamma}^2/(1 - \hat{\gamma}^2))} \end{aligned}$$

The required result follows. □

Formally, this theorem applies to any matrix, not only to those that are in generalized hierarchical form. But we refer explicitly to  $\hat{A}$  to stress that this results is helpful only when the original C.B.S. constant is already bounded away from 1. This is perhaps better seen by rewriting (23) in the following equivalent form:

$$\frac{1}{1 - \tilde{\gamma}^2} - 1 \leq \eta^2 \left( \frac{1}{1 - \hat{\gamma}^2} - 1 \right)$$

On the other hand, this latter relation clearly displays that HBBF and HBMG may indeed improve significantly HBBD when  $Q_{FF}$  is a good approximation to  $A_{FF}$  (note that when  $Q_{FF}$  is symmetric,  $\eta$  is just the spectral radius of  $I - Q_{FF}^{-1}A_{FF}$ ).

Now, ‘ideal’ block-diagonal preconditioning of  $\tilde{A}$  would require to solve a system with  $\tilde{A}_{CC}$  (the bottom right block of  $\tilde{A}$ ). This would be impractical because  $\tilde{A}_{CC}$  is in general much denser than  $\hat{A}_C$ ; it is even a dense matrix as soon as  $Q_{FF}^{-1}$  is dense. Then, using  $\hat{A}_C$  to approximate  $\tilde{A}_{CC}$  is justified as follows. By (2) of Lemma 1,  $\hat{A}_C$  is spectrally equivalent to the Schur complement of  $\hat{A}$ , which is equal to the Schur complement of  $\tilde{A}$ , which itself is (by Lemma 1 again) spectrally equivalent to  $\tilde{A}_{CC}$ ; hence, by transitivity,  $\hat{A}_C$  is spectrally equivalent to  $\tilde{A}_{CC}$ . This reasoning, coupled with the application of Theorems 2 and 4, leads to the following theorems.

*Theorem 5 (analysis of HBBF)*

Let  $A$  be a symmetric positive definite matrix partitioned in  $2 \times 2$  block form, and let  $J_{FC}$  be some interpolation matrix. Let  $\hat{A}$  be the matrix resulting from the application of the generalized basis transformation defined by (9), (10), and let  $\hat{\gamma}$  be the C.B.S. constant associated with  $\hat{A}$ . Let  $\tilde{A}$  be the matrix defined by (22), and let  $\tilde{\gamma}$  be the associated C.B.S. constant. Let  $B_{\text{HBBF}}$  be the HBBF preconditioning matrix, as defined in Section 2.3, with symmetric positive definite approximation  $Q_{FF}$  to  $A_{FF}$  such that

$$\lambda_M = \lambda_{\max}(Q_{FF}^{-1}A_{FF}) < 2$$

Then:

$$\begin{aligned} \kappa(B_{\text{HBBF}}^{-1}A) &\leq \frac{1 + \tilde{\gamma}}{1 - \tilde{\gamma}} \frac{\lambda_M}{\min(1 - \hat{\gamma}^2, \lambda_m)} \\ &\leq \frac{1 + (\eta\hat{\gamma}/\sqrt{1 - \hat{\gamma}^2(1 - \eta^2)})}{1 - (\eta\hat{\gamma}/\sqrt{1 - \hat{\gamma}^2(1 - \eta^2)})} \frac{\lambda_M}{\min(1 - \hat{\gamma}^2, \lambda_m)} \end{aligned}$$

and

$$\kappa(B_{\text{HBBF}}^{-1}A) \geq \max\left(\frac{\lambda_M}{\lambda_m}, \frac{\lambda_M(1 - \tilde{\gamma}^2)}{1 - \hat{\gamma}^2}, \frac{1}{1 - \tilde{\gamma}}\right)$$

where

$$\lambda_m = \lambda_{\min}(Q_{FF}^{-1}A_{FF}), \quad \eta = \max(\lambda_M - 1, 1 - \lambda_m)$$

Further, if  $\lambda_M \leq 1$ ,

$$\begin{aligned} \kappa(B_{\text{HBBF}}^{-1}A) &\leq \frac{1}{1 - \tilde{\gamma}} \frac{1}{\min(1 - \hat{\gamma}^2, \lambda_m)} \\ &\leq \frac{1}{1 - (\eta\hat{\gamma}/\sqrt{1 - \hat{\gamma}^2(1 - \eta^2)})} \frac{1}{\min(1 - \hat{\gamma}^2, \lambda_m)} \end{aligned}$$

*Proof*

The first two inequalities to prove are straightforward corollaries of Theorems 2 and 4 providing that

$$1 - \hat{\gamma}^2 \leq \frac{\mathbf{z}_C^T \tilde{A}_{CC} \mathbf{z}_C}{\mathbf{z}_C^T \hat{A}_C \mathbf{z}_C} \leq 1 \quad \text{for all } \mathbf{z}_C \neq 0 \tag{24}$$

where  $\tilde{A}_{CC} = \hat{A}_C - \hat{A}_{CF}(2Q_{FF}^{-1} - Q_{FF}^{-1}A_{FF}Q_{FF}^{-1})\hat{A}_{FC}$ .

Now,  $\lambda_M < 2$  implies that  $2Q_{FF}^{-1} - Q_{FF}^{-1}A_{FF}Q_{FF}^{-1}$  is positive definite (see Lemma A.1), hence the upper bound in (24). On the other hand, for all  $\mathbf{z}_C \neq 0$ ,

$$\frac{\mathbf{z}_C^T \tilde{A}_{CC} \mathbf{z}_C}{\mathbf{z}_C^T \hat{A}_C \mathbf{z}_C} = \frac{\mathbf{z}_C^T \tilde{A}_{CC} \mathbf{z}_C}{\mathbf{z}_C^T S_{\hat{A}} \mathbf{z}_C} \frac{\mathbf{z}_C^T S_{\hat{A}} \mathbf{z}_C}{\mathbf{z}_C^T \hat{A}_C \mathbf{z}_C} \geq \frac{\mathbf{z}_C^T S_{\hat{A}} \mathbf{z}_C}{\mathbf{z}_C^T \hat{A}_C \mathbf{z}_C} \geq 1 - \hat{\gamma}^2 \tag{25}$$

which shows the lower bound in (24). Moreover,

$$\min_{\mathbf{z}_C \neq 0} \frac{\mathbf{z}_C^T \tilde{A}_{CC} \mathbf{z}_C}{\mathbf{z}_C^T \hat{A}_C \mathbf{z}_C} \leq \left( \max_{\mathbf{z}_C \neq 0} \frac{\mathbf{z}_C^T \tilde{A}_{CC} \mathbf{z}_C}{\mathbf{z}_C^T S_{\hat{A}} \mathbf{z}_C} \right) \left( \min_{\mathbf{z}_C \neq 0} \frac{\mathbf{z}_C^T S_{\hat{A}} \mathbf{z}_C}{\mathbf{z}_C^T \hat{A}_C \mathbf{z}_C} \right) = \frac{1 - \hat{\gamma}^2}{1 - \hat{\gamma}^2} \tag{26}$$

hence the lower bound on  $\kappa(B_{\text{HBBF}}^{-1}A)$ , using Theorem 2 again.

The last two inequalities to prove follow along the same line as the two first ones, using additional fact that

$$\hat{B}_{\text{HBBF}} - \hat{A} = \begin{pmatrix} Q_{FF} - A_{FF} & 0 \\ 0 & \hat{A}_{CF}Q_{FF}^{-1}\hat{A}_{FC} \end{pmatrix}$$

is positive semidefinite when  $\lambda_M \leq 1$ , hence  $\lambda_{\max}(B_{\text{HBBF}}^{-1}A) \leq 1$ , which supersedes the upper bound from Theorem 2. □

*Theorem 6 (analysis of HBMG)*

Let  $A$  be a symmetric positive definite matrix partitioned in  $2 \times 2$  block form, and let  $J_{FC}$  be some interpolation matrix. Let  $\hat{A}$  be the matrix resulting from the application of the generalized basis transformation defined by (9), (10), and let  $\hat{\gamma}$  be the C.B.S. constant associated with  $\hat{A}$ . Let  $\tilde{A}$  be the matrix defined by (22), and let  $\tilde{\gamma}$  be the associated C.B.S. constant. Let  $B_{\text{HBMG}}$  be the HBMG preconditioning matrix, as defined in Section 2.4, with non-singular approximation  $Q_{FF}$  to  $A_{FF}$  such that

$$\eta = \|I - Q_{FF}^{-1}A_{FF}\|_{A_{FF}} < 1$$

Then:

$$\begin{aligned} \kappa(B_{\text{HBMG}}^{-1}A) &\leq \frac{1}{1 - \tilde{\gamma}} \frac{1}{\min(1 - \hat{\gamma}^2, 1 - \eta^2)} \\ &\leq \frac{1}{1 - (\eta\tilde{\gamma}/\sqrt{1 - \hat{\gamma}^2(1 - \eta^2)})} \frac{1}{\min(1 - \hat{\gamma}^2, 1 - \eta^2)} \end{aligned}$$

and

$$\kappa(B_{\text{HBMG}}^{-1}A) \geq \max \left( \frac{1 - \delta^2}{1 - \eta^2}, \frac{(1 - \delta^2)(1 - \hat{\gamma}^2)}{1 - \hat{\gamma}^2}, \frac{1}{1 - \hat{\gamma}} \right)$$

where  $\delta$  is the smallest singular value of  $I - A_{FF}^{1/2}Q_{FF}^{-1}A_{FF}^{1/2}$ .

