NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2000; **7**:243–267

A robust algebraic multilevel preconditioner for non-symmetric *M*-matrices

Yvan Notay*,†

Service de Métrologie Nucléaire (CP 165–84), Université Libre de Bruxelles, 50 Avenue F. D. Roosevelt, B-1050 Brussels, Belgium

SUMMARY

Stable finite difference approximations of convection–diffusion equations lead to large sparse linear systems of equations whose coefficient matrix is an *M*-matrix, which is highly non-symmetric when the convection dominates. For an efficient iterative solution of such systems, it is proposed to consider in the non-symmetric case an algebraic multilevel preconditioning method formerly proposed for pure diffusion problems, and for which theoretical results prove grid independent convergence in this context. These results are supplemented here by a Fourier analysis that applies to constant coefficient problems with periodic boundary conditions whenever using an 'idealized' version of the two-level preconditioner. Within this setting, it is proved that any eigenvalue λ of the preconditioned system satisfies $\left|\lambda^{-1} - 1 - i\,c\right| \le \frac{1}{2}$ for some real constant c such that $|c| \le \frac{1}{4}$. This result holds independently of the grid size and uniformly with respect to the ratio between convection and diffusion. Extensive numerical experiments are conducted to assess the convergence of practical two- and multi-level schemes. These experiments, which include problems with highly variable and rotating convective flow, indicate that the convergence is grid independent. It deteriorates moderately as the convection becomes increasingly dominating, but the convergence factor remains uniformly bounded. This conclusion is supported for both uniform and some non-uniform (stretched) grids. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: iterative methods; preconditioning; convection–diffusion

1. INTRODUCTION

We consider the iterative solution of large sparse linear systems

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

* Correspondence to: Yvan Notay, Service de Métrologie Nucléaire (CP 165–84), Université Libre de Bruxelles, 50 Avenue 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

Received 30 July 1999
Revised 29 December 1999

resulting from the finite difference approximation of PDEs of the form

$$-\nu\,\Delta u \;+\; \overline{v}\cdot\overline{\nabla}u \;=\; f \tag{2}$$

defined in some two dimensional domain Ω (with suitable boundary conditions). In this equation, $\overline{v}$ is a convective flow and the viscosity parameter ν governs the ratio between convection and diffusion. Such equations occur in many applications such as in transport models, but also as part of some iterative approaches to solving the Navier–Stokes equations (e.g. [1]).

As is well known, the standard central finite difference approximation of (2) produces a stable discretization only if

$$h \;\leq\; \frac{\nu}{\sup\limits_{(x,y)} \|\overline{v}\|_\infty} \tag{3}$$

in which case the system matrix A has non-positive off-diagonal entries and non-negative row-sum, i.e. is an M-matrix [2]. When condition (3) is not met, spurious oscillations appear in the discrete solution. To get rid of these, simple remedies resort to artificial viscosity or upwind finite difference approximation. In both cases, the matrix of the resulting linear system is again an M-matrix with non-negative row-sum [3]. The discretization is then only first order accurate, but second order accuracy can be restored by means of a defect correction technique (e.g. References [4–6]), for which a few (say 2–3) systems are to be solved successively with the same first order operator.

Hence, to develop iterative solution techniques for this kind of applications, it not too restrictive to assume that A is an M-matrix with non-negative row-sum. However, even for this class of matrices, many fewer results are available for the general case than for pure diffusion problems ($\overline{v}=0$), for which the system matrix is in addition symmetric and positive definite. To some extent, these results can be extrapolated to diffusion dominated problems, but it is more challenging to design a solver that remains robust when the convection (strongly) dominates, i.e. when the singular perturbation parameter ν goes to zero.

In the latter case, Gauss–Seidel relaxation can be very efficient for moderate grid sizes if the sweep direction follows the underlying flow direction [7]. To remain efficient for diffusion dominated problems and larger grid sizes, it suffices then to combine this approach with a coarse grid correction, i.e. to consider a multigrid scheme with Gauss–Seidel smoothing [3]. However, when the convective flow has variable direction, for instance when it is rotating, it is intrinsically difficult to perform the relaxation 'following' the flow. Consequently, standard multigrid algorithms may lose robustness, as the latter requires that the smoother becomes an exact solver in the limit $\nu \to 0$ [8]. Modern approaches to circumvent this problem include sophisticated renumbering techniques (e.g. Reference [9]), matrix dependent interpolation (e.g. Reference [10]) and algebraic multigrid like methods ([11] and the references therein).

In this paper, we attack problem (1) from the angle of preconditioning. A preconditioner is a matrix B such that solving a system $B\,\mathbf{x}=\mathbf{y}$ is relatively cheap whereas the preconditioned system

$$B^{-1}A\,\mathbf{u} \;=\; B^{-1}\mathbf{b}$$

(left preconditioning) or

$$A\,B^{-1}\widetilde{\mathbf{u}} \;=\; \mathbf{b}$$

(right preconditioning) is much easier to solve by an iterative method based on Krylov subspaces (e.g. References [12,13]). This depends mainly of the eigenvalues of $B^{-1}A$, which are to be as far as possible well clustered away from the origin of the complex plane ‡. To obtain a good preconditioner is particularly important in the non-symmetric case, since none of the many Krylov subspace methods can then guarantee monotone convergence at a reasonable cost if the system to be solved is too poorly (pre)conditioned.

This consideration leads us to focus more particularly on so-called algebraic multilevel preconditioning methods [14–29], which are based on a block incomplete factorization process applied to the system matrix partitioned in hierarchical form. Indeed, the above quoted works present many theoretical and numerical results showing that this approach is efficient to cluster the eigenvalues of the preconditioned system, at least in the symmetric case. Typically, these methods are also relatively cheap per iteration and robust with respect to jumps in the PDE coefficients and anisotropy or grid stretching.

