

Optimal V-cycle Algebraic Multilevel Preconditioning

Yvan Notay*

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

We consider algebraic multilevel preconditioning methods based on the recursive use of a 2×2 block incomplete factorization procedure in which the Schur complement is approximated by a coarse grid matrix. As is well known, for discrete second-order elliptic PDEs, optimal convergence properties are proved for such basic two-level schemes under mild assumptions on the PDE coefficients, but their recursive use in a simple V-cycle algorithm does not generally lead to optimal order convergence.

In the present paper, we analyse the combination of these techniques with a smoothing procedure much the same as the one used in standard multigrid algorithms, except that smoothing is not required on the finest grid. Theoretical results prove optimal convergence properties for the V-cycle under an assumption similar to the ‘approximation property’ of the classical multigrid convergence theory. On the other hand, numerical experiments made on both 2D and 3D problems show that the condition number is close to that of the two-level method. Further, the method appears robust in the presence of discontinuity and anisotropy, even when the material interfaces are not aligned with the coarse grid. © 1998 John Wiley & Sons, Ltd.

KEY WORDS iterative methods for linear systems; acceleration of convergence; preconditioning

1. Introduction

This paper deals with the iterative solution of large sparse symmetric positive (non-negative) definite linear systems

$$A\mathbf{u} = \mathbf{b} \quad (1.1)$$

arising from the discretization of second order elliptic PDEs. In such cases, the conjugate gradient method combined with a suitable preconditioning is a choice method (e.g., [4,5,16,22]). The preconditioner here is a symmetric positive definite matrix B such that solving a system with B is easy whereas the (spectral) condition number of the precondi-

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

tioned system

$$\kappa(B^{-1}A) = \frac{\lambda_{\max}(B^{-1}A)}{\lambda_{\min}(B^{-1}A)}$$

on which depends the convergence rate, has to be as small as possible¹. In particular, a preconditioner has an optimal order of computational complexity when solving a system with B requires $\mathcal{O}(n)$ operations (where n is the number of unknowns), and when $\kappa(B^{-1}A)$ is bounded independently of the grid size and possibly other problem dependent parameters.

To meet both these requirements, many recent works focus on the design of algebraic multilevel preconditioners (e.g., [3,6,7,10–12,31,35–39]). Like standard multigrid [23,40], these are based on the recursive use of a two-level method. However, the basic scheme originates from a block incomplete factorization of the system matrix partitioned in a 2×2 block form

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad (1.2)$$

where the first block of unknowns corresponds to the fine grid nodes and the second block of unknowns to the coarse grid nodes.

For such two-level methods, eigenvalue bounds are generally proved by means of local analyses, that is, in the finite element context, by computing certain quantities associated with element matrices. Consequently, these bounds hold independently of the grid size and of possible jumps in the PDE coefficients, as long as these may only occur across the edges of the coarse mesh (e.g., [2,8,25,28,38]).

Hence, algebraic multilevel methods exhibit a strong potential for robustness. Unfortunately, see [38], their recursive use in a simple V-cycle algorithm does not generally result in an optimal order preconditioning. Moreover, so far, proposed W-cycle schemes (e.g., [3,6,7,10–12,36–38]) require that some parameters are determined in function of eigenvalue estimates. In some cases, one can rely on theoretical results. However, as inaccurate estimates may lead to poor performances or even to an indefinite preconditioner, a better robustness is achieved with run-time eigenvalue estimation. Relevant procedures are proposed and discussed in [10], their main drawback being the additional implementation effort required and an increase of the cost of the preprocessing step.

Here, we develop another approach, in which one adds to a simple V-cycle some smoothing iterations in a similar way as is done in standard multigrid methods, except that smoothing is not required on the finest grid. Theoretically, we prove optimal order convergence of the resulting method for ‘well-behaved’ problems, whereas, in practice, the condition number appears independent of the number of levels and close to that of the two-level preconditioner even for ‘difficult’ problems for which classical multigrid methods would not work.

To our knowledge, the combination of algebraic multilevel methods (based on block incomplete factorization) with smoothing has not been proposed before. Our theoretical results are also new. Compared with algebraic multigrid methods developed along the lines of the results in [34,32,33], our method uses simpler coarse grid matrices and allows fixed (geometric) coarsening, thus giving simpler and cheaper implementations whenever working with regular grids. Alternative approaches using algebraic multigrid with simple coarse grid matrices include methods based on aggregation, see e.g., [18].

¹ Throughout this paper, $\lambda_{\max}(C)$ and $\lambda_{\min}(C)$ denote, respectively, the largest and the smallest eigenvalue of C .

The remainder of this paper is organized as follows: in Section 2, we present algebraic multilevel preconditioning methods and motivate our approach to stabilize the V-cycle; the theoretical analysis is developed in Section 3 and the results of numerical experiments are reported in Section 4.

2. Algebraic multilevel preconditioning and smoothing

An exact factorization of the system matrix (1.2) writes

$$A = \begin{pmatrix} A_{11} & \\ A_{21} & S \end{pmatrix} \begin{pmatrix} I & A_{11}^{-1} A_{12} \\ & I \end{pmatrix} \tag{2.1}$$

where $S = A_{22} - A_{21} A_{11}^{-1} A_{12}$ is the Schur complement. This leads to a direct method, which is impractical because S is generally dense. However, it is most often easy to find a sparse approximation A_c to S that has a similar structure to that of A . One then obtains the following two-level block incomplete factorization preconditioner:

$$B_{(2)} = \begin{pmatrix} A_{11} & \\ A_{21} & A_c \end{pmatrix} \begin{pmatrix} I & A_{11}^{-1} A_{12} \\ & I \end{pmatrix} \tag{2.2}$$

Initially, this method was proposed for finite element matrices computed with the hierarchical basis [2,8,11,12,36–38], generalizing in this way the hierarchical basis multigrid method [13–15,42]. In this context, one uses $A_c = A_{22}^{(hb)}$, which corresponds to the finite element discretization on the coarse mesh.

However, as observed in [36], elementary computation shows that, for any $B_{(2)}$ of the form (2.2),

$$B_{(2)}^{-1} A = \begin{pmatrix} I & * \\ & A_c^{-1} S \end{pmatrix} \tag{2.3}$$

where the Schur complement S is in addition invariant under a basis transformation of the form

$$T = \begin{pmatrix} I & J_{12} \\ & I \end{pmatrix} \tag{2.4}$$

like the one associated with the change from the hierarchical basis to the nodal basis. Consequently, identical convergence results are obtained with a block incomplete factorization (2.2) performed in the usual nodal basis. Since $A^{(hb)} = T^t A T$, one has then

$$A_c = A_{22}^{(hb)} = A_{22} + A_{21} J_{12} + J_{12}^t A_{12} + J_{12}^t A_{11} J_{12}, \tag{2.5}$$

or, equivalently,

$$A_c = r A p \tag{2.6}$$

where

$$p = \begin{pmatrix} J_{12} \\ I \end{pmatrix}, \quad r = p^t \tag{2.7}$$

A_c is thus the ‘Galerkin approximation’ corresponding to the prolongation p .

Nevertheless, whenever working with the nodal basis, it is generally more convenient to compute A_c by means of more algebraic techniques as proposed in [28,31]. For instance,

letting Δ be the diagonal matrix with same row-sum as A_{11} , the method in [28] takes as coarse grid matrix the Schur complement \tilde{S} of

$$\tilde{A} = \begin{pmatrix} \Delta & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

in which zero rows and columns are symmetrically deleted. With this simple method, a 2D five-point or a 3D seven-point matrix is always approximated on the coarse grid by a matrix with, respectively, a five-point or a seven-point structure.

Whatever the choice of A_c , the key features of the preconditioner (2.2) are two-fold. On the one hand, see (2.3), the associated condition number and convergence rate will be independent of the mesh size if and only if A_c is spectrally equivalent to S , which is proved by means of local analyses, thus independent of possible jumps in the PDE coefficients as long as these may only occur across the edges of the coarse mesh (e.g., [2,8,25,28,38]).