*Proof*

Firstly,  $I - A^{1/2}p\hat{A}_C^{-1}p^T A^{1/2}$  is an orthogonal projector [12, p. 431], and therefore positive semidefinite. Hence, from (15), one deduces

$$\mathbf{z}^T B_{\text{HBMG}}^{-1} \mathbf{z} \leq \mathbf{z}^T A^{-1} \mathbf{z} \quad \text{for all } \mathbf{z}$$

which shows  $\lambda_{\max}(B_{\text{HBMG}}^{-1}A) \leq 1$ . Further, (25) and (26) are also valid here. The theorem is then a straightforward corollary of Theorems 2 and 4 if

$$\min_{\mathbf{z}_F \neq 0} \frac{\mathbf{z}_F^T (Q_{FF}^{-1} + Q_{FF}^{-T} - Q_{FF}^{-1}A_{FF}Q_{FF}^{-T}) \mathbf{z}_F}{\mathbf{z}_F^T A_{FF}^{-1} \mathbf{z}_F} = 1 - \eta^2$$

and

$$\max_{\mathbf{z}_F \neq 0} \frac{\mathbf{z}_F^T (Q_{FF}^{-1} + Q_{FF}^{-T} - Q_{FF}^{-1}A_{FF}Q_{FF}^{-T}) \hat{\mathbf{z}}_F}{\mathbf{z}_F^T A_{FF}^{-1} \mathbf{z}_F} = 1 - \delta^2$$

where both can be checked from

$$Q_{FF}^{-1} + Q_{FF}^{-T} - Q_{FF}^{-1}A_{FF}Q_{FF}^{-T} = A_{FF}^{-1} - A_{FF}^{-1/2}(I - A_{FF}^{1/2}Q_{FF}^{-1}A_{FF}^{1/2})(I - A_{FF}^{1/2}Q_{FF}^{-T}A_{FF}^{1/2})A_{FF}^{-1/2}$$

and from the fact that  $\eta^2$  and  $\delta^2$  are, respectively, the largest and the smallest eigenvalue of  $(I - A_{FF}^{1/2}Q_{FF}^{-1}A_{FF}^{1/2})(I - A_{FF}^{1/2}Q_{FF}^{-T}A_{FF}^{1/2})$ .  $\square$

As HBBD, these approaches can thus be efficient if and only if  $\hat{\gamma}$  is bounded away from 1 and  $\eta = \|I - Q_{FF}^{-1}A_{FF}\|_{A_{FF}}$  relatively small. Here again, the latter condition should be seen as a constraint on the  $F/C$  partitioning, to be chosen such that  $A_{FF}$  is reasonably well conditioned. Then, finding approximations for which  $\eta$  is fairly small is not that difficult (see, e.g. Reference [6]). In this case, the main term in the upper bound on the condition number is  $1/(1 - \hat{\gamma}^2)$  for both HBMG and HBBF, and one should not expect much difference between them.

An alternative, more direct, analysis of HBBF preconditioning is developed in References [3, 5, 6]. This analysis assumes  $Q_{FF}$  such that  $\lambda_M \leq 1$ , and leads to a bound whose exact expression is hard to interpret, but which may fortunately be simplified into

$$\kappa(B_{\text{HBBF}}^{-1}A) \leq \frac{1}{\lambda_m} \frac{1}{1 - \hat{\gamma}^2}$$

This is better than the bound of Theorem 5 if  $\hat{\gamma}^2 \leq 1/(2 - \eta^2)$ , but leads anyway to the same general conclusion: when one uses a close enough approximation to  $A_{FF}$ , the condition number is close to  $1/(1 - \hat{\gamma}^2)$ , which is the ‘ideal’ condition number corresponding to exact inversion of  $A_{FF}$  (observe that upper and lower bounds coincide when  $Q_{FF} = A_{FF}$ ).

Our analysis brings an additional light, being more general and including virtually any block-diagonal preconditioning of the transformed matrix  $\tilde{A}$ . In particular, it shows that there is no real need to scale  $Q_{FF}$  so as to satisfy  $\lambda_M \leq 1$ , as suggested in Reference [6].

3.4. Analysis of MBF

This preconditioner is also a block-diagonal one, with respect to the matrix

$$\tilde{A} = \begin{pmatrix} I & \\ -A_{CF}P_{FF}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{FF} & A_{FC} \\ A_{CF} & A_{CC} \end{pmatrix} \begin{pmatrix} I & -P_{FF}^{-1}A_{FC} \\ & I \end{pmatrix} \tag{27}$$

This transformation is similar to the one analysed in Theorem 4, but applied directly to a matrix for which the C.B.S. constant is likely to be poor. A further analysis is therefore needed. The following theorem is restricted to approximations  $P_{FF}$  satisfying some strong assumptions, but allows to bound the C.B.S. constant of the transformed matrix independently of the C.B.S. constant of the original matrix.

Theorem 7

Let  $A$  be a symmetric positive definite matrix partitioned in  $2 \times 2$  block form. Let  $P_{FF}$  be some symmetric positive definite approximation to  $A_{FF}$  such that

$$\lambda_{\min}(P_{FF}^{-1}A_{FF}) \geq 1 \tag{28}$$

and

$$\begin{pmatrix} P_{FF} & A_{FC} \\ A_{CF} & A_{CC} \end{pmatrix} \text{ is positive semidefinite} \tag{29}$$

Then, the C.B.S. constant  $\tilde{\gamma}$  associated with

$$\tilde{A} = \begin{pmatrix} I & \\ -A_{CF}P_{FF}^{-1} & I \end{pmatrix} A \begin{pmatrix} I & -P_{FF}^{-1}A_{FC} \\ & I \end{pmatrix}$$

satisfies

$$\tilde{\gamma} \leq \sqrt{1 - \frac{1}{\lambda_{\max}(P_{FF}^{-1}A_{FF})}}$$

Proof

First, observe that the positive semidefiniteness of the matrix in (29) implies the positive semidefiniteness of its Schur complement, hence

$$\mathbf{z}_C^T A_{CC} \mathbf{z}_C \geq \mathbf{z}_C^T A_{CF} P_{FF}^{-1} A_{FC} \mathbf{z}_C \quad \text{for all } \mathbf{z}_C$$

Let then  $\alpha = 1/\lambda_{\max}(P_{FF}^{-1}A_{FF})$ . One has, for all  $\mathbf{z}_C$ ,

$$\begin{aligned} \mathbf{z}_C^T (S_A - \alpha \tilde{A}_{CC}) \mathbf{z}_C &= (1 - \alpha)(\mathbf{z}_C^T A_{CC} \mathbf{z}_C) + \mathbf{z}_C^T A_{CF} (2\alpha P_{FF}^{-1} - \alpha P_{FF}^{-1} A_{FF} P_{FF}^{-1} - A_{FF}^{-1}) A_{FC} \mathbf{z}_C \\ &\geq \mathbf{z}_C^T A_{CF} ((1 + \alpha) P_{FF}^{-1} - \alpha P_{FF}^{-1} A_{FF} P_{FF}^{-1} - A_{FF}^{-1}) A_{FC} \mathbf{z}_C \\ &= (P_{FF}^{-1/2} A_{FC} \mathbf{z}_C)^T (I - (P_{FF}^{-1/2} A_{FF} P_{FF}^{-1/2})^{-1}) (I - \alpha (P_{FF}^{-1/2} A_{FF} P_{FF}^{-1/2})) (P_{FF}^{-1/2} A_{FC} \mathbf{z}_C) \\ &\geq 0 \end{aligned}$$

The required result follows then from Lemma 1. □

Requirements (28), (29) are somewhat conflictual because  $P_{FF}$  has to be smaller (or not larger), in the positive definite sense, than  $A_{FF}$ , but at the same time sufficiently large (in the same sense) so that exchanging  $A_{FF}$  for  $P_{FF}$  in  $A$  does not spoil its positive semidefiniteness. The following lemma is useful to see how this can be managed when  $A$  is a (non-strictly) diagonally dominant matrix.

*Lemma 8*

Let  $A$  be a (non-strictly) diagonally dominant symmetric matrix with positive diagonal entries, partitioned in  $2 \times 2$  block form. Let  $P_{FF}$  be some symmetric positive definite approximation to  $A_{FF}$  such that

$$P_{FF}^{-1} \geq 0 \quad (30)$$

(entrywise) and,  $\forall i \in F$  such that  $(A_{FC})_{ij} \neq 0$  for some  $j \in C$ ,

$$(P_{FF}^{-1} A_{FF} \mathbf{e}_F)_i \leq \frac{(A_{FF} \mathbf{e}_F)_i}{\sum_{j \in C} |(A_{FC})_{ij}|} \quad (31)$$

where  $\mathbf{e}_F = (1 \dots 1)^T$ . Then:

$$\begin{pmatrix} P_{FF} & A_{FC} \\ A_{CF} & A_{CC} \end{pmatrix} \text{ is positive semidefinite} \quad (32)$$

*Proof*

Let  $F_C = \{i \in F \mid (A_{FC})_{ij} \neq 0 \text{ for some } j \in C\}$  be the set of  $F$  unknowns connected to at least one  $C$  unknown. Let  $K_{FF}$  be such that

$$(K_{FF})_{ij} = \begin{cases} (P_{FF}^{-1})_{ij} & \text{if } i, j \in F_C \\ 0 & \text{otherwise} \end{cases}$$

and let  $G_{FF}$  be the diagonal matrix such that

$$(G_{FF})_{ii} = \begin{cases} \sum_{j \in C} |(A_{FC})_{ij}| & \text{if } i \in F_C \\ 1 & \text{otherwise} \end{cases}$$

Let  $\mathbf{x}_F = A_{FF} \mathbf{e}_F$ , and let  $\bar{\mathbf{x}}_F$  be such that

$$(\bar{\mathbf{x}}_F)_i = \begin{cases} (\mathbf{x}_F)_i & \text{if } i \in F_C \\ 1 & \text{otherwise} \end{cases}$$

Since  $P_{FF}^{-1}$  is non-negative, one has, with (31),

$$((G_{FF}^{-1} - K_{FF})\bar{\mathbf{x}}_F)_i \geq ((G_{FF}^{-1} - P_{FF}^{-1})\mathbf{x}_F)_i \geq 0$$

for all  $i \in F_C$ . Moreover,  $A$  being (non-strictly) diagonally dominant (with positive diagonal entries),  $(\mathbf{x}_F)_i > 0$  for all  $i \in F_C$ , implying that  $\bar{\mathbf{x}}_F$  is a positive vector. Since  $G_{FF}^{-1} - K_{FF}$  has non-positive offdiagonal entries, this, together with  $(G_{FF}^{-1} - K_{FF})\bar{\mathbf{x}}_F \geq 0$ , implies that  $G_{FF}^{-1} - K_{FF}$  is positive semidefinite because it is generalized diagonally dominant with non-negative diagonal entries.