Concerning the non-symmetric case, as these methods are essentially algebraic, they can in principle still be applied with little handling. However, few works address this point. In Reference [45], convection-diffusion problems are considered, but the preconditioner is computed from the discrete diffusion operator, which can of course be efficient only for diffusion dominated problems. In fact, only Reusken's works [22–24] attack directly convection dominated problems, showing that the algebraic multilevel approach can be efficient in this context too. However, he considers different, in general denser, coarse grid matrices than those generally used in the symmetric case, and the method, which is presented in a multigrid like framework, remains relatively costly per iteration.

In this paper, we consider the method presented in Reference [21], where a full theoretical support is developed for the symmetric case, proving grid independence convergence for the five-point approximation of diffusion problems under mild assumptions on the PDE coefficients (which allow arbitrary jumps or anisotropy).

As the construction of the preconditioner is purely algebraic, its extension to the non-symmetric case is straightforward. Our main theoretical result is then a proof (by a Fourier analysis) that all eigenvalues are nicely clustered when applying an 'idealized' version of the two level method (see Section 2) to constant coefficient problems on a regular grid with periodic boundary conditions. Although this result is less general than the one obtained in the symmetric case, it is nevertheless a strong proof of robustness, since it holds for any grid size and any five point stencil with non-positive off-diagonal entries and zero row-sum. Considering PDEs of the form (2), this means that our analysis holds uniformly in the singular perturbation parameter ν. Further, anisotropic diffusion is also allowed, as it may be induced in practice at the discrete level by using different mesh sizes in the x and y directions.

Concerning non-idealized two- and multi-level schemes, we conduct extensive numerical experiments to assess the eigenvalue distribution, from which we conclude that in the constant coefficient case the nice clustering proved for the 'ideal' version is in fact not deeply perturbed.

The method, which is proved robust in Reference [21] for diffusion problems with variable coefficients, is thus also robust for convection dominated problems with constant coefficients. This leads us to expect a good behavior in the general case of convection diffusion problems with variable coefficients. Here again, we resort to numerical experiments to assess this point, considering two difficult benchmark problems from the literature with highly variable and rotating flow.

The remainder of this paper is organized as follows. Section 2 contains our analysis of the 'ideal'

‡ It is generally recognized that the degree of normality also plays a role in the convergence rate, although not yet well understood in realistic applications.

two level scheme; the practical two-level method is presented in Section 3 and its multilevel extension considered in Section 4; numerical results for variable flows are presented in Section 5.

2. THE 'IDEAL' TWO-LEVEL SCHEME

Algebraic two-level preconditioners are based on a partitioning of the unknowns into fine and coarse grid ones. Let

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \tag{4}$$

be the corresponding 2×2 block form of the system matrix, where the first block of unknowns corresponds to the fine grid nodes and the second block of unknowns to the coarse grid nodes. An exact block factorization of A is written

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

where

$$S_A = A_{22} - A_{21} A_{11}^{-1} A_{12}$$

is the Schur complement. In general, A_{11}^{-1} is dense, and the latter is not computable at a reasonable cost. Algebraic multilevel methods have therefore as a key ingredient a technique to derive a (sparse) approximation S to S_A . If systems of the type $A_{11}\,\mathbf{x}_1 = \mathbf{y}_1$ can be solved exactly, the preconditioner is then written

$$B = \begin{pmatrix} A_{11} & \\ A_{21} & S \end{pmatrix} \begin{pmatrix} I & A_{11}^{-1} A_{12} \\ & I \end{pmatrix} \tag{5}$$

A_{11} has usually most of its rows strongly diagonally dominant and for this reason is fairly well conditioned, independently of the problem (mesh) size [30–32]. Hence, systems with A_{11} can in principle be solved relatively cheaply by classical iterative methods. Although in general, for cost efficiency reasons, only approximate solutions are performed, we shall thus consider here the 'ideal' preconditioner (5). Issues related to the approximate inversion of A_{11} are discussed in the next section.

Now, the theoretical foundation of algebraic two-level methods lies in the identity [26]

$$B^{-1} A = \begin{pmatrix} I & * \\ & S^{-1} S_A \end{pmatrix} \tag{6}$$

so that the quality of the preconditioner essentially depends on a good clustering of the eigenvalues of $S^{-1} S_A$.

In the finite element context, most methods proposed so far take S equal to the coarse grid discretization matrix, which is naturally available through a basis transformation (in this way, one generalizes the hierarchical basis multigrid method [32–35]). In the finite difference context, this would require rediscretization, so we propose in Reference [21] a more algebraic alternative. Letting Δ be the diagonal matrix with same row-sum as A_{11}, we take S equal to the Schur complement of

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

in which one has deleted the zero lines and columns. In other words,

$$S = A_{22} - A_{21} K A_{12} \tag{7}$$

where K is the diagonal matrix defined by

$$K_{ii} = \begin{cases} (A_{11}\,\mathbf{e}_1)_i^{-1} & \text{if } (A_{11}\,\mathbf{e}_1)_i \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

$\mathbf{e}_1 = (1 \quad 1 \quad \dots \quad 1)^{\mathrm{T}}$ being the vector with all components equal to unity.

From a practical point of view, note that this method is pretty easy to implement, and that if A originates from a five-point scheme, S will present relative to the coarse grid a similar five-point structure. From a more theoretical point of view, we prove in Lemma 2.1 below that, when A is an M-matrix with non-negative row-sum, then S is itself an M-matrix with non-negative row-sum. Hence, a recursive use of the preconditioning technique is consistent in this context.

Lemma 2.1. *Let A be a square matrix with non-positive off-diagonal entries and non-negative row-sum. Partition A as in Equation (4), and define S by Equations (7), (8). Then, S has also non-positive off-diagonal entries and non-negative row-sum.*

Proof

It is easy to check with Equations (7), (8) that the off-diagonal entries of S are non-positive. Further, Equation (8) implies $K\, A_{11}\, \mathbf{e}_1 \leq \mathbf{e}_1$, whence, noting that $A_{21} \leq 0$,

$$\begin{aligned} S\,\mathbf{e}_2 &= A_{22}\,\mathbf{e}_2 - A_{21}\,K\,A_{12}\,\mathbf{e}_2 \\ &= A_{22}\,\mathbf{e}_2 + A_{21}\,K\,A_{11}\,\mathbf{e}_1 - A_{21}\,K\,(A\mathbf{e})_1 \\ &\geq (A\,\mathbf{e})_2 \end{aligned}$$

■

Of course, these nice properties would be useless without a good clustering of the eigenvalues of $S^{-1}S_A$ (Lemma 2.1 does not even precludes S from being singular). In Reference [21], considering the symmetric case, we show that these eigenvalues, necessarily real, are bounded below by 1 and above by 2 when the PDE coefficients are piecewise constant on the coarse mesh. This result is partly based on the observation that, under the given assumptions, S is equal to the coarse grid discretization matrix times some scaling factor.