On the other hand, A_{11} has condition number $\mathcal{O}(1)$ [2,8,14], so that the two required solutions of a system with A_{11} can be done in a number of operations proportional to the number of unknowns in block 1. In practice, these solutions are most often made only approximately, i.e., A_{11} has to be exchanged in the formulation (2.2) for some preconditioner P . Whenever working with the hierarchical basis, it clearly follows from the analysis in [4,8] that the condition number associated with

$$B_{(2)}^{(P)} = \begin{pmatrix} P & \\ A_{21} & A_c \end{pmatrix} \begin{pmatrix} I & P^{-1} A_{12} \\ & I \end{pmatrix} \quad (2.8)$$

cannot be much worse than $\kappa(P^{-1} A_{11}) \cdot \kappa(A_c^{-1} S)$.

Whenever working with the nodal basis, one should be more cautious, as inappropriate choice of P may lead to the loss of the spectral equivalence with A [29]. However, problems are prevented when P satisfies some given algebraic requirements, which, when A is an M matrix, are in particular met by letting P be an MILU factorization of A_{11}^1 . Results in [29] prove then for $\lambda_{\min}(B_{(2)}^{(P)-1} A)$ and $\lambda_{\max}(B_{(2)}^{(P)-1} A)$ bounds that are close to, respectively, $\lambda_{\min}(P^{-1} A_{11}) \cdot \lambda_{\min}(A_c^{-1} S)$ and $\lambda_{\max}(P^{-1} A_{11}) \cdot \lambda_{\max}(A_c^{-1} S)$.

In the theoretical analysis developed below, we assume $P = A_{11}$ from the second level. One possible approach consists then of inverting A_{11} to some sufficient precision by a proper iterative scheme, the extra cost being acceptable since the finest grid is not involved. On the other hand, satisfactory numerical results are obtained in Section 4 when using an MILU factorization of A_{11} at every level. In some sense, this approach relies on the fact that the multilevel method is likely to be stable when the basic two-level scheme itself is not deeply perturbed by the approximate inversion of A_{11} .

Combination with a smoother

If

$$p_A = \begin{pmatrix} -P^{-1} A_{12} \\ I \end{pmatrix}, \quad r_A = \begin{pmatrix} -A_{21} P^{-1} & I \end{pmatrix} \quad (2.9)$$

¹ By MILU we refer to an incomplete LU decomposition in which discarded fill entries are lumped to the diagonal to force P to have the same row-sum as A_{11} [4,16].

can be considered as, respectively, (matrix dependent) prolongation and restriction,

$$B_{(2)}^{(P)^{-1}} = \begin{pmatrix} P^{-1} & 0 \\ 0 & 0 \end{pmatrix} + p_A A_c^{-1} r_A \tag{2.10}$$

appears similar to a coarse grid correction term as used in multigrid methods [21,31]. This observation led us to investigate the combined use of algebraic multilevel preconditioning with a smoothing procedure. At first sight, this is not a clever approach, since smoothing is costly on the finest grid whereas the preconditioner (2.2) has already grid independent convergence. But, in the context of the recursive use of the two-level method, smoothing can help to stabilize the condition number, multigrid schemes being usually optimal with a simple V-cycle [23]. On the other hand, such a stabilization does not require smoothing on the finest grid, so that the extra cost is moderate and in fact smaller than that involved by a W-cycle.

In this respect, note however that, whereas the efficiency of the two-level method alone is largely independent of the scaling of A_c (see (2.3) when $P = A_{11}$), its proper interaction with a smoother is not. Roughly speaking, the eigenvalues of $A_c^{-1}S$ corresponding to ‘smooth’ eigenvectors should be close to 1. Now, the theoretical results below show that ‘all goes right’ when A_c is the Galerkin operator (2.6) corresponding to a prolongation and a restriction of the form (2.7), for which the classical multigrid theory would apply. A practical approach to fix the scaling factor consists then of comparing for some model problem the tentative coarse grid matrix with the Galerkin matrix resulting from linear interpolation. For instance, considering \tilde{S} and the standard finite difference approximation of the Laplacian, this procedure yields

$$A_c = \begin{cases} 2\tilde{S} & \text{for 2D five-point matrices} \\ 4\tilde{S} & \text{for 3D seven-point matrices} \end{cases} \tag{2.11}$$

For non-model problems, there is not much theoretical certainty in using these rules, but the robustness of our method is anyway more or less empirical since the theoretical results are based on the classical multigrid theory, which itself does not apply as soon as the PDE coefficients have jumps inside coarse grid cells.

Remark 1

In algebraic multigrid methods (AMG, e.g., [34,32,33]), the interpolation is defined by solving approximately the fine grid residual equations, which means that the prolongation and the restriction have also the form (2.7) with P a (rough) approximation of A_{11} . This further confirms our guess that the two-level preconditioner (2.8) effectively acts as a coarse grid correction, and should thus work well in combination with a smoother.

Besides this point, it should be clear that our approach has little in common with AMG. Indeed, these methods exclusively consider Galerkin coarse grid matrices (see [33, p. 77]), that is

$$A_c = r_A A p_A = A_{22} - 2A_{21} P^{-1} A_{12} + A_{21} P^{-1} A_{11} P^{-1} A_{12} \tag{2.12}$$

This not only implies that coarse grid matrices tend to be denser, but also that only very rough approximate inverses of A_{11} can be used, to avoid prohibitive complexity in A_c . A further side effect seems then that some restrictions are put on the coarsening process, which

entail that AMG methods *require* an adaptive coarsening strategy, whereas our method has been designed with a fixed (geometric) coarsening in mind.

Choice of the smoother

Many works on multigrid methods focus on the design of efficient smoothers. Various possibilities are tested and compared in, e.g., [24]. In the present context, however, the results will depend much less critically on the numerical cost involved by the smoothing scheme, since no smoothing is performed on the finest grid. This allows one to choose a ‘robust’ smoother, without having to investigate for cheaper alternatives.

In the numerical experiments reported below, we used one pre- and one post-smoothing iteration with the relaxed ILU($\omega = -1$) smoother [9], which is robust for anisotropic problems, at least in 2D [41]. Performances are generally improved by optimizing the parameter ω , but, to keep the method simple, we did not include such refinements in our tests.

3. Theoretical analysis

First, we need to formulate the multilevel method derived from the two-level preconditioner. If ℓ levels are considered, we let the system matrix

$$A = A^{(\ell)} = \begin{pmatrix} A_{11}^{(\ell)} & A_{12}^{(\ell)} \\ A_{21}^{(\ell)} & A_{22}^{(\ell)} \end{pmatrix} \quad (3.1)$$

correspond to level ℓ , and, for $k = \ell, \dots, 2$, we partition $A^{(k)}$ similarly to $A^{(\ell)}$ and set

$$A^{(k-1)} = A_c^{(k)} \quad (3.2)$$

where $A_c^{(k)}$ is the approximation of the Schur complement of $A^{(k)}$.

Since we assume exact inversion of A_{11} except on the finest grid, the preconditioner $B^{(\ell)}$ of $A^{(\ell)}$ is then defined by the recursion

$$B^{(k)} = \begin{pmatrix} A_{11}^{(k)} & \\ A_{21}^{(k)} & M^{(k-1)-1} \end{pmatrix} \begin{pmatrix} I & A_{11}^{(k)-1} A_{12}^{(k)} \\ & I \end{pmatrix}, \quad k = 2, \dots, \ell - 1 \quad (3.3)$$

$$B^{(\ell)} = \begin{pmatrix} P^{(\ell)} & \\ A_{21}^{(\ell)} & M^{(k-1)-1} \end{pmatrix} \begin{pmatrix} I & P^{(\ell)-1} A_{12}^{(\ell)} \\ & I \end{pmatrix} \quad (3.4)$$

where

$$M^{(1)} = A^{(1)-1} \quad (3.5)$$

and where $M^{(k)}$, $k = 2, \dots, \ell - 1$, is an approximate inverse of $A^{(k)}$ defined in function of $B^{(k)-1}$.