On the other hand, the matrix

$$\begin{pmatrix} G_{FF} & A_{FC} \\ A_{CF} & A_{CC} \end{pmatrix}$$

is symmetric and non-strictly diagonally dominant with positive diagonal entries. It is therefore positive semidefinite, as well as its Schur complement  $A_{CC} - A_{CF}G_{FF}^{-1}A_{FC}$ . Hence, for all  $\mathbf{z}_C$ ,

$$\mathbf{z}_C^T A_{CC} \mathbf{z}_C \geq \mathbf{z}_C^T A_{CF} G_{FF}^{-1} A_{FC} \mathbf{z}_C \geq \mathbf{z}_C^T A_{CF} K_{FF} A_{FC} \mathbf{z}_C = \mathbf{z}_C^T A_{CF} P_{FF}^{-1} A_{FC} \mathbf{z}_C$$

This shows that the Schur complement of the matrix in (32) is positive semidefinite, hence this matrix has to be positive semidefinite too, since its top left block is positive definite. □

Now, observe first that the right-hand side of (31) cannot be smaller than 1, on account of the diagonal dominance assumption. Hence, (31) is always satisfied by approximations  $P_{FF}$  satisfying the row-sum criterion

$$P_{FF} \mathbf{e}_F = A_{FF} \mathbf{e}_F \tag{33}$$

or, equivalently,

$$P_{FF}^{-1} A_{FF} \mathbf{e}_F = \mathbf{e}_F \tag{34}$$

If  $A_{FF}$  has in addition non-positive offdiagonal entries (i.e. is an  $M$ -matrix), then it is easy to see that all requirements (28), (30), (31), are satisfied with a (possibly blockwise) modified ILU factorization of  $A_{FF}$  [3, 67–69]. Indeed, such factorizations are computed so as to satisfy the row-sum criterion (33), whereas (28), (30) follow from the standard analysis of these methods (see, e.g. Reference [70] for a proof that includes the case of a blockwise factorization).

On the other hand, if  $A_{FF}$  is monotone and  $A_{FF} \mathbf{e}_F > 0$ , another possibility is to start from any weak regular splitting of  $A_{FF}$ , and add a diagonal correction so as to satisfy (34). Indeed, a weak regular splitting is a splitting  $A_{FF} = C_{FF} - N_{FF}$  such that

$$\begin{aligned} C_{FF}^{-1} &\geq 0 \\ C_{FF}^{-1} N_{FF} &\geq 0 \end{aligned}$$

and it turns out that (28), (30) are then satisfied with  $P_{FF}$  defined by

$$P_{FF}^{-1} = C_{FF}^{-1} + \Delta_{FF}$$

where  $\Delta_{FF}$  is the diagonal matrix such that (34) holds, i.e.

$$\Delta_{FF} (A_{FF} \mathbf{e}_F) = N_{FF} \mathbf{e}_F$$

To see this, observe that, since weak regular splittings are convergent (e.g. Reference [3, Corollary 6.17]),

$$A_{FF}^{-1} = C_{FF}^{-1} + C_{FF}^{-1} N_{FF} C_{FF}^{-1} + \dots \geq C_{FF}^{-1}$$

Since  $(P_{FF}^{-1} - A_{FF}^{-1})(A_{FF}\mathbf{e}_F) = 0$  and  $A_{FF}\mathbf{e}_F > 0$ , this shows that  $P_{FF}^{-1} - A_{FF}^{-1}$  is generalized diagonally dominant with non-negative diagonal entries. It is therefore positive semidefinite, which shows both (28) and  $\text{diag}(P_{FF}^{-1}) \geq 0$ , implying (30) since the offdiagonal entries of  $P_{FF}^{-1}$  are non-negative by construction.

Here, it is also worth mentioning that, starting from any basic approximation satisfying requirements (28), (29), one may improve it by some inner iterations based on Chebyshev polynomials, and still satisfy these requirements; see Reference [10] for details.

Beyond these algebraic developments, we would like to stress more informally the role of the row-sum criterion (33) or (34). In our eyes, it essentially ensures that  $-P_{FF}^{-1}A_{FC}$  acts as a correct interpolation matrix, being close to exact for the constant vector, which is the only (potential) low-energy mode for (non-strictly) diagonally dominant matrices. For more general matrices, even if there is no supporting theory, a very first thing to do is therefore to satisfy a similar criterion with respect to all low energy modes.

Note that in Reference [10], numerical results are reported for two-level MBF preconditioning, that illustrate the behaviour of the method when one does not take care of this condition, using standard ILU instead of modified ILU to approximate  $A_{FF}$ . It turns out that the resulting condition number grows faster than  $h^{-2}$ , even though the ILU factorization of  $A_{FF}$  is on its own of excellent quality, with condition number approximately equal to 1.25.

We now exploit Theorem 7 to develop our analysis of MBF preconditioning.

*Theorem 9 (analysis of MBF)*

Let  $A$  be a symmetric positive definite matrix partitioned in  $2 \times 2$  block form. Let  $B_{\text{MBF}}$  be the MBF preconditioning matrix, as defined in Section 2.5 (with symmetric positive definite approximations  $P_{FF}$  to  $A_{FF}$  and  $S$  to  $S_A = A_{CC} - A_{CF}A_{FF}^{-1}A_{FC}$ , respectively). Assume that

$$\lambda_m = \lambda_{\min}(P_{FF}^{-1}A_{FF}) \geq 1$$

and

$$\begin{pmatrix} P_{FF} & A_{FC} \\ A_{CF} & A_{CC} \end{pmatrix} \text{ is positive semidefinite}$$

Let  $\bar{A}$  be the matrix defined by (27), and let  $\bar{\gamma}$  be the associated C.B.S. constant. Then:

$$\begin{aligned} \kappa(B_{\text{MBF}}^{-1}A) &\leq \frac{1 + \bar{\gamma}}{1 - \bar{\gamma}} \frac{\max(v_M/(1 - \bar{\gamma}^2), \lambda_M)}{\min(v_m, \lambda_m)} \\ &\leq \left(1 + \sqrt{1 - \lambda_M^{-1}}\right)^2 \frac{\lambda_M^2 v_M}{\min(v_m, \lambda_m)} \end{aligned}$$

and

$$\kappa(B_{\text{MBF}}^{-1}A) \geq \max\left(\frac{\max(\lambda_M, v_M)}{\min(\lambda_m, v_m/(1 - \bar{\gamma}^2))}, \frac{1}{1 - \bar{\gamma}}\right)$$

where

$$v_M = \lambda_{\max}(S^{-1}S_A), \quad v_m = \lambda_{\min}(S^{-1}S_A)$$

and

$$\lambda_M = \lambda_{\max}(P_{FF}^{-1}A_{FF})$$

Further, if  $S = \hat{A}_C$ , where  $\hat{A}_C$  is the Galerkin matrix (4) associated with some interpolation matrix  $J_{FC}$ , then,

$$\begin{aligned} \kappa(B_{MBF}^{-1}A) &\leq \frac{1 + \bar{\gamma}}{1 - \bar{\gamma}} \frac{\max((1 - \bar{\gamma}^2)^{-1}, \lambda_M)}{1 - \hat{\gamma}^2} \\ &\leq \left(1 + \sqrt{1 - \lambda_M^{-1}}\right)^2 \frac{\lambda_M^2}{1 - \hat{\gamma}^2} \end{aligned}$$

and

$$\kappa(B_{MBF}^{-1}A) \geq \max\left(\frac{\lambda_M}{\lambda_m}, \frac{\lambda_M(1 - \bar{\gamma}^2)}{1 - \hat{\gamma}^2}, \frac{1}{1 - \bar{\gamma}}\right)$$

where  $\hat{\gamma}$  is the C.B.S. constant associated with the matrix  $\hat{A}$  resulting from the application of the generalized basis transformation defined by (9), (10).

*Proof*

The first inequality to prove is a straightforward corollary of Theorem 2 if

$$v_m \leq \frac{\mathbf{z}_C^T \bar{A}_{CC} \mathbf{z}_C}{\mathbf{z}_C^T S \mathbf{z}_C} \leq \frac{v_M}{1 - \bar{\gamma}^2} \quad \text{for all } \mathbf{z}_C \neq 0$$

which one proves with Lemma 1 and

$$\frac{\mathbf{z}_C^T \bar{A}_{CC} \mathbf{z}_C}{\mathbf{z}_C^T S \mathbf{z}_C} = \frac{\mathbf{z}_C^T \bar{A}_{CC} \mathbf{z}_C}{\mathbf{z}_C^T S_{\bar{A}} \mathbf{z}_C} \frac{\mathbf{z}_C^T S_{\bar{A}} \mathbf{z}_C}{\mathbf{z}_C^T S \mathbf{z}_C} \quad \text{for all } \mathbf{z}_C \neq 0$$

The second inequality then follows from Theorem 7, which implies  $(1 - \bar{\gamma}^2)^{-1} \leq \lambda_M$  and  $(1 + \bar{\gamma})/(1 - \bar{\gamma}) \leq \lambda_M(1 + \bar{\gamma})^2 \leq \lambda_M(1 + \sqrt{1 - \lambda_M^{-1}})^2$ . With the help of Lemma 1 again, the further two upper bounds are particular cases of the previous ones, for  $v_M \leq 1$  and  $v_m = 1 - \hat{\gamma}^2$ .