Now, in the general case, S may have little in common with a coarse discretization and the analysis of the spectrum of $S^{-1}S_A$ has to be based on different arguments. Since the eigenvalues are complex, we found that the only viable way was a Fourier analysis (see e.g. Reference [3]). As usual, the latter comes with some limitations, namely we are restricted to constant coefficient problems and we have in addition to assume periodic boundary conditions instead of the physical ones. Considering PDEs of the form (2), this also means that the discrete system becomes singular since the corresponding stencil has zero row-sum. It is then easy to see that S, and therefore the preconditioner (5), become necessarily singular too. Although the behavior of iterative solvers for singular systems in now well understood (see Reference [36] for a treatment of singular preconditioners), this raises an additional difficulty since, in fact, we are not interested in solving the singular system, but we want to extrapolate the convergence properties observed in this case to a nearby system, in general non-singular. That is why in the following theorem we assume that the stencil has row-sum $\varepsilon > 0$, and report on $\lim_{\varepsilon\to 0} \lambda_\varepsilon$, where λ_ε denotes an eigenvalue of $B_\varepsilon^{-1} A_\varepsilon$. Of course, for the regular eigenvalues, this limit converges to the corresponding eigenvalue of the pencil $A - \lambda B$, but an additional indication is obtained for the eigenmode associated with the constant vector, which in fact spans the null space of A.

Theorem 2.1. *Assume that some PDE with periodic boundary conditions is defined in $\Omega =]-1,1] \times]-1,1]$. Let Ω be discretized by means of a uniform regular grid of mesh size $h = 1/2N$, where N is some positive integer. Assume that the discretization scheme and the PDE are such that the latter is approximated on this grid by a constant five-point stencil*

$$\begin{array}{ccc} & -a_N & \\ -a_W & a & -a_E \\ & -a_S & \end{array} , \quad a = (a_N + a_S + a_W + a_E)(1+\varepsilon) \tag{9}$$

with $\varepsilon > 0$ and a_N, a_S, a_W, $a_E \geq 0$.

Denote by A_ε the resulting matrix, and partition it as in Equation (4) on the basis of a partitioning of the nodes for which the second block is formed by the nodes belonging to the uniform grid of mesh size $2h$. Let B_ε be defined by Equation (5) with S given by Equations (7), (8).

Then, B_ε is non-singular, and, for any eigenvalue λ_ε of $B_\varepsilon^{-1} A_\varepsilon$, letting $\lambda = \lim_{\varepsilon\to 0} \lambda_\varepsilon$, one has

$$\left| \lambda^{-1} - 1 - i\,c \right| \leq \frac{1}{2} \tag{10}$$

for some real constant c such that $|c| \leq \frac{1}{4}$.

Proof
We first gather the notation that will be used in this proof (k, ℓ denote some integers, see below).

$$\begin{array}{llll}
\sigma_x & = a_W + a_E & \sigma_y & = a_S + a_N \\
\delta_x & = \sigma_x^{-1} (a_E - a_W) & \delta_y & = \sigma_y^{-1} (a_N - a_S) \\
\delta_+ & = \frac{1}{2} (\delta_x + \delta_y) & \delta_- & = \frac{1}{2} (\delta_x - \delta_y) \\
w_x & = a_W \,\mathrm{e}^{-i\pi k h} + a_E \,\mathrm{e}^{i\pi k h} & w_y & = a_S \,\mathrm{e}^{-i\pi \ell h} + a_N \,\mathrm{e}^{i\pi \ell h} \\
c_k & = \cos(\pi k h) & s_k & = \sin(\pi k h) \\
c_+ & = \cos\left(\pi h \frac{k+\ell}{2}\right) & c_- & = \cos\left(\pi h \frac{k-\ell}{2}\right) \\
s_+ & = \sin\left(\pi h \frac{k+\ell}{2}\right) & s_- & = \sin\left(\pi h \frac{k-\ell}{2}\right) \\
c_m & = \max(|c_+|, |c_-|) & s_m & = \max(|s_+|, |s_-|)
\end{array}$$

Note that, according to this notation,

$$w_x = \sigma_x \, (c_k + i\,\delta_x\, s_k), \quad w_y = \sigma_y \, \left(c_\ell + i\,\delta_y\, s_\ell\right)$$

whereas $|\delta_x|$, $|\delta_y| \leq 1$. On the other hand, multiplying A_ε by some factor entails the multiplication of B_ε by the same factor. Hence, the scaling of the stencil is unimportant, and we may assume without loss of generality that

$$a_N + a_S + a_W + a_E \;=\; \sigma_x + \sigma_y \;=\; 1$$

We now start the proof by noting that S is the matrix associated to the constant stencil

$$\begin{array}{ccc} & -\dfrac{a_N^2}{\sigma_y+\varepsilon} & \\ -\dfrac{a_W^2}{\sigma_x+\varepsilon} & \sigma_S & -\frac{a_E^2}{\sigma_x+\varepsilon} \\ & -\dfrac{a_S^2}{\sigma_y+\varepsilon} & \end{array}\;, \quad \sigma_S = 1+\varepsilon - \frac{2\,a_W\,a_E}{\sigma_x+\varepsilon} - \frac{2\,a_N\,a_S}{\sigma_y+\varepsilon}$$

applied to the periodic grid formed by the coarse nodes. It follows that the eigenvectors of S are the $4\,N^2$ functions