Using $M^{(k)} = B^{(k)-1}$ leads to a simple V-cycle which is usually not optimal [38]. Previous works on algebraic multilevel preconditioners [3,6,7,10–12,36–38] consider then $M^{(k)} = \mathcal{P}_\nu(B^{(k)-1} A^{(k)}) A^{(k)-1}$ where $\mathcal{P}_\nu(t)$ is a ν degree polynomial such that $\mathcal{P}_\nu(0) = 0$ and whose remaining coefficients are determined as a function of eigenvalue estimates for

$B^{(k)-1}A^{(k)}$.

Here, we consider $M^{(k)}$ defined by

$$(I - M^{(k)} A^{(k)}) = (I - R^{(k)} A^{(k)}) (I - B^{(k)-1} A^{(k)}) (I - R^{(k)} A^{(k)}) \quad (3.6)$$

where $R^{(k)}$ is a (symmetric) relaxation operator for $A^{(k)}$, which possibly corresponds to several smoothing iterations, although in practice we found that one iteration was enough. Note that

$$x^{(k)} = M^{(k)} y^{(k)}$$

is easily computed with

$$\begin{aligned} x_{(1)}^{(k)} &= R^{(k)} y^{(k)} & ; & \quad y_{(1)}^{(k)} = y^{(k)} - A^{(k)} x_{(1)}^{(k)} \\ x_{(2)}^{(k)} &= B^{(k)-1} y_{(1)}^{(k)} & ; & \quad y_{(2)}^{(k)} = y_{(1)}^{(k)} - A^{(k)} x_{(2)}^{(k)} \\ x_{(3)}^{(k)} &= R^{(k)} y_{(2)}^{(k)} \\ x^{(k)} &= x_{(1)}^{(k)} + x_{(2)}^{(k)} + x_{(3)}^{(k)} \end{aligned}$$

The basic properties of $M^{(k)}$ are proved in the following lemma.

Lemma 3.1. *Let $M^{(k)}$ be defined by (3.6) for given symmetric positive matrices $A^{(k)}$, $B^{(k)}$, $R^{(k)}$.*

If $\rho(I - R^{(k)} A^{(k)}) \leq 1$, then $M^{(k)}$ is symmetric positive definite and satisfies

$$\lambda_{\min}(M^{(k)} A^{(k)}) \geq \min(1, \lambda_{\min}(B^{(k)-1} A^{(k)})) \quad (3.7)$$

$$\lambda_{\max}(M^{(k)} A^{(k)}) \leq \max(1, \lambda_{\max}(B^{(k)-1} A^{(k)})) \quad (3.8)$$

Proof

Letting $A^{(k)} = L^{(k)} U^{(k)}$ be a symmetric factorization of $A^{(k)}$ and defining

$$\begin{aligned} \tilde{M}^{(k)} &= U^{(k)} M^{(k)} L^{(k)} \\ \tilde{R}^{(k)} &= U^{(k)} R^{(k)} L^{(k)} \\ \tilde{B}^{(k)-1} &= U^{(k)} B^{(k)-1} L^{(k)} \end{aligned}$$

one obtains

$$\tilde{M}^{(k)} = I + (I - \tilde{R}^{(k)}) (\tilde{B}^{(k)-1} - I) (I - \tilde{R}^{(k)})$$

from which the symmetry of $M^{(k)}$ is obvious.

Further, let \mathbf{v} be any vector and set $\mathbf{w} = (I - \tilde{R}^{(k)})\mathbf{v}$. One has

$$\frac{\mathbf{v}^t \tilde{M}^{(k)} \mathbf{v}}{\mathbf{v}^t \mathbf{v}} = 1 + \frac{\mathbf{w}^t (\tilde{B}^{(k)-1} - I) \mathbf{w}}{\mathbf{w}^t \mathbf{w}} \frac{\mathbf{w}^t \mathbf{w}}{\mathbf{v}^t \mathbf{v}}$$

whereas $\rho(I - R^{(k)} A^{(k)}) \leq 1$ implies $\frac{\mathbf{w}^t \mathbf{w}}{\mathbf{v}^t \mathbf{v}} \leq 1$. Hence,

$$\lambda_{\min}(M^{(k)} A^{(k)}) = \min_{\mathbf{v}} \frac{\mathbf{v}^t \tilde{M}^{(k)} \mathbf{v}}{\mathbf{v}^t \mathbf{v}}$$

$$\geq 1 + \min \left(0, \min_{\mathbf{w}} \frac{\mathbf{w}^t (\tilde{B}^{(k)-1} - I) \mathbf{w}}{\mathbf{w}^t \mathbf{w}} \right)$$

i.e., (3.7) and

$$\begin{aligned} \lambda_{\max} \left(M^{(k)} A^{(k)} \right) &= \max_{\mathbf{v}} \frac{\mathbf{v}^t \tilde{M}^{(k)} \mathbf{v}}{\mathbf{v}^t \mathbf{v}} \\ &\leq 1 + \max \left(0, \max_{\mathbf{w}} \frac{\mathbf{w}^t (\tilde{B}^{(k)-1} - I) \mathbf{w}}{\mathbf{w}^t \mathbf{w}} \right) \end{aligned}$$

i.e., (3.8). Note that (3.7) proves also the positive definiteness of $M^{(k)}$. ■

By induction, Lemma 3.1 ensures the positive definiteness of $B^{(\ell)}$ if the splittings associated with $R^{(k)}$ are convergent for all k . Further, inequalities (3.7), (3.8) show that smoothing cannot have an adverse effect on the eigenvalue distribution.

Proving a beneficial effect is of course harder. Since

$$\lambda_{\min} \left(M^{(\ell-1)} S^{(\ell)} \right) \geq \lambda_{\min} \left(M^{(\ell-1)} A^{(\ell-1)} \right) \cdot \lambda_{\min} \left(A^{(\ell-1)-1} S^{(\ell)} \right) \quad (3.9)$$

$$\lambda_{\max} \left(M^{(\ell-1)} S^{(\ell)} \right) \leq \lambda_{\max} \left(M^{(\ell-1)} A^{(\ell-1)} \right) \cdot \lambda_{\max} \left(A^{(\ell-1)-1} S^{(\ell)} \right) \quad (3.10)$$

the convergence rate will approach that of the two-level method if $\kappa(M^{(\ell-1)} A^{(\ell-1)})$ is close to one. To analyse this, we resort to some arguments of the multigrid convergence theory. More particularly, the following theorem rewrites and adapts to our context Hackbush’s result for the V-cycle in the symmetric positive definite case [23, Lemma 7.2.1 and Theorem 7.2.2].

Theorem 3.1. *Let $M^{(k)}$, $k = 1, \dots, \ell - 1$ be defined by (3.5),(3.6), where $A^{(k)}$ and $R^{(k)}$ are symmetric positive definite matrices and where $B^{(k)}$, $k = 2, \dots, \ell - 1$ is defined by (3.3) for some 2×2 partitioning of $A^{(k)}$. Assume that*

$$\lambda_{\max} \left(R^{(k)} A^{(k)} \right) \leq 1 \quad k = 2, \dots, \ell - 1 \quad (3.11)$$

If, for some positive constants c_A , C_A such that $c_A < 2$, one has, for $k = 2, \dots, \ell - 1$,

$$-c_A \mathbf{v}^t R^{(k)} \mathbf{v} \leq \mathbf{v}^t \left(A^{(k)-1} - B_{(2)}^{(k)-1} \right) \mathbf{v} \leq C_A \mathbf{v}^t R^{(k)} \mathbf{v} \quad \forall \mathbf{v} \quad (3.12)$$

where

$$B_{(2)}^{(k)} = \begin{pmatrix} A_{11}^{(k)} & \\ A_{21}^{(k)} & A^{(k-1)} \end{pmatrix} \begin{pmatrix} I & A_{11}^{(k)-1} A_{12}^{(k)} \\ & I \end{pmatrix} \quad (3.13)$$

then, for $k = 1, \dots, \ell - 1$,

$$\lambda_{\min} \left(M^{(k)} A^{(k)} \right) \geq \frac{2}{2 + C_A} \quad (3.14)$$

$$\lambda_{\max} \left(M^{(k)} A^{(k)} \right) \leq \frac{2}{2 - c_A} \quad (3.15)$$