Since

$$\max_{\mathbf{z}_C \neq 0} \frac{\mathbf{z}_C^T \bar{A}_{CC} \mathbf{z}_C}{\mathbf{z}_C^T S \mathbf{z}_C} \geq \left( \min_{\mathbf{z}_C \neq 0} \frac{\mathbf{z}_C^T \bar{A}_{CC} \mathbf{z}_C}{\mathbf{z}_C^T S_{\bar{A}} \mathbf{z}_C} \right) \left( \max_{\mathbf{z}_C \neq 0} \frac{\mathbf{z}_C^T S_{\bar{A}} \mathbf{z}_C}{\mathbf{z}_C^T S \mathbf{z}_C} \right) = v_M$$

and

$$\min_{\mathbf{z}_C \neq 0} \frac{\mathbf{z}_C^T \bar{A}_{CC} \mathbf{z}_C}{\mathbf{z}_C^T S \mathbf{z}_C} \geq \left( \max_{\mathbf{z}_C \neq 0} \frac{\mathbf{z}_C^T \bar{A}_{CC} \mathbf{z}_C}{\mathbf{z}_C^T S_{\bar{A}} \mathbf{z}_C} \right) \left( \min_{\mathbf{z}_C \neq 0} \frac{\mathbf{z}_C^T S_{\bar{A}} \mathbf{z}_C}{\mathbf{z}_C^T S \mathbf{z}_C} \right) = \frac{v_m}{1 - \bar{\gamma}^2}$$

the lower bounds are also corollaries of Theorem 2. □

Here again, we have to mention that slightly better bounds are obtained with the more direct analysis developed in Reference [10]. Their expression, however, is not that easy to

interpret, so we do not display them. The present approach has also several other merits: it applies more widely to any block-diagonal preconditioning of the transformed matrix  $\bar{A}$ , and, most important, shows that this type of preconditioning can be efficient if *and only if*  $\bar{\gamma}$  is bounded away from 1, that is, if *and only if*  $P_{FF}$ , besides being a close approximation to  $A_{FF}$ , is also such that  $-P_{FF}^{-1}A_{FC}$  acts as a ‘correct’ interpolation.

### 3.5. Algebraic interpolation and the C.B.S. constant

Methods based on generalized hierarchical bases (HBBD, HBBF and HBMG) can be efficient only if  $\hat{\gamma}$  is bounded away from 1. This raises the question of the construction of the interpolation matrix  $J_{FC}$ , which we consider in this subsection.

Theorem 7 and Lemma 8 open the way to interpolations of the form  $-K_{FF}A_{FC}$ , where  $K_{FF}$  stands for some approximate inverse of  $A_{FF}$ . However, it is essential here to keep  $K_{FF}$  as sparse as possible, to avoid excessive complexity in  $\hat{A}_C$ . The sparsest  $K_{FF}$  is diagonal. If  $A$  is (non-strictly) diagonally dominant with positive diagonal entries and  $A_{FF}$  strictly diagonally dominant, diagonal  $K_{FF}$  such that

$$\frac{1}{(A_{FF})_{ii} - \sum_{j \in F \setminus \{i\}} |(A_{FF})_{ij}|} \leq (K_{FF})_{ii} \leq \frac{1}{\sum_{j \in C} |(A_{FC})_{ij}|} \quad \text{for all } i \in F \quad (35)$$

satisfies all assumption of Lemma 8 and Theorem 7,<sup>‡</sup> which yield

$$\hat{\gamma} \leq \sqrt{1 - \frac{1}{\lambda_{\max}(K_{FF}A_{FF})}}$$

Here, we would like to point out that, if  $A$  is a (non-strictly) diagonally dominant  $M$ -matrix and if all coupling in  $A_{FC}$  are classified ‘strong’, then the standard interpolation in AMG is defined with

$$(J_{FC})_{ij} = \frac{-(\sum_{j \neq i} |(A)_{ij}|)(A_{FC})_{ij}}{(A_{FF})_{ii}(\sum_{j \in C} |(A_{FC})_{ij}|)} \quad \text{for all } i \in F, j \in C$$

[12, p. 448]. Interesting enough, this interpolation is thus of the form  $-K_{FF}A_{FC}$  with diagonal  $K_{FF}$ ; moreover, the corresponding  $K_{FF}$  also satisfies (35), the upper bound following from the diagonal dominance assumption, whereas, letting  $\delta_i = (A_{FF})_{ii} - \sum_{j \neq i} |(A)_{ij}|$ , one has

$$\begin{aligned} \frac{\sum_{j \neq i} |(A)_{ij}|}{(A_{FF})_{ii}(\sum_{j \in C} |(A_{FC})_{ij}|)} &= \frac{(A_{FF})_{ii} - \delta_i}{(A_{FF})_{ii}((A_{FF})_{ii} - \sum_{j \in F \setminus \{i\}} |(A_{FF})_{ij}| - \delta_i)} \\ &= \frac{1}{(A_{FF})_{ii} - \sum_{j \in F \setminus \{i\}} |(A_{FF})_{ij}|} \frac{1 - (\delta_i/(A_{FF})_{ii})}{1 - (\delta_i/(A_{FF})_{ii} - \sum_{j \in F \setminus \{i\}} |(A_{FF})_{ij}|)} \end{aligned}$$

from which one sees that the lower bound in (35) is also satisfied,  $\delta_i$  being non-negative.

However, in general, AMG does not use interpolations of the form  $-K_{FF}A_{FC}$ , even for  $M$ -matrices. Indeed, couplings are classified in ‘strong’ and ‘weak’, according their magnitude,

<sup>‡</sup>The lower bound in (35) implies that  $A_{FF} - K_{FF}^{-1}$  is (non-strictly) diagonally dominant, hence (28).

and only strong couplings in  $A_{FC}$  lead to a non-zero term in  $J_{FC}$ . As this yields sparser  $J_{FC}$ , this raises the question whether such interpolations are also appropriate for the definition of generalized hierarchical basis transformations. A detailed answer to this question would perhaps require a thorough analysis of all interpolation schemes that have been proposed with AMG. Here we consider another viewpoint, based on the measure

$$\tau = \max_{\mathbf{z} \neq 0} \frac{(\mathbf{z}_F - J_{FC}\mathbf{z}_C)^T \text{diag}(A_{FF})(\mathbf{z}_F - J_{FC}\mathbf{z}_C)}{\mathbf{z}^T A \mathbf{z}} \tag{36}$$

that is often used to assess the quality of an interpolation scheme in AMG, the smaller  $\tau$ , the better the interpolation [12]. The following theorem just shows that  $\tau$  is in fact strongly connected to  $\hat{\gamma}$ . More precisely, it shows that  $\tau$  is reasonably bounded if and only if  $\hat{\gamma}$  is bounded away from 1 and  $A_{FF}$  is well conditioned. Note that the theorem is not restricted to  $D_{FF} = \text{diag}(A_{FF})$ , which is the intended application here. This wider scope is used in the next subsection.

*Theorem 10*

Let  $A$  be a symmetric positive definite matrix partitioned in  $2 \times 2$  block form, and let  $J_{FC}$  be some interpolation matrix. Let  $\hat{A}$  be the matrix resulting from the application of the generalized basis transformation defined by (9), (10), and let  $\hat{\gamma}$  be the C.B.S. constant associated with  $\hat{A}$ .

Let

$$\tau = \max_{\mathbf{z} \neq 0} \frac{(\mathbf{z}_F - J_{FC}\mathbf{z}_C)^T D_{FF}(\mathbf{z}_F - J_{FC}\mathbf{z}_C)}{\mathbf{z}^T A \mathbf{z}}$$

where  $D_{FF}$  is a symmetric positive definite matrix of same size as  $A_{FF}$ . Then:

$$\tau \leq \frac{1}{\lambda_{\min}(D_{FF}^{-1}A_{FF})} \frac{1}{1 - \hat{\gamma}^2}$$

and

$$\tau \geq \max \left( \frac{1}{\lambda_{\max}(D_{FF}^{-1}A_{FF})} \frac{1}{1 - \hat{\gamma}^2}, \frac{1}{\lambda_{\min}(D_{FF}^{-1}A_{FF})} \right)$$

*Proof*

Let  $J$  and  $q$  be defined by (9) and (12), respectively. Since  $\mathbf{z}_F - J_{FC}\mathbf{z}_C = q^T J^{-1} \mathbf{z}$  there holds (with  $\mathbf{w} = J^{-1} \mathbf{z}$ ),

$$\tau = \max_{\mathbf{w} \neq 0} \frac{\mathbf{w}_F^T D_{FF} \mathbf{w}_F}{\mathbf{w}^T \hat{A} \mathbf{w}}$$

On the other hand, from the factorization

$$\begin{pmatrix} A_{FF} & \hat{A}_{FC} \\ \hat{A}_{CF} & \hat{A}_C \end{pmatrix} = \begin{pmatrix} I & \hat{A}_{FC} \hat{A}_C^{-1} \\ & I \end{pmatrix} \begin{pmatrix} S_A^{(F)} & \\ & \hat{A}_C \end{pmatrix} \begin{pmatrix} I & \\ \hat{A}_C^{-1} \hat{A}_{CF} & I \end{pmatrix}$$

(with  $S_A^{(F)} = A_{FF} - \hat{A}_{FC} \hat{A}_C^{-1} \hat{A}_{CF}$ ), one deduces

$$\mathbf{w}^T \hat{A} \mathbf{w} = \mathbf{w}_F^T S_A^{(F)} \mathbf{w}_F + (\mathbf{w}_C + \hat{A}_C^{-1} \hat{A}_{CF} \mathbf{w}_F)^T \hat{A}_C (\mathbf{w}_C + \hat{A}_C^{-1} \hat{A}_{CF} \mathbf{w}_F)$$

and therefore

$$\min_{\mathbf{w}_C}(\mathbf{w}^T \hat{A} \mathbf{w}) = \mathbf{w}_F^T S_{\hat{A}}^{(F)} \mathbf{w}_F \quad (37)$$