$$v_{k\ell} \;=\; \mathrm{e}^{i\,\pi\,k\,x}\,\mathrm{e}^{i\,\pi\,\ell\,y}\,, \qquad 1-N \leq k\,,\ \ell \leq N \tag{11}$$

evaluated at the (coarse) grid points [3, Section 8.1.2]. It is then easy to see that the corresponding eigenvalues can be written

$$\begin{aligned} S_{k\ell}^{(\varepsilon)} \;&=\; \sigma_S - \frac{a_W^2}{\sigma_x+\varepsilon}\,\mathrm{e}^{-i\,\pi\,k\,(2h)} - \frac{a_E^2}{\sigma_x+\varepsilon}\,\mathrm{e}^{i\,\pi\,k\,(2h)} - \frac{a_S^2}{\sigma_y+\varepsilon}\,\mathrm{e}^{-i\,\pi\,\ell\,(2h)} - \frac{a_N^2}{\sigma_y+\varepsilon}\,\mathrm{e}^{i\,\pi\,\ell\,(2h)} \\ &=\; 1+\varepsilon - \frac{w_x^2}{\sigma_x+\varepsilon} - \frac{w_y^2}{\sigma_y+\varepsilon} \end{aligned}$$

Since $|w_x| \leq \sigma_x$, $|w_y| \leq \sigma_y$, one has $\left|S_{k\ell}^{(\varepsilon)}\right| \geq \varepsilon > 0$, proving that S is invertible and therefore that B_ε is non-singular.

Then, by Equation (6), Equation (10) holds if and only if it holds with respect to the eigenvalues of $S^{-1}S_A$. On the other hand, it is shown in Reference [22] that the vectors (11) are also eigenvectors of S_A^{-1} with corresponding eigenvalue[§]

$$\begin{aligned} T_{k\ell}^{(\varepsilon)} \;&=\; \frac{1}{4}\left(\frac{1}{1+\varepsilon-(w_x+w_y)} + \frac{1}{1+\varepsilon+(w_x+w_y)} + \frac{1}{1+\varepsilon-(w_x-w_y)}\right. \\ &\qquad \left. + \frac{1}{1+\varepsilon+(w_x-w_y)}\right) \\ &=\; \frac{1}{2}\left(\frac{1+\varepsilon}{(1+\varepsilon)^2-(w_x+w_y)^2} + \frac{1+\varepsilon}{(1+\varepsilon)^2-(w_x-w_y)^2}\right) \end{aligned}$$

[§] Strictly speaking, the proof is given only for $\varepsilon = 0$, but the extension is straightforward.

We first examine the case where $\sigma_x = 0$. Then $w_x = 0$, whence

$$S_{k\ell}^{(\varepsilon)} = 1 + \varepsilon - \frac{w_y^2}{\sigma_y + \varepsilon}, \quad T_{k\ell}^{(\varepsilon)} = \frac{1+\varepsilon}{(1+\varepsilon)^2 - w_y^2}$$

i.e. $S_{k\ell}^{(\varepsilon)}\, T_{k\ell}^{(\varepsilon)} = 1$, showing that $B_\varepsilon = A_\varepsilon$. As a similar result holds when $\sigma_y = 0$, we pursue the proof assuming that both σ_x and σ_y are positive.

Then, since c_k, $c_\ell \ge 0$ (remember that $h = 1/2\,N$), $w_x - w_y$ can never be equal to 1 whereas $w_x + w_y = 1$ if and only if $k = \ell = 0$. We consider this case separately as it is the only one for which $T_{k\ell}^{(\varepsilon)}$ does not have a well defined limit for $\varepsilon \to 0$. Nevertheless,

$$\begin{aligned} S_{00}^{(\varepsilon)} &= 1 + \varepsilon - \frac{\sigma_x^2}{\sigma_x + \varepsilon} - \frac{\sigma_y^2}{\sigma_y + \varepsilon} \\ T_{00}^{(\varepsilon)} &= \frac{1}{2}\left(\frac{1+\varepsilon}{\varepsilon\,(\varepsilon+2)} + \frac{1+\varepsilon}{(1+\varepsilon)^2 - (\sigma_x - \sigma_y)^2} \right) \end{aligned}$$

so that, since $|\sigma_x - \sigma_y| < 1$,

$$\lambda_{00}^{-1} = \lim_{\varepsilon\to 0} S_{00}^{(\varepsilon)}\, T_{00}^{(\varepsilon)} = \lim_{\varepsilon\to 0} \frac{1 + \varepsilon - \dfrac{\sigma_x}{1+\dfrac{\varepsilon}{\sigma_x}} - \dfrac{\sigma_y}{1+\frac{\varepsilon}{\sigma_y}}}{4\,\varepsilon} = \frac{3}{4}$$

Obviously, this is compatible with Equation (10).

We shall thus assume in the remaining of this proof that either $k \neq 0$ or $\ell \neq 0$. Therefore,

$$\lambda_{k\ell}^{-1} = \lim_{\varepsilon\to 0} S_{k\ell}^{(\varepsilon)}\, T_{k\ell}^{(\varepsilon)} = S_{k\ell}\, T_{k\ell}$$

where

$$S_{k\ell} = 1 - \frac{w_x^2}{\sigma_x} - \frac{w_y^2}{\sigma_y}, \quad T_{k\ell} = \frac{1}{2}\left(\frac{1}{1-(w_x+w_y)^2} + \frac{1}{1-(w_x-w_y)^2} \right).$$

One may check that

$$\begin{aligned} S_{k\ell} &= 1 - (w_x+w_y)^2 - \left(\sqrt{\frac{\sigma_y}{\sigma_x}}\, w_x - \sqrt{\frac{\sigma_x}{\sigma_y}}\, w_y \right)^2 \\ &= 1 - (w_x-w_y)^2 - \left(\sqrt{\frac{\sigma_y}{\sigma_x}}\, w_x + \sqrt{\frac{\sigma_x}{\sigma_y}}\, w_y \right)^2 \end{aligned}$$

whence

$$1-\lambda_{k\ell}^{-1} \;=\; \frac{1}{2}\left(\frac{\left(\sqrt{\frac{\sigma_y}{\sigma_x}}\,w_x-\sqrt{\frac{\sigma_x}{\sigma_y}}\,w_y\right)^2}{1-(w_x+w_y)^2}\;+\;\frac{\left(\sqrt{\frac{\sigma_y}{\sigma_x}}\,w_x+\sqrt{\frac{\sigma_x}{\sigma_y}}\,w_y\right)^2}{1-(w_x-w_y)^2}\right)$$