Proof

Obviously, (3.14), (3.15) hold for $k = 1$ and we will show by induction that they hold for any k . In this view, let

$$U^{(k)} = \begin{pmatrix} A_{11}^{(k)\frac{1}{2}} & \\ & S^{(k)\frac{1}{2}} \end{pmatrix} \begin{pmatrix} I & A_{11}^{(k)-1} A_{12}^{(k)} \\ & I \end{pmatrix}$$

where $S^{(k)}$ is the Schur complement of $A^{(k)}$. Note that $A^{(k)} = L^{(k)} U^{(k)}$ where $L^{(k)} = U^{(k)t}$. Let further

$$\begin{aligned} \tilde{M}^{(k)} &= U^{(k)} M^{(k)} L^{(k)} \\ \tilde{R}^{(k)} &= U^{(k)} R^{(k)} L^{(k)} \\ \tilde{B}^{(k)-1} &= U^{(k)} B^{(k)-1} L^{(k)} \\ \tilde{B}_{(2)}^{(k)-1} &= U^{(k)} B_{(2)}^{(k)-1} L^{(k)} \end{aligned}$$

One has

$$(I - \tilde{M}^{(k)}) = (I - \tilde{R}^{(k)}) (I - \tilde{B}^{(k)-1}) (I - \tilde{R}^{(k)})$$

On the other hand,

$$\begin{aligned} \tilde{B}^{(k)-1} &= \begin{pmatrix} I & \\ & S^{(k)\frac{1}{2}} M^{(k-1)} S^{(k)\frac{1}{2}} \end{pmatrix} \\ \tilde{B}_{(2)}^{(k)-1} &= \begin{pmatrix} I & \\ & S^{(k)\frac{1}{2}} A^{(k-1)-1} S^{(k)\frac{1}{2}} \end{pmatrix} \end{aligned}$$

whence, by the induction argument,

$$\frac{2}{2+c_A} \mathbf{v}^t \tilde{B}_{(2)}^{(k)-1} \mathbf{v} \leq \mathbf{v}^t \tilde{B}^{(k)-1} \mathbf{v} \leq \frac{2}{2-c_A} \mathbf{v}^t \tilde{B}_{(2)}^{(k)-1} \mathbf{v} \quad \forall \mathbf{v} \quad (3.16)$$

The right inequality implies

$$\mathbf{v}^t (I - \tilde{M}^{(k)}) \mathbf{v} \geq \mathbf{v}^t (I - \tilde{R}^{(k)}) \left(I - \frac{2}{2-c_A} \tilde{B}_{(2)}^{(k)-1} \right) (I - \tilde{R}^{(k)}) \mathbf{v} \quad \forall \mathbf{v}$$

which gives, using the left inequality (3.12),

$$\mathbf{v}^t (I - \tilde{M}^{(k)}) \mathbf{v} \geq \mathbf{v}^t (I - \tilde{R}^{(k)}) \left(\frac{-c_A}{2-c_A} I - \frac{2c_A}{2-c_A} \tilde{R}^{(k)} \right) (I - \tilde{R}^{(k)}) \mathbf{v} \quad \forall \mathbf{v}$$

Therefore,

$$\begin{aligned} \lambda_{\max} (M^{(k)} A^{(k)}) &= \lambda_{\max} (\tilde{M}^{(k)}) \\ &\leq 1 + \frac{c_A}{2 - c_A} \max_{0 \leq \xi \leq 1} (1 - \xi)^2 (1 + 2\xi) \end{aligned}$$

whence (3.15) since the maximum corresponds to $\xi = 0$.

Similarly, the left inequality (3.16) implies

$$\mathbf{v}^t \left(I - \tilde{M}^{(k)} \right) \mathbf{v} \leq \mathbf{v}^t \left(I - \tilde{R}^{(k)} \right) \left(I - \frac{2}{2+c_A} \tilde{B}_{(2)}^{(k)-1} \right) \left(I - \tilde{R}^{(k)} \right) \mathbf{v} \quad \forall \mathbf{v}$$

which gives, using the right inequality (3.12),

$$\mathbf{v}^t \left(I - \tilde{M}^{(k)} \right) \mathbf{v} \leq \mathbf{v}^t \left(I - \tilde{R}^{(k)} \right) \left(\frac{c_A}{2+c_A} I + \frac{2c_A}{2+c_A} \tilde{R}^{(k)} \right) \left(I - \tilde{R}^{(k)} \right) \mathbf{v} \quad \forall \mathbf{v}$$

whence

$$\begin{aligned} \lambda_{\min} \left(M^{(k)} A^{(k)} \right) &= \lambda_{\min} \left(\tilde{M}^{(k)} \right) \\ &\geq 1 - \frac{C_A}{2 + C_A} \max_{0 \leq \xi \leq 1} (1 - \xi)^2 (1 + 2\xi) \end{aligned}$$

i.e., (3.14). ■

For the method defined in the hierarchical basis with $A^{(k-1)} = A_{22}^{(hb)^{(k)}}$, it is easy to show that $\lambda_{\max}(A^{(k-1)-1} S^{(k)}) \leq 1$ [36], whence $c_A = 0$ in (3.12). For 2D five-point finite difference matrices, it is proved in [28] that $\lambda_{\max}(\tilde{S}^{(k)-1} S^{(k)}) \leq 2$ when the PDE coefficients are piecewise constant on the coarsest mesh, whence again $c_A = 0$ when using the rule (2.11).

The analysis of C_A is harder. In fact, the multigrid convergence theory developed in [23] for the symmetric positive definite case makes use of the following similar ‘approximation property’:

$$\mathbf{v}^t \left(A^{(k)-1} - p^{(k)} A^{(k-1)-1} r^{(k)} \right) \mathbf{v} \leq \bar{C}_A \mathbf{v}^t R^{(k)} \mathbf{v} \quad \forall \mathbf{v} \quad (3.17)$$

General analyses of \bar{C}_A are rather involved and extending them to our context lies beyond the scope of the present paper. However, it is possible to rest on the developments in [23] for the proof of the approximation property and derive simple upper bounds for C_A assuming that (3.17) holds. This is the purpose of the the following theorem and lemma, where we address the case of damped Jacobi smoothing ($R^{(k)-1} = \omega \text{diag}(A^{(k)})$).

Of course, this approach does not tell us anything for problems that do not fit well in classical multigrid theory because standard multigrid algorithms tend to perform poorly. In such cases, our method may still be successful since algebraic multilevel preconditioners (and our main theoretical assumption (3.12)) depend only on the coarse grid matrix and do not refer to any interpolation procedure, whereas works that develop more robust multigrid algorithms (e.g., [1,19,20]) stress that the problems of standard multigrid are due to an inappropriate interpolation rather than to a bad coarse grid approximation. Referring to Lemma 3.2 below, this means that (3.19) may hold in cases for which (3.18) does not hold or holds only for a prolongation $p^{(k)}$ that is not known a priori or impractical to use.

Lemma 3.2. *Let $A^{(k)}$ and $A^{(k-1)}$ be symmetric positive definite matrices. Let $D^{(k)} = \text{diag}(A^{(k)})$ and $D_2^{(k)} = \text{diag}(A_{22}^{(k)})$.*

If, for some $p^{(k)}$ of the form (2.7),

$$\mathbf{v}^t \left(A^{(k)-1} - p^{(k)} A^{(k-1)-1} r^{(k)} \right) \mathbf{v} \leq \bar{C}_A \omega^{-1} \mathbf{v}^t D^{(k)-1} \mathbf{v} \quad \forall \mathbf{v} \quad (3.18)$$

where $r^{(k)} = p^{(k)t}$, then

$$\mathbf{v}_2^t \left(S^{(k)-1} - A^{(k-1)-1} \right) \mathbf{v}_2 \leq \bar{C}_A \omega^{-1} \mathbf{v}_2^t D_2^{(k)-1} \mathbf{v}_2 \quad \forall \mathbf{v}_2 \quad (3.19)$$

where $S^{(k)}$ is the Schur complement of $A^{(k)}$ related to the partitioning induced by $p^{(k)}$.