Hence,

$$\tau = \frac{1}{\lambda_{\min}(D_{FF}^{-1} S_{\hat{A}}^{(F)})} \quad (38)$$

and the required result follows from Lemma 1, using

$$\lambda_{\min}(D_{FF}^{-1} S_{\hat{A}}^{(F)}) \geq \lambda_{\min}(A_{FF}^{-1} S_{\hat{A}}^{(F)}) \lambda_{\min}(D_{FF}^{-1} A_{FF})$$

to prove the upper bound and

$$\begin{aligned} \lambda_{\min}(D_{FF}^{-1} S_{\hat{A}}^{(F)}) &\leq \lambda_{\max}(A_{FF}^{-1} S_{\hat{A}}^{(F)}) \lambda_{\min}(D_{FF}^{-1} A_{FF}) \\ \lambda_{\min}(D_{FF}^{-1} S_{\hat{A}}^{(F)}) &\leq \lambda_{\min}(A_{FF}^{-1} S_{\hat{A}}^{(F)}) \lambda_{\max}(D_{FF}^{-1} A_{FF}) \end{aligned}$$

to prove the lower bound. □

### 3.6. Analysis of AMG

Firstly, note that  $I - A^{1/2} p \hat{A}_C^{-1} p^T A^{1/2}$  is an orthogonal projector [12, p. 431]. It is therefore positive semidefinite and singular, implying that  $B_{\text{AMG}}$  defined from (8) satisfies

$$\lambda_{\max}(B_{\text{AMG}}^{-1} A) = 1 \quad (39)$$

Hence,

$$\begin{aligned} \kappa(B_{\text{AMG}}^{-1} A) &= (\lambda_{\min}(B_{\text{AMG}}^{-1} A))^{-1} \\ &= (1 - \|(I - M^{-1} A)(I - p \hat{A}_C^{-1} p^T A)(I - M^{-T} A)\|_A)^{-1} \end{aligned} \quad (40)$$

that is, we are left with the analysis of the  $A$ -norm of the iteration matrix.

Since

$$(I - p \hat{A}_C^{-1} p^T A)^2 = (I - p \hat{A}_C^{-1} p^T A)$$

one has, letting  $C = (I - p \hat{A}_C^{-1} p^T A)(I - M^{-T} A)$ ,

$$\begin{aligned} \|(I - M^{-1} A)(I - p \hat{A}_C^{-1} p^T A)(I - M^{-T} A)\|_A &= \|A^{-1} C^T A C\|_A \\ &= \|(A^{1/2} C A^{-1/2})^T (A^{1/2} C A^{-1/2})\| \\ &= \|A^{1/2} C A^{-1/2}\|_2^2 \\ &= \|(I - p \hat{A}_C^{-1} p^T A)(I - M^{-T} A)\|_A^2 \end{aligned} \quad (41)$$

Thus the  $A$ -norm of the iteration matrix is equal to the square of that of a simplified AMG scheme with post-smoothing only.<sup>§</sup>

Now, the classical analysis of  $\|(I - M^{-1}A)(I - p\hat{A}_C^{-1}p^T A)\|_A$ , as developed in References [7, 11, 12], is based on the measure  $\tau$  introduced in the previous subsection (see (36)). Here we follow the more general approach developed in Reference [9]. The latter analysis also introduces some measure, which depends upon a restriction operator, for which one has some freedom. We restrict our attention to the choice  $r=(0 I)$  which is the most natural in the context of this paper, referring to Reference [9] for a wider discussion. With this choice, the considered measure is

$$\mu = \max_{\mathbf{z} \neq 0} \frac{(\mathbf{z}_F - J_{FC}\mathbf{z}_C)^T X_{FF}(\mathbf{z}_F - J_{FC}\mathbf{z}_C)}{\mathbf{z}^T A \mathbf{z}} \tag{42}$$

where  $X_{FF}$  is the top left block of

$$X = M(M + M^T - A)^{-1}M^T \tag{43}$$

Note that, as in the above-quoted works, we assume that  $M + M^T - A$  (and therefore  $X$ ) is positive definite, which, see Lemma A.1, amounts to assume

$$\|I - M^{-1}A\|_A < 1 \tag{44}$$

or, equivalently,

$$\rho(I - (\frac{1}{2}(M + M^T))^{-1}A) < 1$$

That is, the symmetric part of the smoother has to define a convergent iterative process.

As proved in Reference [9] (see also Theorem 12 below),  $1 - \mu^{-1}$  is in fact an upper bound on  $\|(I - M^{-1}A)(I - p\hat{A}_C^{-1}p^T A)\|_A^2$ . It is thus interesting to see that  $\tau$  and  $\mu$  are closely related. Letting  $D = \text{diag}(A)$ , commonly used smoothers satisfy the *smoothing property*

$$\|(I - M^{-1}A)\mathbf{z}\|_A^2 \leq \|\mathbf{z}\|_A^2 - \sigma \|\mathbf{z}\|_{AD^{-1}A}^2 \quad \text{for all } \mathbf{z}$$

for some non-trivial  $\sigma > 0$  (see References [7, 11, 12] for examples). This is equivalent to

$$\frac{\mathbf{z}^T(M + M^T - A)\mathbf{z}}{\mathbf{z}^T M^T D^{-1} M \mathbf{z}} \geq \sigma \quad \text{for all } \mathbf{z} \neq 0$$

[11, 12], which shows that  $\sigma^{-1}$  is an upper bound on  $\lambda_{\max}(D_{FF}^{-1}X_{FF})$ , entailing  $\mu \leq \tau/\sigma$ .

This, however, depends upon the smoothing property. The following lemma allows to prove a relation between  $\tau$  and  $\mu$  independently of the latter. Part of its proof is inspired from the proof of Lemma 2.3 in Reference [9].

*Lemma 11*

Let  $A$  be a symmetric positive definite matrix partitioned in  $2 \times 2$  block form, let  $M$  be some non-singular matrix of same size as  $A$ , and let  $X$  be defined by (43). Let

$$Y = \frac{1}{2}(M + M^T), \quad Z = \frac{1}{2}(M - M^T)$$

<sup>§</sup>This fact and its proof were pointed out by some anonymous referee.

If  $X$  is positive definite, then,

$$\frac{1}{2} \leq \frac{\mathbf{z}^T X \mathbf{z}}{\mathbf{z}^T Y \mathbf{z}} \leq \frac{1 + \|Y^{-1/2} Z Y^{-1/2}\|^2}{2 - \lambda_{\max}(Y^{-1} A)} \quad \text{for all } \mathbf{z} \neq 0 \quad (45)$$

and, for any symmetric positive definite matrix  $D_{FF}$  of same size as  $A_{FF}$ ,

$$\begin{aligned} \lambda_{\max}(D_{FF}^{-1} X_{FF}) &\leq \lambda_{\max}(D_{FF}^{-1} Y_{FF}) \frac{1 + \|Y^{-1/2} Z Y^{-1/2}\|^2}{2 - \lambda_{\max}(Y^{-1} A)} \\ \lambda_{\min}(D_{FF}^{-1} X_{FF}) &\geq \max\left(\frac{1}{2} \lambda_{\min}(D_{FF}^{-1} Y_{FF}), \lambda_{\min}(D_{FF}^{-1} A_{FF})\right) \end{aligned}$$

*Proof*

Firstly,

$$\begin{aligned} (M^{-1} + M^{-T})^{-1} &= (M^{-1}(M + M^T)M^{-T})^{-1} \\ &= M^T(M + M^T)^{-1}M \\ &= \frac{1}{4}(2M^T - (M + M^T))(M + M^T)^{-1}(2M - (M + M^T)) \\ &\quad + \frac{1}{4}(M + M^T) \\ &= \frac{1}{2}(Z^T Y^{-1} Z + Y) \end{aligned}$$

Therefore, for all  $\mathbf{z} \neq 0$ ,

$$\begin{aligned} \frac{\mathbf{z}^T Y^{-1} \mathbf{z}}{\mathbf{z}^T X^{-1} \mathbf{z}} &= \frac{\mathbf{z}^T Y^{-1} \mathbf{z}}{\mathbf{z}^T (\frac{1}{2}(M^{-1} + M^{-T})) \mathbf{z}} \frac{\mathbf{z}^T (\frac{1}{2}(M^{-1} + M^{-T})) \mathbf{z}}{\mathbf{z}^T (M^{-1} + M^{-T} - M^{-T} A M^{-1}) \mathbf{z}} \\ &= \frac{\mathbf{z}^T Y^{-1} \mathbf{z}}{\mathbf{z}^T (Y + Z^T Y^{-1} Z)^{-1} \mathbf{z}} \frac{1}{2 - \frac{\mathbf{z}^T M^{-T} A M^{-1} \mathbf{z}}{\mathbf{z}^T (\frac{1}{2}(M^{-1} + M^{-T})) \mathbf{z}}} \end{aligned}$$

The upper bound in (45) follows then from the fact that  $(\frac{1}{2}(M^{-1} + M^{-T}))^{-1} M^{-T} A M^{-1}$  is similar to  $Y^{-1} A$ . On the other hand, the lower bound is proved by checking that, for all  $\mathbf{z}$ ,

$$\mathbf{z}^T X \mathbf{z} \geq \mathbf{z}^T M (M + M^T)^{-1} M^T \mathbf{z} = \mathbf{z}^T (\frac{1}{2}(Z^T Y^{-1} Z + Y)) \mathbf{z} \geq \frac{1}{2} \mathbf{z}^T Y \mathbf{z}$$