Now, noting that $\Re\mathrm{e}\left(z^2\right)\le(\Re\mathrm{e}(z))^2\;\;\forall\, z\in\mathscr{C}\;\left((z_r+i\,z_i)^2=z_r^2-z_i^2+2\,i\,z_r\,z_i\right)$, one has

$$\begin{aligned}
\Re\mathrm{e}\left(1-(w_x+w_y)^2\right) &\ge 1-\left(\Re\mathrm{e}(w_x+w_y)\right)^2\\
&= (\sigma_x+\sigma_y)^2-(\sigma_x\,c_k+\sigma_y\,c_\ell)^2\\
&= \sigma_x^2\,s_k^2+\sigma_y^2\,s_\ell^2+2\,\sigma_x\,\sigma_y\,(1-c_k\,c_\ell)\\
&\ge 2\,\sigma_x\,\sigma_y\,(1-c_k\,c_\ell+|s_k\,s_\ell|)\\
&= 2\,\sigma_x\,\sigma_y\left(1-\max\left(\cos(\pi\,h(k-\ell))\,,\;\cos(\pi\,h(k+\ell))\right)\right)\\
&= 4\,\sigma_x\,\sigma_y\,s_m^2
\end{aligned}$$

Similarly, one finds

$$\begin{aligned}
\Re\mathrm{e}\left(1-(w_x-w_y)^2\right) &\ge 2\,\sigma_x\,\sigma_y\left(1+\max\left(\cos(\pi\,h(k-\ell))\,,\;\cos(\pi\,h(k+\ell))\right)\right)\\
&= 4\,\sigma_x\,\sigma_y\,c_m^2
\end{aligned}$$

We exploit these latter results in the following way: for any $z\in\mathscr{C}$, $\Re\mathrm{e}(z)\ge 1$ implies that z^{-1} belongs to the disc with centre $(\frac{1}{2}\,,\;0)$ and radius $\frac{1}{2}$, i.e. $z^{-1}=\frac{1}{2}(1+\eta)$ for some η such that $|\eta|\le 1$. Therefore, letting

$$\frac{\sqrt{\frac{\sigma_y}{\sigma_x}}\,w_x-\sqrt{\frac{\sigma_x}{\sigma_y}}\,w_y}{2\,s_m\,\sqrt{\sigma_x\,\sigma_y}}=\alpha_r+i\,\alpha_i\,,\qquad \frac{\sqrt{\frac{\sigma_y}{\sigma_x}}\,w_x+\sqrt{\frac{\sigma_x}{\sigma_y}}\,w_y}{2\,c_m\,\sqrt{\sigma_x\,\sigma_y}}=\beta_r+i\,\beta_i$$

(with α_r , α_i , β_r , β_i real), one has

$$1-\lambda_{k\ell}^{-1}\;=\;\frac{1}{4}\left((\alpha_r+i\,\alpha_i)^2\;(1+\eta_1)\;+\;(\beta_r+i\,\beta_i)^2\;(1+\eta_2)\right)$$

for some η_1 , η_2 such that $|\eta_1|$, $|\eta_2|\le 1$. One then easily finds

$$\begin{aligned}
\left|1-\lambda_{k\ell}^{-1}-i\,\frac{\alpha_r\,\alpha_i+\beta_r\,\beta_i}{2}\right| &\le \frac{1}{4}\left|\alpha_r^2-\alpha_i^2+\beta_r^2-\beta_i^2\right|+\frac{1}{4}\left(\alpha_r^2+\alpha_i^2+\beta_r^2+\beta_i^2\right)\\
&= \frac{1}{2}\,\max\left(\alpha_r^2+\beta_r^2\,,\;\alpha_i^2+\beta_i^2\right)\;.
\end{aligned}$$

Hence the required result (10) is proved if

$$\alpha_r^2+\beta_r^2 \;\le\; 1 \tag{12}$$

$$\alpha_i^2+\beta_i^2 \;\le\; 1 \tag{13}$$

$$|\alpha_r\, \alpha_i + \beta_r\, \beta_i| \quad \leq \quad \frac{1}{2} \tag{14}$$

Now,

$$\alpha_r = \frac{1}{2\, s_m}\,(c_k - c_\ell) = -s_m^{-1}\,(s_+\, s_-), \qquad \beta_r = \frac{1}{2\, c_m}\,(c_k + c_\ell) = c_m^{-1}\,(c_+\, c_-)$$

from which one easily deduces Equation (12) (using either $|s_-| \leq s_m$ and $|c_-| \leq c_m$ or $|s_+| \leq s_m$ and $|c_+| \leq c_m$). On the other hand,

$$\begin{aligned}
\alpha_i &= \frac{1}{2\, s_m}\,(\delta_x\, s_k - \delta_y\, s_\ell) \;=\; \frac{1}{s_m}\,(\delta_+\, s_-\, c_+ + \delta_-\, s_+\, c_-) \\
\beta_i &= \frac{1}{2\, c_m}\,(\delta_x\, s_k + \delta_y\, s_\ell) \;=\; \frac{1}{c_m}\,(\delta_+\, s_+\, c_- + \delta_-\, s_-\, c_+)
\end{aligned}$$

whence (since $|s_-\, s_+ + c_-\, c_+| = |c_\ell| \leq 1$ and $|s_-\, s_+ - c_-\, c_+| = |c_k| \leq 1$)

$$\begin{aligned}
\alpha_i^2 + \beta_i^2 &= \delta_+^2 \left(\frac{s_-^2\, c_+^2}{s_m^2} + \frac{s_+^2\, c_-^2}{c_m^2}\right) + \delta_-^2 \left(\frac{s_+^2\, c_-^2}{s_m^2} + \frac{s_-^2\, c_+^2}{c_m^2}\right) \\
&\qquad + 2\, \delta_+\, \delta_-\, (s_-\, c_-\, s_+\, c_+) \left(\frac{1}{s_m^2} + \frac{1}{c_m^2}\right) \\
&\leq \delta_+^2 + \delta_-^2 + 2\, |\delta_+\, \delta_-|\, (|c_-\, c_+| + |s_-\, s_+|) \\
&\leq (|\delta_+| + |\delta_-|)^2 \\
&= \big(\max(|\delta_x|\,,\; |\delta_y|)\big)^2 \\
&\leq 1
\end{aligned}$$