Proof

(3.19) is (3.18) restricted to vectors of the form $\mathbf{v} = \begin{pmatrix} 0 \\ \mathbf{v}_2 \end{pmatrix}$. ■

Theorem 3.2. Let $A^{(k)}$ be a symmetric positive definite matrix partitioned in a 2×2 block form, let $S^{(k)}$ be the corresponding Schur complement, and let $D^{(k)} = \text{diag}(A^{(k)})$, $D_1^{(k)} = \text{diag}(A_{11}^{(k)})$, $D_2^{(k)} = \text{diag}(A_{22}^{(k)})$. Let $B_{(2)}^{(k)}$ be defined by (3.13), where $A^{(k-1)}$ is symmetric positive definite.

If

$$\mathbf{v}_2^t \left(S^{(k)-1} - A^{(k-1)-1} \right) \mathbf{v}_2 \leq \omega^{-1} \widehat{C}_A \mathbf{v}_2^t D_2^{(k)-1} \mathbf{v}_2 \quad \forall \mathbf{v}_2 \quad (3.20)$$

then

$$\mathbf{v}^t \left(A^{(k)-1} - B_{(2)}^{(k)-1} \right) \mathbf{v} \leq \omega^{-1} C_A \mathbf{v}^t D^{(k)-1} \mathbf{v} \quad \forall \mathbf{v} \quad (3.21)$$

where

$$C_A \leq \widehat{C}_A \left(1 + \lambda_{\min}^{-1} \left(D_1^{(k)-1} A_{11}^{(k)} \right) \lambda_{\min}^{-1} \left(D_2^{(k)-1} A_{22}^{(k)} \right) \right) \quad (3.22)$$

Proof

Define $U^{(k)}$, $L^{(k)}$, $\widetilde{B}_{(2)}^{(k)}$ as in the proof of Theorem 3.1. For any vector \mathbf{v} and positive number α ,

$$\begin{aligned} \mathbf{v}^t \left(I - \widetilde{B}_{(2)}^{(k)-1} \right) \mathbf{v} &= \mathbf{v}^t \begin{pmatrix} 0 & 0 \\ 0 & I - S^{(k)\frac{1}{2}} A^{(k-1)-1} S^{(k)\frac{1}{2}} \end{pmatrix} \mathbf{v} \\ &\leq \widehat{C}_A \omega^{-1} \mathbf{v}^t \begin{pmatrix} 0 & 0 \\ 0 & S^{(k)\frac{1}{2}} D_2^{(k)-1} S^{(k)\frac{1}{2}} \end{pmatrix} \mathbf{v} \\ &\leq \widehat{C}_A \omega^{-1} \mathbf{v}^t \begin{pmatrix} \alpha^{-1} I & 0 \\ 0 & S^{(k)\frac{1}{2}} D_2^{(k)-1} S^{(k)\frac{1}{2}} \end{pmatrix} \mathbf{v} \\ &= \widehat{C}_A \omega^{-1} \mathbf{v}^t U^{(k)} G_\alpha^{-1} L^{(k)} \mathbf{v}, \end{aligned}$$

where

$$\begin{aligned} G_\alpha &= L^{(k)} \begin{pmatrix} \alpha I & \\ & S^{(k)-\frac{1}{2}} D_2^{(k)} S^{(k)-\frac{1}{2}} \end{pmatrix} U^{(k)} \\ &= \begin{pmatrix} \alpha A_{11}^{(k)} & \alpha A_{12}^{(k)} \\ \alpha A_{21}^{(k)} & D_2^{(k)} + \alpha A_{21}^{(k)} A_{11}^{(k)-1} A_{12}^{(k)} \end{pmatrix} \end{aligned}$$

Therefore, (3.21) holds with $C_A \leq \nu^{-1} \widehat{C}_A$ if

$$\mathbf{v}^t G_\alpha^{-1} \mathbf{v} \leq \nu^{-1} \mathbf{v}^t D^{(k)-1} \mathbf{v} \quad \forall \mathbf{v}$$

or, equivalently, if $G_\alpha - \nu D^{(k)}$ is non-negative definite. Since $\alpha A_{11}^{(k)} - \nu D_1^{(k)}$ is positive definite for α sufficiently large, $G_\alpha - \nu D^{(k)}$ will be non-negative definite if and only if so is its Schur complement

$$S_{G_\alpha - \nu D^{(k)}} = (1 - \nu) D_2^{(k)} + A_{21}^{(k)} \left(\alpha A_{11}^{(k)-1} - \alpha^2 \left(\alpha A_{11}^{(k)} - \nu D_1^{(k)} \right)^{-1} \right) A_{12}^{(k)}$$

On the other hand, letting $\lambda_2 = \lambda_{\min}(D_2^{(k)-1} A_{22}^{(k)})$,

$$\mathbf{v}_2^t D_2^{(k)} \mathbf{v}_2 \geq \lambda_2 \mathbf{v}_2^t A_{22}^{(k)} \mathbf{v}_2 \geq \lambda_2 \mathbf{v}_2^t A_{21}^{(k)} A_{11}^{(k)-1} A_{12}^{(k)} \mathbf{v}_2 \quad \forall \mathbf{v}_2$$

Hence, $S_{G_\alpha - \nu D^{(k)}}$ is non-negative definite if

$$((1-\nu)\lambda_2 + \alpha) \mathbf{v}_1^t A_{11}^{(k)-1} \mathbf{v}_1 \geq \alpha^2 \mathbf{v}_1^t \left(\alpha A_{11}^{(k)} - \nu D_1^{(k)} \right)^{-1} \mathbf{v}_1 \quad \forall \mathbf{v}_1$$

or, equivalently, if

$$((1-\nu)\lambda_2 + \alpha) \left(\alpha A_{11}^{(k)} - \nu D_1^{(k)} \right) - \alpha^2 A_{11}^{(k)} = \alpha \left((1-\nu)\lambda_2 A_{11}^{(k)} - \nu D_1^{(k)} \right) - \nu(1-\nu)\lambda_2 D_1^{(k)}$$

is non-negative definite. Setting

$$\nu = \frac{\lambda_1 \lambda_2}{1 + \lambda_1 \lambda_2} - \frac{\lambda_2}{\alpha}$$

where $\lambda_1 = \lambda_{\min}(D_1^{(k)-1} A_{11}^{(k)})$, the latter matrix writes

$$\frac{\alpha \lambda_2}{1 + \lambda_1 \lambda_2} \left(A_{11}^{(k)} - \lambda_1 D_1^{(k)} \right) + \lambda_2^2 A_{11}^{(k)} + \lambda_2(1-\nu(1-\nu)) D_1^{(k)}$$

which is clearly non-negative definite since $\max_t t(1-t) = \frac{1}{2}$. Therefore,

$$C_A \leq \nu^{-1} \widehat{C}_A = \left(\frac{\lambda_1 \lambda_2}{1 + \lambda_1 \lambda_2} - \frac{\lambda_2}{\alpha} \right)^{-1} \widehat{C}_A$$

and the proof is completed by taking the limit for α going to the infinity. ■

Note that if $A_{22}^{(k)}$ is diagonal, then $\lambda_{\min}(D_2^{(k)-1} A_{22}^{(k)}) = 1$. On the other hand, $A_{11}^{(k)}$ has condition number $\mathcal{O}(1)$ [2,8,14]. In practice, $\lambda_{\min}(D_1^{(k)-1} A_{11}^{(k)})$ is unacceptably small when the problem is strongly anisotropic, but this shortcoming has to be related to the fact that damped Jacobi smoothing performs poorly anyway in such cases.