The upper bound and the first lower bound on the eigenvalues of  $D_{FF}^{-1} X_{FF}$  are straightforward corollaries of (45), whereas the second lower bound follows from the fact that

$$A^{-1} - X^{-1} = (I - M^{-T} A) A^{-1} (I - A M^{-1})$$

is positive semidefinite, implying  $\lambda_{\min}(A_{FF}^{-1} X_{FF}) \geq 1$ .  $\square$

Since

$$\tau \lambda_{\min}(D_{FF}^{-1} X_{FF}) \leq \mu \leq \tau \lambda_{\max}(D_{FF}^{-1} X_{FF})$$

(with  $D_{FF} = \text{diag}(A_{FF})$ ), this lemma allows to bound  $\mu$  in function of  $\tau$ . That is, the analysis of the smoothing property may be replaced by an analysis of

$$\lambda_{\max}(D_{FF}^{-1}Y_{FF}) \frac{1 + \|Y^{-1/2}ZY^{-1/2}\|^2}{2 - \lambda_{\max}(Y^{-1}A)}$$

Such an analysis should not be that difficult,  $\|Y^{-1/2}ZY^{-1/2}\|$  being essentially a measure on how far  $M$  is from a symmetric matrix, whereas  $\lambda_{\max}(Y^{-1}A)$  is bounded away from 2 when the smoother is properly scaled. For instance, with damped Jacobi smoothing, that is,  $M = \omega^{-1}D$ , one has  $Z = 0$ ,  $Y = M$ , and the lemma yields

$$\frac{\tau}{2\omega} \leq \mu \leq \frac{\tau}{\omega(2 - \omega\lambda_{\max}(D^{-1}A))}$$

Now, we want to go somewhat further and exploit the fact that Theorem 10 may be applied with  $D_{FF} = X_{FF}$ , yielding

$$\frac{1}{\lambda_{\max}(X_{FF}^{-1}A_{FF})} \frac{1}{1 - \hat{\gamma}^2} \leq \mu \leq \frac{1}{\lambda_{\min}(X_{FF}^{-1}A_{FF})} \frac{1}{1 - \hat{\gamma}^2} \tag{46}$$

whereas Lemma 11 applied with  $D_{FF} = A_{FF}$  gives

$$\lambda_{\min}(X_{FF}^{-1}A_{FF}) \geq \lambda_{\min}(Y_{FF}^{-1}A_{FF}) \frac{2 - \lambda_{\max}(Y^{-1}A)}{1 + \|Y^{-1/2}ZY^{-1/2}\|^2} \tag{47}$$

and

$$\lambda_{\max}(X_{FF}^{-1}A_{FF}) \leq 1$$

Here, it should be noted that the smoother is often symmetric and scaled in such a way that  $\lambda_{\max}(M^{-1}A) \leq 1$ . Then, (47) gives  $\lambda_{\min}(X_{FF}^{-1}A_{FF}) \geq \lambda_{\min}(M_{FF}^{-1}A_{FF})$ , and the latter quantity is easily bounded below if (and only if)  $A_{FF}$  is well conditioned.

Theorem 12 below contains our analysis of AMG. The lower bound is new, whereas the upper bounds are obtained by combining (39), (40), (41), (46) and (47) with the known fact that  $\|(I - M^{-1}A)(I - p\hat{A}_C^{-1}p^T A)\|_A^2$  is bounded by  $1 - \mu^{-1}$  (fact for which we give an alternative proof).

*Theorem 12 (analysis of AMG)*

Let  $A$  be a symmetric positive definite matrix partitioned in  $2 \times 2$  block form, and let  $J_{FC}$  be some interpolation matrix. Let  $B_{AMG}$  be the AMG preconditioning matrix as defined by (8), with non-singular smoother  $M$  such that  $\|I - M^{-1}A\|_A < 1$ . Let  $\hat{A}$  be the matrix resulting from the application of the generalized basis transformation defined by (9), (10), and let  $\hat{\gamma}$  be the C.B.S. constant associated with  $\hat{A}$ . Let  $X$  and  $\mu$  be defined by (43), (42), respectively. Then:

$$\begin{aligned} \kappa(B_{AMG}^{-1}A) &\leq \mu \\ &\leq \frac{1}{(1 - \hat{\gamma}^2)\lambda_{\min}(X_{FF}^{-1}A_{FF})} \\ &\leq \frac{1 + \|Y^{-1/2}ZY^{-1/2}\|^2}{(1 - \hat{\gamma}^2)(2 - \lambda_{\max}(Y^{-1}A))\lambda_{\min}(Y_{FF}^{-1}A_{FF})} \end{aligned}$$

where  $Y = \frac{1}{2}(M + M^T)$ ,  $Z = \frac{1}{2}(M - M^T)$ . Moreover, for any symmetric positive definite block-diagonal matrix

$$D = \begin{pmatrix} D_{FF} & \\ & D_{CC} \end{pmatrix}$$

there holds

$$\kappa(B_{\text{AMG}}^{-1}A) \geq \max \left( \frac{\lambda_{\min}(D^{-1}Y)}{2(1 - \hat{\gamma}^2)\lambda_{\max}(A_{FF}(D_{FF}^{-1} + J_{FC}D_{CC}^{-1}J_{FC}^T))}, \frac{\lambda_{\min}(D^{-1}Y)}{2\lambda_{\min}(A_{FF}(D_{FF}^{-1} + J_{FC}D_{CC}^{-1}J_{FC}^T))} \right)$$

*Proof*

Let  $K = I - p\hat{A}_C^{-1}p^T A$ . One has

$$\begin{aligned} \|(I - M^{-1}A)(I - p\hat{A}_C^{-1}p^T A)\|_A^2 &= \max_{z \neq 0} \frac{z^T K^T (I - AM^{-T})A(I - M^{-1}A)Kz}{z^T Az} \\ &= \max_{z \neq 0} \frac{z^T (K^T AK - K^T AX^{-1}AK)z}{z^T Az} \\ &= \max_{w \neq 0} \frac{w^T (\hat{K}^T \hat{A} \hat{K} - \hat{K}^T \hat{A} \hat{X}^{-1} \hat{A} \hat{K})w}{w^T \hat{A} w} \end{aligned}$$

where

$$\begin{aligned} \hat{K} &= J^{-1}KJ \\ &= I - J^{-1}p\hat{A}_C^{-1}p^T J^{-T}\hat{A} \\ &= I - \begin{pmatrix} 0 & 0 \\ 0 & \hat{A}_C^{-1} \end{pmatrix} \hat{A} \\ &= \begin{pmatrix} I & 0 \\ -\hat{A}_C^{-1}\hat{A}_{CF} & 0 \end{pmatrix} \end{aligned}$$

Therefore,

$$\hat{A}\hat{K} = \begin{pmatrix} S_{\hat{A}}^{(F)} & 0 \\ 0 & 0 \end{pmatrix} = qS_{\hat{A}}^{(F)}q^T$$

with  $q$  defined by (12). One then obtains, remembering (37),

$$\begin{aligned} \|(I - M^{-1}A)(I - p\hat{A}_C^{-1}p^T A)\|_A^2 &= \max_{w \neq 0} \frac{w^T (qS_{\hat{A}}^{(F)}q^T - qS_{\hat{A}}^{(F)}q^T \hat{X}^{-1} qS_{\hat{A}}^{(F)}q^T)w}{w^T \hat{A} w} \\ &= 1 - \min_{w_F \neq 0} \frac{w_F^T q^T \hat{X}^{-1} q w_F}{w_F^T (S_{\hat{A}}^{(F)})^{-1} w_F} \end{aligned}$$

Further,

$$q^T \hat{X}^{-1} q = (S_{\hat{X}}^{(F)})^{-1} = (X_{FF} - \hat{X}_{FC} \hat{X}_{CC}^{-1} \hat{X}_{CF})^{-1}$$

hence  $\mathbf{w}_F^T q^T \hat{X}^{-1} q \mathbf{w}_F \geq \mathbf{w}_F^T X_{FF}^{-1} \mathbf{w}_F$ . The upper bounds then follow from (39), (40), (41), (46), (47) and (38), which may be read  $\mu = 1/\lambda_{\min}(X_{FF}^{-1} S_{\hat{A}}^{(F)})$ .

To check the lower bounds, observe that, with (45),

$$\mathbf{z}^T X^{-1} \mathbf{z} \leq 2 \mathbf{z}^T Y^{-1} \mathbf{z} \leq \frac{2}{\lambda_{\min}(D^{-1}Y)} \mathbf{z}^T D^{-1} \mathbf{z} \quad \text{for all } \mathbf{z}$$

Hence, letting  $\hat{D} = J^T D J$ ,

$$\begin{aligned} \min_{\mathbf{w}_F \neq 0} \frac{\mathbf{w}_F^T q^T \hat{X}^{-1} q \mathbf{w}_F}{\mathbf{w}_F^T (S_{\hat{A}}^{(F)})^{-1} \mathbf{w}_F} &\leq \frac{2}{\lambda_{\min}(D^{-1}Y)} \left( \min_{\mathbf{w}_F \neq 0} \frac{\mathbf{w}_F^T q^T \hat{D}^{-1} q \mathbf{w}_F}{\mathbf{w}_F^T (S_{\hat{A}}^{(F)})^{-1} \mathbf{w}_F} \right) \\ &\leq \frac{2}{\lambda_{\min}(D^{-1}Y)} \left( \min_{\mathbf{w}_F \neq 0} \frac{\mathbf{w}_F^T A_{FF}^{-1} \mathbf{w}_F}{\mathbf{w}_F^T (S_{\hat{A}}^{(F)})^{-1} \mathbf{w}_F} \right) \left( \max_{\mathbf{w}_F \neq 0} \frac{\mathbf{w}_F^T q^T \hat{D}^{-1} q \mathbf{w}_F}{\mathbf{w}_F^T A_{FF}^{-1} \mathbf{w}_F} \right) \\ &= \frac{2}{\lambda_{\min}(D^{-1}Y)} (1 - \hat{\gamma}^2) \lambda_{\max}(A_{FF}(q^T \hat{D}^{-1} q)) \end{aligned}$$

and, similarly,

$$\begin{aligned} \min_{\mathbf{w}_F \neq 0} \frac{\mathbf{w}_F^T q^T \hat{X}^{-1} q \mathbf{w}_F}{\mathbf{w}_F^T (S_{\hat{A}}^{(F)})^{-1} \mathbf{w}_F} &\leq \frac{2}{\lambda_{\min}(D^{-1}Y)} \left( \max_{\mathbf{w}_F \neq 0} \frac{\mathbf{w}_F^T A_{FF}^{-1} \mathbf{w}_F}{\mathbf{w}_F^T (S_{\hat{A}}^{(F)})^{-1} \mathbf{w}_F} \right) \left( \min_{\mathbf{w}_F \neq 0} \frac{\mathbf{w}_F^T q^T \hat{D}^{-1} q \mathbf{w}_F}{\mathbf{w}_F^T A_{FF}^{-1} \mathbf{w}_F} \right) \\ &= \frac{2}{\lambda_{\min}(D^{-1}Y)} \lambda_{\min}(A_{FF}(q^T \hat{D}^{-1} q)) \end{aligned}$$