Finally,

$$\begin{aligned}
\alpha_r\, \alpha_i &= \frac{-1}{s_m^2}\, \Big(\delta_+\, s_-^2\, s_+\, c_+ + \delta_-\, s_+^2\, s_-\, c_-\Big) \\
\beta_r\, \beta_i &= \frac{1}{c_m^2}\, \Big(\delta_+\, c_-^2\, s_+\, c_+ + \delta_-\, c_+^2\, s_-\, c_-\Big)
\end{aligned}$$

whence

$$\alpha_r\, \alpha_i + \beta_r\, \beta_i \quad = \quad \delta_+\, s_+\, c_+ \left(\frac{c_-^2}{c_m^2} - \frac{s_-^2}{s_m^2}\right) \;+\; \delta_-\, s_-\, c_- \left(\frac{c_+^2}{c_m^2} - \frac{s_+^2}{s_m^2}\right)$$

Now,

$$\left|\frac{c_-^2}{c_m^2} - \frac{s_-^2}{s_m^2}\right| \leq 1 \text{ and } \left|\frac{c_+^2}{c_m^2} - \frac{s_+^2}{s_m^2}\right| \leq 1$$

(in both cases, one subtracts two numbers in $[0\,,\; 1]$). Using $s_+\, c_+ = \frac{1}{2}\sin(\pi\, h(k + \ell))$ and $s_-\, c_- =$

$\frac{1}{2}\sin(\pi\, h(k-\ell))$, one then easily finds

$$|\alpha_r\, \alpha_i + \beta_r\, \beta_i| \;\leq\; \frac{1}{2}\;(|\delta_+| + |\delta_-|) \;=\; \frac{1}{2}\;\max(|\delta_x|\,,\;|\delta_y|)$$

whence Equation (14). ■

Note that, despite the limitations inherent in Fourier analysis, Theorem 2.1 provides a strong proof of robustness, as Equation (10) holds independently of the grid size and for any stencil with non-positive off-diagonal entries and zero row-sum. In particular, if the analysis is applicable to some matrix A, it will also be applicable to its approximate Schur complement S, i.e. our result is consistent with a recursive use of the preconditioning technique as arises in a multilevel framework.

Finally, to provide a graphical illustration, we have represented on Figure 1 the region that has to contain the eigenvalues, together with the actual eigenvalues as computed for the PDE (2) with constant flow $\overline{v} = (\cos\beta\,,\ \sin\beta)$, $\Omega =]-1,1]\times]-1,1]$, periodic boundary conditions and a five-point up-wind scheme applied on a uniform grid of mesh size $h = 1/64$. One may see that our analysis is relatively sharp, although still slightly too pessimistic.

3. A PRACTICAL TWO-LEVEL SCHEME

Although from a theoretical point of view, it is comfortable to assume the exact inversion of A_{11}, in practice two such inversions are required for each application of the preconditioner (5). Hence, the overall cost will be kept reasonable only if the number of operations involved remains a small multiple of the number of non-zero elements in A_{11}. This means that practical two-level schemes have to rely on an approximate inversion of A_{11}. The preconditioner has thus to be rewritten

$$B \;=\; \begin{pmatrix} P & \\ A_{21} & S \end{pmatrix} \begin{pmatrix} I & P^{-1}A_{12} \\ & I \end{pmatrix}$$

where P is a preconditioner for A_{11}, possibly implicitly defined by a very few steps of some iterative procedure.

Here, we have to recall that with a naive choice of P, eigenvalues of $B^{-1}A$ may be unbounded even though those of $P^{-1}A_{11}$ and of $S^{-1}S_A$ are both nicely clustered [37]. However, we show in the same paper that, at least in the symmetric case, problems are prevented when P satisfies some given algebraic requirements that are in particular met when A is an M-matrix and P is computed from a *modified* ILU (MILU) [12,38] factorization of A_{11}. This algebraic analysis is further refined in Reference [21] for the specific choice of approximate Schur complement S discussed in Section 2, leading to the proof that

$$\lambda_{\min}\left(B^{-1}A\right) \;\;\geq\;\; 1 \tag{15}$$

$$\lambda_{\max}\left(B^{-1}A\right) \;\;\leq\;\; \lambda_{\max}\left(P^{-1}A_{11}\right)\cdot\lambda_{\max}\left(S^{-1}S_A\right) \tag{16}$$

under the only assumption that A is a symmetric M-matrix with non-negative row-sum. On the other hand, a further analysis of five-point finite difference matrices shows that $\lambda_{\max}\left(P^{-1}A_{11}\right)$ cannot ex-