Model problem analysis

Consider the model problem corresponding to the five-point finite difference discretization of the 2D Laplacian on the unit square with Dirichlet boundary conditions and uniform mesh size $h = 2^{-\ell}$ in both directions. Let the nodes for level k correspond to the uniform grid of mesh size $h_k = 2^{-k}$, and let $A^{(k)}$, $k = \ell - 1, \dots, 1$ be defined by the rule (2.11), or, equivalently, be the Galerkin matrix corresponding to linear interpolation. As noted above,

$\lambda_{\max}(A^{(k-1)^{-1}}S^{(k)}) \leq 1$, i.e., $c_A = 0$ in (3.12).

On the other hand, $S^{(k)}$ and $A^{(k-1)}$ have as common set of eigenvectors the discrete sine functions $\sin \mu\pi x h_k \sin \nu\pi y h_k$, $\mu = 1, \dots, \frac{h_k^{-1}}{2} - 1$, $\nu = 1, \dots, \frac{h_k^{-1}}{2} - 1$. The corresponding eigenvalues write (see [30])

$$\begin{aligned} S^{(\mu\nu)} &= 16 \left(\frac{1}{1 - \frac{c_\mu}{2} - \frac{c_\nu}{2}} + \frac{1}{1 + \frac{c_\mu}{2} - \frac{c_\nu}{2}} + \frac{1}{1 - \frac{c_\mu}{2} + \frac{c_\nu}{2}} + \frac{1}{1 + \frac{c_\mu}{2} + \frac{c_\nu}{2}} \right)^{-1} \\ A_c^{(\mu\nu)} &= 4 - 2c_{2\mu} - 2c_{2\nu} \end{aligned}$$

where $c_\mu = \cos \mu\pi h_k$, $c_\nu = \cos \nu\pi h_k$, $c_{2\mu} = \cos 2\mu\pi h_k$, $c_{2\nu} = \cos 2\nu\pi h_k$.

Since $\text{diag}(A^{(k)}) = 4I$, the constant \widehat{C}_A in (3.20) satisfies

$$\widehat{C}_A = 4\omega \max_{\mu,\nu} \left(S^{(\mu\nu)^{-1}} - A_c^{(\mu\nu)^{-1}} \right)$$

and can therefore be analysed directly. Using $c_{2\mu} = 2c_\mu^2 - 1$, $c_{2\nu} = 2c_\nu^2 - 1$, and letting $c_\mu^2 = s_\mu$, $c_\nu^2 = s_\nu$, tedious manipulation (easier with computer algebra) show that

$$\begin{aligned} 4 \left(S^{(\mu\nu)^{-1}} - A_c^{(\mu\nu)^{-1}} \right) &= \frac{16(1 - c_\mu^2 - c_\nu^2) + 3(c_\mu^4 + c_\nu^4) + 10c_\mu^2 c_\nu^2}{\left(4 - (c_\mu + c_\nu)^2\right) \left(4 - (c_\mu - c_\nu)^2\right) (2 - c_\mu^2 - c_\nu^2)} \\ &= \frac{3s_\mu^2 + 10s_\mu s_\nu + 3s_\nu^2}{(s_\mu + s_\nu) (8(s_\mu + s_\nu) + (s_\mu - s_\nu)^2)} \\ &= \frac{\alpha^{(\mu\nu)}}{\beta^{(\mu\nu)}} \end{aligned}$$

where $\alpha^{(\mu\nu)} = 3s_\mu^2 + 10s_\mu s_\nu + 3s_\nu^2$, $\beta^{(\mu\nu)} = (s_\mu + s_\nu) (8(s_\mu + s_\nu) + (s_\mu - s_\nu)^2)$. Since

$$\beta^{(\mu\nu)} - 2\alpha^{(\mu\nu)} = (s_\mu - s_\nu)^2 (2 + s_\mu + s_\nu)$$

is always non-negative, we deduce

$$\widehat{C}_A \leq \frac{\omega}{2} \tag{3.23}$$

On the other hand, letting $\mathbf{x}_1^{(k)}$ be the positive vector defined on the fine grid nodes at level k by

$$\left(\mathbf{x}_1^{(k)} \right)_i = \begin{cases} \sqrt{2} & \text{if } i \text{ is centred in a coarse grid square} \\ 1 & \text{otherwise} \end{cases}$$

it is easy to see that

$$\left(-\text{offdiag} \left(A_{11}^{(k)} \right) \right) \mathbf{x}_1^{(k)} \leq 2\sqrt{2} \mathbf{x}_1^{(k)}$$

Hence, $\rho(\text{offdiag}(A_{11}^{(k)})) \leq 2\sqrt{2}$ and therefore $\lambda_{\min}(D_1^{(k)^{-1}}A_{11}^{(k)}) = \frac{1}{4}\lambda_{\min}(A_{11}^{(k)}) \geq 1 - \frac{\sqrt{2}}{2}$. Theorem 3.2 implies then

$$C_A \leq \frac{3 + \sqrt{2}}{2} \omega \tag{3.24}$$

To satisfy $\lambda_{\max}(R^{(k)}A^{(k)}) \leq 1$, $\omega = 2$ is the proper choice. Since $\lambda_{\min}(A^{(\ell-1)^{-1}}S^{(\ell)}) \geq \frac{1}{2}$ and $\lambda_{\max}(A^{(\ell-1)^{-1}}S^{(\ell)}) \leq 1$ [25], Theorem 3.1 and (3.9), (3.10) finally yield

$$\lambda_{\min}\left(M^{(\ell-1)^{-1}}S^{(\ell)}\right) \geq \frac{1}{5 + \sqrt{2}} \tag{3.25}$$

$$\lambda_{\max}\left(M^{(\ell-1)^{-1}}S^{(\ell)}\right) \leq 1 \tag{3.26}$$

independently of the number of levels. If subsystems with $A_{11}^{(\ell)}$ are solved exactly too, these bounds directly apply to the extremal eigenvalues of $B^{(\ell)^{-1}}A^{(\ell)}$ (see 2.3), whereas the results in [29] prove that they deteriorate only slightly when $P^{(\ell)}$ is a MILU factorization of $A_{11}^{(\ell)}$.

4. Numerical results

We have tested our multilevel preconditioner on linear systems resulting from the finite difference discretization of

$$-\partial_x a_x \partial_x u - \partial_y a_y \partial_y u - \partial_z a_z \partial_z u = f \quad \text{in } \Omega$$

$$\begin{cases} u = 0 & \text{on } \Gamma_0 \subset \partial \Omega \\ \frac{\partial u}{\partial n} = 0 & \text{on } \Gamma_1 = \partial \Omega \setminus \Gamma_0 \end{cases}$$

where Ω is the unit square (2D) or unit cube (3D). We use the point mesh box integration scheme [26] with, except on Problem 3D1, uniform mesh size h in each direction.

In all experiments, we ran the preconditioned conjugate gradient algorithm with the zero vector as initial approximation, and the reported the number of iterations. is the one needed to reduce the relative residual error $\frac{\|r_k\|}{\|b\|}$ below 10^{-8} . The reported CPU times are given in seconds and include preprocessing, that is, the operations needed to compute the successive incomplete factorizations of A_{11} , approximate Schur complements and relaxation operators. No particular optimization effort was made, so that the timing results cannot pretend to be optimal and should be used only as relative criteria. Whenever the CPU time is not reported, this is because, for technical reasons, the corresponding test was run with a program using a completely different data structure, so that a comparison would be meaningless. The considered machine is a SUN SPARC20/514 workstation with 128 Mb of RAM memory.

In the tables, *AML* refers to algebraic multilevel preconditioners as presented in Section 2, with, at every level, A_{11} approximated by a MILU factorization without fill-in, considering a natural ordering of the concerned nodes. A_c was defined by the rule (2.11) at levels where smoothing is performed, whereas we set $A_c = \tilde{S}$ at level(s) where smoothing is not performed, further numerical tests showing that this is slightly better than using the rule (2.11) everywhere.