Since  $q^T \hat{D}^{-1} q = (q^T J^{-1}) D^{-1} (q^T J^{-1})^T = D_{FF}^{-1} + J_{FC} D_{CC}^{-1} J_{FC}^T$ , the required results follow.  $\square$

The connection between the AMG theory and the C.B.S. constant is already implicit in some previous works, see Reference [12, Section A.5]. Our analysis, however, shows more clearly that the corresponding two-level scheme converges fast if *and only if*  $\hat{\gamma}$  is bounded away from 1 and  $A_{FF}$  is well conditioned.

#### 4. CONCLUSIONS

Remarkably, it turns out that, despite their differences, all the schemes are efficient under the same general conditions, that is, they work well if and only if  $A_{FF}$  is well conditioned and  $\hat{\gamma}$  bounded away from 1, where  $\hat{\gamma}$  is the C.B.S. constant associated with the matrix  $\hat{A}$  resulting from the generalized basis transformation (10). The conditioning of  $A_{FF}$  is even

doubly important because when it is good, it is also easier to define interpolation matrices  $J_{FC}$  for which the corresponding  $\hat{\gamma}$  is nicely bounded away from 1 (see Section 3.5).

Now, this requirement on  $A_{FF}$  should be seen as a constraint on the  $F/C$  partitioning. In other words, the coarsening process has to be designed such as to guarantee the good conditioning of  $A_{FF}$ , in one way or another. This is just what is behind the compatible relaxation idea developed in References [9, 18]: an  $F/C$  partitioning is tested, and possibly iteratively improved, by assessing the convergence of a basic iterative solution scheme for a system with  $A_{FF}$ .

A detailed comparison of the bounds is difficult. However, a rough comparison is easily obtained by considering only the dependence with respect to  $\hat{\gamma}$ , assuming all other terms sufficiently close to 1. On this basis, the situation is somewhat less favourable for HBBD, with a condition number of order  $1/(1 - \hat{\gamma})$ , whereas all other methods exhibit a condition number close to  $1/(1 - \hat{\gamma}^2)$ . Comparing AMG with HBMG, one may be prone to believe that AMG should converge at least as fast, since relaxing *in addition* the  $C$  unknowns should not have an adverse effect. However the analysis does not support this claim, and if it does not prove the converse, it is remarkable that the bound for AMG mainly depends on the way the  $A_{FF}$  block is approximated by the smoother, and very little on the global properties of the latter.

On the whole, we did not thus find enough differences to state that one method is better than another. That is, all two-level schemes (except HBBD) have roughly comparable convergence properties. In practice, the relative behaviour of the associated methods will mainly depend on factors that are beyond this analysis: their behaviour in the context of a *multilevel* cycle and the computational cost of each cycle, for instance. From that point of view, let us just mention that AMG tends to be more costly, but is often viable with a simple V cycle, whereas other methods require in general W cycles, often polynomially accelerated. On the other hand, maybe because of this use of V cycles, AMG seems more demanding with respect to the quality of the interpolation. For instance, see Reference [12, p. 524], it is not advised to use AMG with a simple aggregation scheme, which corresponds to the crudest but cheapest possible interpolation. Nevertheless, the associated measure  $\tau$  (36) is not necessarily that bad for  $M$ -matrices (see Reference [12, Equation (A.4.21)]), and good results are obtained in Reference [55] with an MBF scheme that uses coarse grid matrices built by aggregation.

## APPENDIX A

### *Lemma A.1*

Let  $A$  be a symmetric positive definite  $n \times n$  matrix, and  $R$  some  $n \times n$  matrix. The following two propositions are equivalent.

- (1)  $\|I - RA\|_A \leq 1$  ( $< 1$ ).
- (2)  $R + R^T - RAR^T$  is positive semidefinite (positive definite).

Moreover, if  $R$  is non-singular and  $M = R^{-1}$ , they are also equivalent to:

- (3)  $M + M^T - A$  is positive semidefinite (positive definite).
- (4)  $\rho(I - (\frac{1}{2}(M + M^T))^{-1}A) \leq 1$  ( $< 1$ ).

*Proof*

We develop only the proof for the case of non-strict inequalities and positive semidefiniteness, the proof of the other case being similar. Firstly, (1)  $\iff$  (2) because

$$\begin{aligned} \|I - RA\|_A \leq 1 &\iff \|I - A^{1/2}RA^{1/2}\| \leq 1 \\ &\iff \lambda_{\max}((I - A^{1/2}RA^{1/2})(I - A^{1/2}R^T A^{1/2})) \leq 1 \\ &\iff \lambda_{\max}(I - A^{1/2}(R + R^T - RAR^T)A^{1/2}) \leq 1 \\ &\iff R + R^T - RAR^T \text{ is positive semidefinite} \end{aligned}$$

On the other hand, if  $R$  is non-singular and  $M = R^{-1}$ , one has  $R + R^T - RAR^T = M^{-1}(M + M^T - A)M^{-T}$ . Hence,

$$\begin{aligned} R + R^T - RAR^T \text{ is positive semidefinite} \\ &\iff M + M^T - A \text{ is positive semidefinite} \\ &\iff \mathbf{z}^T(M + M^T)\mathbf{z} \geq \mathbf{z}^T A \mathbf{z} \geq 0 \quad \forall \mathbf{z} \in \mathbb{C}^n \\ &\iff 0 \leq \frac{\mathbf{z}^T A \mathbf{z}}{\mathbf{z}^T (\frac{1}{2}(M + M^T))\mathbf{z}} \leq 2 \end{aligned}$$

which proves (2)  $\iff$  (3)  $\iff$  (4). □

#### ACKNOWLEDGEMENTS

I was initiated to this field by a patient reading of papers by Owe Axelsson and his coworkers. I dedicate this work to him with great pleasure.

#### REFERENCES

1. Hackbusch W. *Multi-grid Methods and Applications*. Springer: Berlin, 1985.
2. Trottenberg U, Oosterlee CW, Schüller A. *Multigrid*. Academic Press: London, 2001.
3. Axelsson O. *Iterative Solution Methods*. Cambridge University Press: Cambridge, 1994.
4. Axelsson O. A survey of algebraic multilevel iterations (AMLI) methods. *BIT* 2003; **43**:863–879.
5. Axelsson O, Gustafsson I. Preconditioning and two-level multigrid methods of arbitrary degree of approximation. *Mathematics of Computation* 1983; **40**:214–242.
6. Axelsson O, Vassilevski PS. Algebraic multilevel preconditioning methods, II. *SIAM Journal on Numerical Analysis* 1990; **27**:1569–1590.
7. Brandt A. Algebraic multigrid theory: the symmetric case. *Applied Mathematics and Computation* 1986; **19**: 23–56.
8. Eijkhout V, Vassilevski PS. The role of the strengthened c.b.s. inequality in multilevel methods. *SIAM Review* 1991; **33**:405–419.
9. Falgout RD, Vassilevski PS. On generalizing the AMG framework. *Technical Report UCRL-JC-150807*, Lawrence Livermore National Laboratory, 2003.
10. Notay Y. Using approximate inverses in algebraic multilevel methods. *Numerische Mathematik* 1998; **80**: 397–417.
11. Ruge JW, Stüben K. Algebraic multigrid (AMG). In *Multigrid Methods, Frontiers in Applied Mathematics*, vol. 3, McCormick SF (ed.). SIAM: Philadelphia, 1987; 73–130.
12. Stüben K. An Introduction to Algebraic Multigrid. In *Multigrid*, Trottenberg *et al.* (eds), Appendix A. Academic Press: London, 2001; 413–532.
13. Vassilevski PS. Nearly optimal iterative methods for solving finite element elliptic equations based on the multilevel splitting of the matrix. *Technical Report # 1989-09*, Institute for Scientific Computation, University of Wyoming, Laramie, U.S.A., 1989.