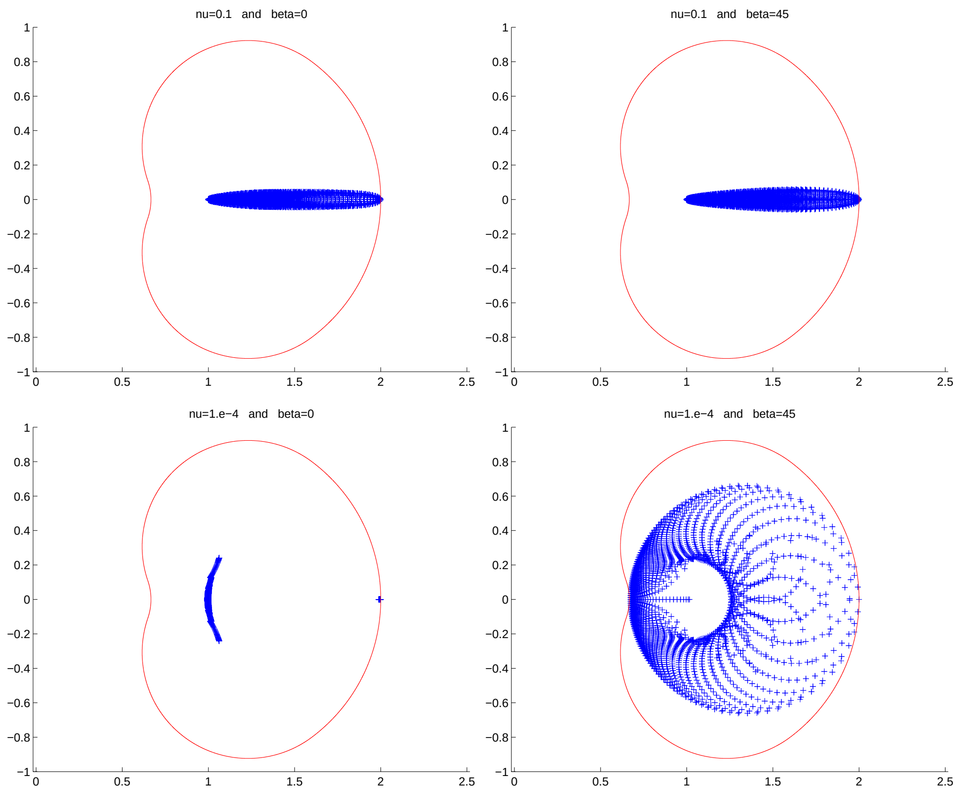

Figure 1. Spectrum of the preconditioned system for the constant flow problem with periodic boundary conditions; eigenvalues are represented by a $+$; the continuous line delimits the region that contains the eigenvalues according to Equation (10); angles β are given in degrees.

ceed $\frac{3}{2}$ even with the simplest scheme of an MILU factorization without fill-in. Hence, although the approximation used is very cheap, the eigenvalue distribution is not deeply perturbed from that of the 'ideal' two-level scheme. Clearly, such an approach can be much better, in terms of cost efficiency, than those that attempt to mimic the exact inversion of A_{11} by performing several inner iterations.

Concerning the non-symmetric case, MILU factorization is still well defined since A_{11} is an M-matrix with most of its rows strongly diagonally dominant [12, Section 7.1]. For completeness, let us recall that, discarding all fill-in, P can be written

$$P = (Q - E)Q^{-1}(Q - F)$$

where $(-E)$ is the strictly lower part of A_{11}, $(-F)$ is the strictly upper part of A_{11} and Q is diagonal and computed such that

$$P\,\mathbf{e}_1 \;=\; A_{11}\,\mathbf{e}_1$$

As in the symmetric case, the good conditioning properties of A_{11} together with the robustness of ILU methods should be reflected in the eigenvalues of $P^{-1}A_{11}$, which we expect therefore to be very well clustered. So, as in the symmetric case, the main danger comes from the global effect on the eigenvalues of $B^{-1}A$, which, as shown in Reference [37], can be very sensitive to the inexact inversion of A_{11}, depending on the type of preconditioner used.

Since an extension of the theoretical results for the symmetric case is presently out of reach, we assessed the eigenvalue distribution through extensive numerical experiments. More precisely, we considered the five-point up-wind finite difference approximation of respectively

$$-\nu\,\Delta u \;+\; \overline{v}\cdot\overline{\nabla}u \;=\; 0$$

and

$$-\nu\left(\partial^2_{xx}\,u \;+\; 10^{-2}\,\partial^2_{yy}\,u\right) \;+\; \overline{v}\cdot\overline{\nabla}u \;=\; 0$$

in the unit square, with Dirichlet boundary conditions in each case, constant flow $\overline{v} = (\cos\beta\,,\ \sin\beta)$, and a uniform grid of mesh size $h = 1/128$. Instead of reporting on realistic convergence rates with some well chosen Krylov accelerator, we computed the asymptotic convergence factor ρ associated with stationary iterations of the form

$$\mathbf{u}_{k+1} \;=\; \mathbf{u}_k + \tau^{-1}\,B^{-1}\,(\mathbf{b} - A\,\mathbf{u}_k) \tag{17}$$

with $\tau = 1.5$ (minimum 200 iterations). Since

$$\rho \;=\; \max_{\lambda\,\in\,\sigma\left(B^{-1}A\right)}\left|\frac{\lambda}{\tau} - 1\right|$$

we can indeed have a better idea of the eigenvalue distribution with such an experiment, and check if the clustering of around 1.5 observed for the ‘ideal’ two-level scheme is deeply perturbed or not.

The results are reported in Tables I and II. One may see that ρ is at worst equal to 0.55, which means that all eigenvalues are inside a circle of centre $(1.5, 0)$ and radius 0.825. One may also observe that this worst case value is also reached in some near symmetric situations ($\nu = 1$), hence, in the two-level case, non-symmetric problems are not necessarily more difficult than symmetric ones, for which there exists full theoretical support.

4. THE MULTILEVEL METHOD

The two-level preconditioner requires solving a system $S\,\mathbf{x}_2 \;=\; \mathbf{y}_2$. In the multilevel case, this is exchanged for a multiplication by

$$M \;=\; \mathcal{P}_m\left(B^{(S)\,-1}\,S\right)\,S^{-1} \tag{18}$$

where $\mathcal{P}_m$ is a polynomial such that $\mathcal{P}_m(0) = 0$ and $B^{(S)}$ is a preconditioner for S. Using for $B^{(S)}$ the same type of preconditioning as for A leads then to a multilevel scheme in which the global preconditioner is defined only implicitly through a recursion. This recursion is followed until the number

Table I. Asymptotic convergence factors for the practical two-level scheme (isotropic diffusion).