Besides *AML-V-cycle with smoothing*, which refers to the method discussed throughout this paper, we included three different variants for comparison purposes: *AML-V-cycle (no smoothing)* is the same method with smoothing iterations skipped; *AML-W-cycle (no smoothing)* is the method proposed in [28] for 2D problems that uses a W-cycle with parameters fixed according eigenvalue bounds proved assuming that the PDE coefficients

Table 1. Results for Problem 2D1

Preconditioner	λ_{\min}	λ_{\max}	κ	Number of iterations	CPU
$h^{-1} = 128$					
AML-V-cycle with smoothing	1.02	2.54	2.50	14	2.42
AML-W-cycle (no smoothing)	0.86	2.00	2.31	14	
AML-V-cycle (no smoothing)	1.01	8.53	8.42	32	
AML-two level	1.02	2.11	2.07	14	
ILU	1E-6	1.21	9E+5	196	18.0
MILU	1.00	371	371	150	13.8
DRILU	2E-3	46.4	2E+4	87	7.96
$h^{-1} = 512$					
AML-V-cycle with smoothing	1.02	2.54	2.50	15	73.3
AML-W-cycle (no smoothing)	0.86	2.00	2.33	15	
AML-V-cycle (no smoothing)	1.02	17.5	17.1	48	
DRILU	2E-3	193	8E+4	192	427

are piecewise constant on the coarsest mesh; *AML-two level* corresponds to the reference two-level scheme (2.8), with thus the first coarse grid matrix inverted exactly.

We always considered standard coarsening and, for *AML-V-cycle with smoothing*, performed full reduction, leaving only one node in the coarsest level, this node being the first one in the natural ordering of the fine grid. For *AML-V-cycle (no smoothing)*, to limit the effect of the dependency of the number of levels on the condition number, the coarsest grid considered was 16×16 (2D, $h^{-1} = 128$), 32×32 (2D, $h^{-1} = 512$) or $5 \times 5 \times 5$ (3D), these grid sizes being limited on account of the involved factorization cost. For *AML-W-cycle (no smoothing)*, to get exactly the same method as the one tested in [28], $A^{(k)}$ was factorized exactly as soon as $n_k \leq 81$ was reached.

We also compared the multilevel variants with some reference incomplete factorization preconditioners: *ILU* is the standard scheme without any fill-in and *MILU* is the ‘modified’ version in which discarded fill entries are lumped to the main diagonal [4,16]; *DRILU* is the perturbed variant of MILU introduced in [27].

Problem 2D1

Here, Ω is the unit square and

$$\begin{aligned}
 a_x &= a_y = \begin{cases} 10^3 & \text{in } \left(\frac{1}{4}, \frac{3}{4}\right) \times \left(\frac{1}{4}, \frac{3}{4}\right) \\ 1 & \text{elsewhere} \end{cases} \\
 f &= \begin{cases} 1 & \text{in } \left(\frac{1}{4}, \frac{3}{4}\right) \times \left(\frac{1}{4}, \frac{3}{4}\right) \\ 0 & \text{elsewhere} \end{cases} \\
 \Gamma_0 &= \{(x, y) \mid 0 \leq x \leq 1, y = 1\}
 \end{aligned}$$

Results are reported in Table 1. Clearly, using smoothing is successful in stabilizing the condition number associated with the V-cycle. The latter is independent of the mesh size and approaches that of the corresponding two-level method. On the other hand, the W-cycle without smoothing confirms its good behaviour [28], whereas the V-cycle without smooth-

Table 2. Results for Problem 2D2 with AML–V-cycle with smoothing

h^{-1}	λ_{\min}	λ_{\max}	κ	Number of iterations	CPU
100	1.01	3.81	3.78	18	1.84
400	1.01	3.19	3.16	17	38.6
500	1.01	4.26	4.22	18	69.1

ing is far from being competitive even though the number of levels was limited by factorizing exactly a rather large coarse grid matrix. ILU-type preconditioners are as expected not optimal, further tests showing that the best variant (DRILU) becomes competitive with our multilevel method when h^{-1} is not larger than 16.

Problem 2D2

Here, Ω is the unit square and

$$\begin{aligned}
 a_x &= 100, a_y = 1, f = 0 && \text{in } (0.65, 0.95) \times (0.05, 0.65) \\
 a_x &= 1, a_y = 100, f = 0 && \text{in } (0.25, 0.45) \times (0.25, 0.45) \\
 a_x &= 100, a_y = 100, f = 1 && \text{in } (0.05, 0.25) \times (0.65, 0.95) \\
 a_x &= 1, a_y = 1, f = 0 && \text{elsewhere} \\
 \Gamma_0 &= \{(x, y) \mid 0 \leq x \leq 1, y = 1\}
 \end{aligned}$$

i.e., the problem has jumps and anisotropy in both directions. Further, when h^{-1} is an odd multiple of 20 (e.g., $h^{-1} = 100$ or $h^{-1} = 500$), one may check that *none of the discontinuity lines is aligned with the coarse grid generated by the first reduction*.

For this problem, the W-cycle method from [28] fails to produce a positive definite preconditioner. This illustrates a shortcoming of versions using parameters predetermined in function of theoretical estimates, whose use is dangerous as soon as the assumptions needed in the proofs of these estimates are not met anymore.

On the other hand, the results for the V-cycle method with smoothing are reported in Table 2. Compared with Table 1, one observes only a slight deterioration of the performances, despite this the problem would be very difficult to solve with a standard multigrid algorithm.

Problem 3D1

Here, Ω is the unit cube and

$$\begin{aligned}
 a_x &= a_y = a_z = \begin{cases} 10^3 & \text{in } \left(\frac{1}{4}, \frac{3}{4}\right) \times \left(\frac{1}{4}, \frac{3}{4}\right) \times \left(\frac{1}{4}, \frac{3}{4}\right) \\ 1 & \text{elsewhere} \end{cases} \\
 f &= \begin{cases} 1 & \text{in } \left(\frac{1}{4}, \frac{3}{4}\right) \times \left(\frac{1}{4}, \frac{3}{4}\right) \times \left(\frac{1}{4}, \frac{3}{4}\right) \\ 0 & \text{elsewhere} \end{cases} \\
 \Gamma_0 &= \{(x, y) \mid 0 \leq x, y \leq 1, z = 0 \text{ or } z = 1\}
 \end{aligned}$$

Table 3. Results for Problem 3D1

preconditioner	λ_{\min}	λ_{\max}	κ	Number of iterations	CPU
$h^{-1} = 40$					
AML-V-cycle with smoothing	1.01	7.39	7.34	23	21.7
AML-V-cycle (no smoothing)	1.01	110	109	79	
ILU	1E-5	1.12	83 599	100	56.3
MILU	1.00	13 304	13 302	255	143
DRILU	3E-3	19.4	677	63	35.6
$h^{-1} = 80$					
AML-V-cycle with smoothing	1.01	7.24	7.20	24	241
DRILU	3E-2	41.9	1531	96	579
non uniform $68 \times 68 \times 68$ mesh					
AML-V cycle with smoothing	1.01	12.4	12.3	33	208
DRILU	1E-3	27.7	1911	88	342

The problem thus presents a strong discontinuity. Further, since more often than in 2D, simulations make use of varying mesh sizes, we also investigated the effect of moderate anisotropy induced by grid stretching. In this view, besides standard uniform grids, we considered a $68 \times 68 \times 68$ mesh which is structurally regular but refined near the interfaces between different materials, the mesh sizes $h_x = h(x)$, $h_y = h(y)$, $h_z = h(z)$ varying according to

$$h(t) = \begin{cases} 0.025 & \text{if } 0.00 \leq t \leq 0.20 \text{ or } 0.30 \leq t \leq 0.50 \\ 0.010 & \text{if } 0.20 \leq t \leq 0.21, \text{ or } 0.29 \leq t \leq 0.30 \\ 0.005 & \text{if } 0.21 \leq t \leq 0.29 \end{cases}$$

with $h(t) = h(1 - t)$ for $0.5 \leq t \leq 1$. In this case, one may also check that *none of the material interfaces is aligned with the (pseudo) coarse grid generated by the first reduction.*

The results are reported in Table 3. As in 2D, our approach to stabilize the V-cycle appears successful and ILU type preconditioners are outmatched even though 3D meshes are necessarily less refined than 2D ones (further tests show here again that DRILU becomes competitive when $h^{-1} \leq 16$). On the other hand, performances remain satisfactory on the non uniform mesh despite the anisotropy induced by the grid stretching [17].