14. Ruge JW, Stüben K. Efficient solution of finite difference and finite element equations by algebraic multigrid (AMG). In *Multigrid Methods for Integral and Differential Equations*, Paddon DJ, Holstein H (eds), The Institute of Mathematics and its Applications Conference Series. Clarendon Press: Oxford, 1985; 169–212.
15. Stüben K. Algebraic multigrid (AMG): experiences and comparisons. *Applied Mathematics and Computation* 1983; **13**:419–452.
16. Bank RE, Smith RK. An algebraic multilevel multigraph algorithm. *SIAM Journal on Scientific Computing* 2002; **23**:1572–1592.
17. Braess D. Towards algebraic multigrid for elliptic problems of second order. *Computing* 1995; **55**:379–393.
18. Brandt A. General highly accurate algebraic coarsening. *Electronic Transactions on Numerical Analysis* 2000; **10**:1–20.
19. Brezina M, Cleary AJ, Falgout RD, Henson VE, Jones JE, Manteuffel TA, McCormick SF, Ruge JW. Algebraic multigrid based on element interpolation (AMGe). *SIAM Journal on Scientific Computing* 2000; **22**: 1570–1592.
20. Cleary AJ, Falgout RD, Henson VE, Jones JE, Manteuffel TA, McCormick SF, Miranda GN, Ruge JW. Robustness and scalability of algebraic multigrid. *SIAM Journal on Scientific Computing* 2000; **21**:1886–1908.
21. Chartier T, Falgout RD, Henson VE, Jones J, Manteuffel T, McCormick S, Ruge J, Vassilevski PS. Spectral AMGe ( $\rho$ AMGe). *SIAM Journal on Scientific Computing* 2004; **25**:1–26.
22. Henson VE, Vassilevski PS. Element-free AMGe: general algorithms for computing interpolation weights in AMG. *SIAM Journal on Scientific Computing* 2002; **23**:629–650.
23. Jones JE, Vassilevski PS. AMGE based on element agglomeration. *SIAM Journal on Scientific Computing* 2001; **23**:109–133.
24. Vaněk P, Mandel J, Brezina M. Algebraic multigrid based on smoothed aggregation for second and fourth order problems. *Computing* 1996; **56**:179–196.
25. Bank RE. Hierarchical bases and the finite element method. *Acta Numerica* 1996; **5**:1–43.
26. Bank RE, Dupont TF. Analysis of a two-level scheme for solving finite element equations. *Technical Report CNA-159*, Center for Numerical Analysis, The University of Texas at Austin, Texas, U.S.A., 1980.
27. Yserentant H. On the multi-level splitting of finite element spaces. *Numerische Mathematik* 1986; **49**:379–412.
28. Chow E, Vassilevski PS. Multilevel block factorizations in generalized hierarchical bases. *Numerical Linear Algebra with Applications* 2003; **10**:105–127.
29. Bramble JH, Pasciak JE, Xu JC. Parallel multilevel preconditioners. *Mathematics of Computation* 1990; **55**(191):1–22.
30. Axelsson O. Stabilization of algebraic multilevel iteration methods: additive methods. *Numerical Algorithms* 1999; **21**:23–47.
31. Axelsson O, Padiy A. On the additive version of the algebraic multilevel iteration method for anisotropic elliptic problems. *SIAM Journal on Scientific Computing* 1999; **20**:1807–1830.
32. Axelsson O. On multigrid methods of the two-level type. In *Multigrid Methods*, Hackbusch W, Trottenberg U (eds), Lectures Notes in Mathematics, vol. 960. Springer: Berlin, Heidelberg, New York, 1982; 352–367.
33. Axelsson O. The stabilized V-cycle method. *Journal of Computational and Applied Mathematics* 1996; **74**: 33–50.
34. Axelsson O, Vassilevski PS. Algebraic multilevel preconditioning methods, I. *Numerische Mathematik* 1989; **56**:157–177.
35. Axelsson O, Vassilevski PS. Variable-step multilevel preconditioning methods. I. selfadjoint and positive definite elliptic problems. *Numerical Linear Algebra with Applications* 1994; **1**:75–101.
36. Vassilevski PS. Hybrid V-cycle algebraic multilevel preconditioners. *Mathematics of Computation* 1992; **58**: 489–512.
37. Vassilevski PS. On two ways of stabilizing the hierarchical basis multilevel methods. *SIAM Review* 1997; **39**:18–53.
38. Bank RE, Dupont TF, Yserentant H. The hierarchical basis multigrid method. *Numerische Mathematik* 1988; **52**:427–458.
39. Axelsson O. The method of diagonal compensation of reduced matrix entries and multilevel iteration. *Journal of Computational and Applied Mathematics* 1991; **38**:31–43.
40. Axelsson O. On algebraic multilevel iterations for methods for selfadjoint elliptic problems with anisotropy. In *Fascicolo Speciale, Numerical Methods*, Rend. Sem. Mat. Univers. Politech. Torino, 1991; 31–61.
41. Axelsson O, Eijkhout V. Analysis of recursive 5-point/9-point factorization method. In *Preconditioned Conjugate Gradient Methods*, Axelsson O, Kolotilina LYU (eds), Lectures Notes in Mathematics, vol. 1457. Springer: Berlin, Heidelberg, New York, 1990; 154–173.
42. Axelsson O, Eijkhout V. The nested recursive two level factorization for nine-point difference matrices. *SIAM Journal on Scientific and Statistical Computing* 1991; **12**:1373–1400.
43. Axelsson O, Neytcheva M. Algebraic multilevel iterations for Stieltjes matrices. *Numerical Linear Algebra with Applications* 1994; **1**:213–236.
44. Bank RE, Smith RK. The incomplete factorization multigraph algorithm. *SIAM Journal on Scientific Computing* 1999; **20**:1349–1364.

45. Bank RE, Wagner C. Multilevel ILU decomposition. *Numerische Mathematik* 1999; **82**:543–576.
46. Botta EF, van der Ploeg A. Preconditioning techniques for matrices with arbitrary sparsity patterns. Department of Mathematics, University of Groningen, The Netherlands, 1995, preprint.
47. Botta EFF, Wubs FW. Matrix renumbering ILU: an effective algebraic multilevel ILU preconditioner for sparse matrices. *SIAM Journal on Matrix Analysis and Applications* 1999; **20**:1007–1026.
48. Kuznetsov YuA. Algebraic multigrid domain decomposition methods. *Soviet Journal of Numerical Analysis and Mathematical modelling* 1989; **4**:361–392.
49. Li Z, Saad Y, Sosonkina M. pARMS: a parallel version of the algebraic recursive multilevel solver. *Numerical Linear Algebra with Applications* 2003; **10**:485–509.
50. Notay Y, Ould Amar Z. A nearly optimal preconditioning based on recursive red-black orderings. *Numerical Linear Algebra with Applications* 1997; **4**:369–391.
51. Notay Y. Optimal V cycle algebraic multilevel preconditioning. *Numerical Linear Algebra with Applications* 1998; **5**:441–459.
52. Notay Y. A multilevel block incomplete factorization preconditioning. *Applied Numerical Mathematics* 1999; **31**:209–225.
53. Notay Y. A robust algebraic multilevel preconditioner for nonsymmetric  $M$ -matrices. *Numerical Linear Algebra with Applications* 2000; **7**:243–267.
54. Notay Y. Optimal order preconditioning of finite difference matrices. *SIAM Journal on Scientific Computing* 2000; **21**:1991–2007.
55. Notay Y. Robust parameter free algebraic multilevel preconditioning. *Numerical Linear Algebra with Applications* 2002; **9**:409–428.
56. Reusken A. On the approximate cyclic reduction preconditioner. *SIAM Journal on Scientific Computing* 1999; **21**:565–590.
57. Saad Y. ILUM: a parallel multi-elimination ILU preconditioner for general sparse matrices. *SIAM Journal on Scientific Computing* 1996; **17**:830–847.
58. Saad Y, Suchomel B. ARMS: an algebraic recursive multilevel solver for general sparse linear systems. *Numerical Linear Algebra with Applications* 2002; **9**:359–378.
59. Saad Y, Zhang J. BILUM: block versions of multielimination and multilevel ILU preconditioner for general sparse linear systems. *SIAM Journal on Scientific Computing* 1999; **20**:2103–2121.
60. van der Ploeg A, Botta EFF, Wubs FW. Nested grids ILU-decomposition (NGILU). *Journal of Computational and Applied Mathematics* 1996; **66**:515–526.
61. Achchab B, Axelsson O, Laayouni L, Souissi A. Strengthened Cauchy–Bunyakovski–Schwarz inequality for a three-dimensional elasticity system. *Numerical Linear Algebra with Applications* 2001; **8**:191–205.
62. Achchab B, Maître JF. Estimate of the constant in two strengthened C.B.S. inequalities for F.E.M. systems of 2D elasticity: application to multilevel methods and a posteriori error estimators. *Numerical Linear Algebra with Applications* 1996; **3**:147–159.
63. Axelsson O, Blaheta R. Two simple derivations of universal bounds for the C.B.S. inequality constant. *Applications of Mathematics* 2004; **49**:57–72.
64. Blaheta R. Nested tetrahedral grids and strengthened C.B.S. inequality. *Numerical Linear Algebra with Applications* 2003; **10**:619–637.
65. Maître JF, Musy F. The contraction number of a class of two-level methods: an exact evaluation for some finite element subspaces and model problems. In *Multigrid Methods*, Hackbusch W, Trottenberg U (eds), Lectures Notes in Mathematics, vol. 960. Springer: Berlin, Heidelberg, New York, 1982; 535–544.
66. Margenov S. Upper bound of the constant in the strengthened C.B.S. inequality for FEM 2D elasticity equations. *Numerical Linear Algebra with Applications* 1994; **1**:65–74.
67. Chan TF, van der Vorst H. Approximate and incomplete factorizations. In *Parallel Numerical Algorithms*, Keyes DE, Samed A, Venkatakrishnan V (eds), ICASE/LaRC Interdisciplinary Series in Science and Engineering, vol. 4. Kluwer Academic Publishers: Dordrecht, 1997; 167–202.
68. Concus P, Golub GH, Meurant G. Block preconditioning for the conjugate gradient method. *SIAM Journal on Scientific and Statistical Computing* 1985; **6**:220–252.
69. Dupont T, Kendall RP, Rachford HH. An approximate factorization procedure for solving self-adjoint elliptic difference equations. *SIAM Journal on Numerical Analysis* 1968; **5**:559–573.
70. Magolu MM, Notay Y. On the conditioning analysis of block approximate factorization methods. *Linear Algebra and its Applications* 1991; **154–156**:583–599.