Note that we deliberately omitted 3D tests with strongly anisotropic coefficients because the kind of smoother considered here cannot pretend to be efficient in such cases, and alternative approaches resorting to plane relaxation or semi-coarsening should be considered instead (see, e.g., [24]).

REFERENCES

1. R. E. Alcouffe, A. Brandt, J. E. Dendy and J. W. Painter. The multi-grid methods for the diffusion equation with strongly discontinuous coefficients. *SIAM J. Sci. Stat. Comput.*, 2, 430–454, 1981.
2. O. Axelsson. On multigrid methods of the two-level type. In W. Hackbusch and U. Trottenberg, editors, *Multigrid Methods*, Lectures Notes in Mathematics, pages 352–367, Springer-Verlag, Berlin, Heidelberg, New York, 1981.
3. O. Axelsson. The method of diagonal compensation of reduced matrix entries and multilevel iteration. *J. Comput. Appl. Math.*, 38, 31–43, 1991.

4. O. Axelsson. *Iterative Solution Methods*, University Press, Cambridge, 1994.
5. O. Axelsson and V. A. Barker. *Finite Element Solution of Boundary Value Problems. Theory and Computation*, Academic Press, New York, 1984.
6. O. Axelsson and V. Eijkhout. Analysis of recursive 5-point/9-point factorization method. In O. Axelsson and L. Kolotilina, editors, *Preconditioned Conjugate Gradient Methods*, Vol. 1457 of *Lecture Notes in Mathematics* pages 154–173. Springer-Verlag, Berlin, 1990.
7. O. Axelsson and V. Eijkhout. The nested recursive two level factorization for nine-point difference matrices. *SIAM J. Sci. Stat. Comput.*, 12, 1373–1400, 1991.
8. O. Axelsson and I. Gustafsson. Preconditioning and two-level multigrid methods of arbitrary degree of approximation. *Math. Comp.*, 40, 214–242, 1983.
9. O. Axelsson and G. Lindskog. On the eigenvalue distribution of a class of preconditioning methods. *Numer. Math.*, 48, 479–498, 1986.
10. O. Axelsson and M. Neytcheva. Algebraic multilevel iterations for Stieltjes matrices. *Num. Lin. Alg. Appl.*, 1, 213–236, 1994.
11. O. Axelsson and P. S. Vassilevski. Algebraic multilevel preconditioning methods. i. *Numer. Math.*, 56, 157–177, 1989.
12. O. Axelsson and P. S. Vassilevski. Algebraic multilevel preconditioning methods, II. *SIAM J. Numer. Anal.*, 27, 1569–1590, 1990.
13. R. E. Bank. Hierarchical bases and the finite element method. *Acta Numerica*, 5, 1–43, 1996.
14. R. E. Bank and T. F. Dupont. 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, USA, 1980.
15. R. E. Bank, T. F. Dupont and H. Yserentant. The hierarchical basis multigrid method. *Numer. Math.*, 52, 427–458, 1988.
16. R. Barrett, M. Berry, T. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine and H. van der Vorst. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, SIAM Publications, Philadelphia, 1993.
17. E. F. F. Botta and F. W. Wubs. The convergence behaviour of iterative methods on severely stretched grids. *Int. j. numer. methods eng.*, 36, 3333–3350, 1993.
18. D. Braess. Towards algebraic multigrid for elliptic problems of second order. *Computing*, 55, 379–393, 1995.
19. P. M. de Zeeuw. Matrix-dependent prolongations and restrictions in a blackbox multigrid solver. *J. Comput. Appl. Math.*, 33, 1–27, 1990.
20. J. E. Dendy. Black box multigrid. *J. Comput. Phys.*, 48, 366–386, 1982.
21. J. Fuhrman. Outlines of a modular algebraic multilevel method. In *AMLI'96: Proceedings of the Conference on Algebraic Multilevel Iteration Methods with Applications*, volume 1, pages 141–152, University of Nijmegen, Nijmegen, The Netherlands, 1996.
22. G. H. Golub and C. F. van Loan. *Matrix Computations*, third edition. The John Hopkins University Press, Baltimore, Maryland, 1996.
23. W. Hackbusch. *Multi-grid Methods and Applications*. Springer, Berlin, 1985.
24. R. Kettler. Linear multigrid methods for numerical reservoir simulation. Ph.D. thesis, Technical University Delft, Delft, 1987.
25. J. Maitre and F. Musy. The contraction number of a class of two-level methods; an exact evaluation for some finite element subspaces and model problems. In W. Hackbusch and U. Trottenberg, editors, *Multigrid Methods, Lectures Notes in Mathematics*, pages 535–544, Springer-Verlag, Berlin, Heidelberg, New York, 1981.
26. S. Nakamura. *Computational Methods in Engineering and Science*, John Wiley & Sons, New York, 1977.
27. Y. Notay. DRIC: a dynamic version of the RIC method. *J. Num. Lin. Alg. with Appl.*, 1, 511–532, 1994.
28. Y. Notay. Optimal order preconditioning of finite difference matrices. Technical Report GANMN 97-02, Université Libre de Bruxelles, Brussels, Belgium, 1997. Available at <http://homepages.ulb.ac.be/~ynotay>.
29. Y. Notay. Using approximate inverses in algebraic multilevel methods. *Numer. Math.*, 80, 397–417, 1998.
30. A. Reusken. Fourier analysis of a robust multigrid method for convection-diffusion equations. *Numer. Math.*, 71, 365–398, 1995.

31. A. Reusken. A multigrid method based on incomplete Gaussian elimination. *J. Num. Lin. Alg. with Appl.*, 3, 369–390, 1996.
32. J. W. Ruge and K. Stüben. Efficient solution of finite difference and finite element equations by algebraic multigrid (AMG). In D. J. Paddon and H. Holstein, editors, *Multigrid Methods for Integral and Differential Equations*, The Institute of Mathematics and its Applications Conference Series, pages 169–212. Clarendon Press, Oxford, 1985.
33. J. W. Ruge and K. Stüben. Algebraic multigrid (AMG). In S. F. McCormick, editor, *Multigrid Methods*, volume 3 of *Frontiers in Applied Mathematics*, pages 73–130, SIAM, Philadelphia, PA, 1987.
34. K. Stüben. Algebraic multigrid (AMG): experiences and comparisons. *Appl. Math. Comput.*, 13, 419–452, 1983.
35. A. van der Ploeg, E. Botta and F. Wubs. Nested grids ILU-decomposition (NGILU). *J. Comput. Appl. Math.*, 66, 515–526, 1996.
36. P. Vassilevski. 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, USA, 1989.
37. P. Vassilevski. Hybrid V-cycle algebraic multilevel preconditioners. *Math. Comp.*, 58, 489–512, 1992.
38. P. Vassilevski. On two ways of stabilizing the hierarchical basis multilevel methods. *SIAM Review*, 39, 18–53, 1997.
39. C. Wagner, W. Kinzelbach and G. Wittum. Schur-complement multigrid. *Numer. Math.*, 75, 523–545, 1997.
40. P. Wesseling. *An Introduction to Multigrid Methods*. J. Wiley and Sons, Chichester, 1992.
41. G. Wittum. On the robustness of ILU-smoothing. *SIAM J. Sci. Statist. Comput.*, 10, 699–717, 1989.
42. H. Yserentant. On the multi-level splitting of finite element spaces. *Numer. Math.*, 49, 379–412, 1986.