β	ν = 1	10^{-2}	10^{-4}	10^{-6}
0	0.41	0.33	0.33	0.33
$\pi/8$	0.41	0.33	0.51	0.52
$\pi/4$	0.41	0.33	0.54	0.55
$3\pi/8$	0.41	0.33	0.51	0.52
$\pi/2$	0.41	0.33	0.33	0.33
$5\pi/8$	0.41	0.37	0.50	0.51
$3\pi/4$	0.41	0.38	0.54	0.55
$7\pi/8$	0.41	0.37	0.49	0.50
π	0.41	0.33	0.33	0.33
$9\pi/8$	0.41	0.33	0.51	0.52
$5\pi/4$	0.41	0.33	0.54	0.55
$11\pi/8$	0.41	0.33	0.50	0.52
$3\pi/2$	0.41	0.33	0.33	0.33
$13\pi/8$	0.41	0.37	0.49	0.50
$7\pi/4$	0.41	0.38	0.54	0.55
$15\pi/8$	0.41	0.37	0.50	0.51

Table II. Asymptotic convergence factors for the practical two-level scheme (anisotropic diffusion).

β	ν = 1	10^{-2}	10^{-4}	10^{-6}
0	0.51	0.33	0.33	0.33
$\pi/8$	0.49	0.38	0.51	0.52
$\pi/4$	0.47	0.41	0.55	0.55
$3\pi/8$	0.46	0.42	0.51	0.52
$\pi/2$	0.46	0.43	0.33	0.33
$5\pi/8$	0.46	0.43	0.51	0.51
$3\pi/4$	0.47	0.41	0.54	0.55
$7\pi/8$	0.49	0.33	0.49	0.50
π	0.51	0.33	0.33	0.33
$9\pi/8$	0.53	0.38	0.51	0.52
$5\pi/4$	0.54	0.41	0.55	0.55
$11\pi/8$	0.55	0.42	0.51	0.52
$3\pi/2$	0.55	0.43	0.33	0.33
$13\pi/8$	0.55	0.43	0.49	0.50
$7\pi/4$	0.54	0.41	0.54	0.55
$15\pi/8$	0.53	0.33	0.51	0.51

of nodes left in the second block is sufficiently small to allow an exact factorization of the approximate Schur complement S .

Now Equation (18) means in practice that one uses a few steps of an iterative method to solve the system approximately with S :

$$\begin{aligned}\mathbf{x}_2^{(0)} &= 0\\ \mathbf{x}_2^{(k+1)} &= \mathbf{x}_2^{(k)} + \tau_k^{-1}\ B^{(S)\,-1}\left(\mathbf{y}_2 - S\,\mathbf{x}_2^{(k)}\right), \quad k = 0, \dots, m-1\end{aligned}$$

In the symmetric case [21], we pay particular attention to the choice $m = 2$, for which the preconditioner remains fairly cheap (see below). We prove that the convergence rate is bounded independently of the number of levels when $\tau_0 = \tau_1 > \overline{\lambda}/2$, where $\overline{\lambda} = \lambda_{\max}\left(B_2^{-1}A\right)$ is the largest eigenvalue for the two-level method alone (note that this should hold for all successive levels considered in the recursion). On the other hand, fastest convergence is obtained by setting $\tau_0 = \tau_1 \approx \overline{\lambda}^{1/2}$. Since, assuming that the PDE coefficients are piecewise constant on the coarsest grid, we prove $\overline{\lambda} \leq 3$, whereas $\overline{\lambda} \approx 8/3$ is a more realistic estimate, we conclude that $\tau_0 = \tau_1 = \sqrt{8/3} = 1.63$ offers the best compromise between efficiency and robustness.

Here again, in the non-symmetric case, little can be said from a theoretical point of view. Heuristically, we nevertheless expect that the best possible choice for τ_k is something close to the centre of the eigenvalue distribution for the two-level method. Considering our results in Section 2, and looking in particular at Figure 1, keeping $\tau_0 = \tau_1 = 1.63$ appears therefore very reasonable, especially if one takes into account that the MILU preconditioning of A_{11} tends to shift the eigenvalue distribution somewhat to the right (see Equations (15), (16)). Moreover, it is highly desirable to obtain a method whose basic parameters need not to be adapted as a function of the degree of symmetry. We therefore did not consider further optimization of these parameters, and assessed the eigenvalue distribution corresponding to $m = 2$ and $\tau_0 = \tau_1 = 1.63$ through numerical tests.

More precisely, we repeated the two experiments of the preceding section, now using the multilevel method with six levels (the coarsest grid is then 3×3). The results are reported in Tables III and IV. Here, the worst value is $\rho = 0.64$, which corresponds to a circle of centre $(1.5,\ 0)$ and radius 0.96 . The situation is somewhat less favourable in highly non-symmetric situations than in near-symmetric ones, but the deterioration for decreasing ν is not sufficient really to spoil the convergence. We explain this behaviour by the fact that polynomial iterative methods cannot be as efficient when the eigenvalues have relatively large imaginary parts, hence the approximation (18) to S^{-1} is necessarily less accurate for small ν .

Computational complexity

Let n be the order of the system (1), n_1 the number of nodes in block 1 and n_2 the number of nodes in block 2. Let $nz(C)$ denote the number of non-zero entries in (sub)matrix C, and a *flop* be a basic ($+$ or $*$) arithmetic operation.

One application of the multilevel preconditioner requires

- two solutions with P, each of them involving $2\,nz(A_{11})$ *flops* ;
- one multiplication by A_{12} ($2\,nz(A_{12})$ *flops*) and one multiplication by A_{12} ($2\,nz(A_{21})$ *flops*);

Table III. Asymptotic convergence factor for the multilevel scheme (isotropic diffusion).

β	ν = 1	10^{-2}	10^{-4}	10^{-6}
0	0.43	0.44	0.53	0.54
$\pi/8$	0.43	0.44	0.59	0.59
$\pi/4$	0.43	0.44	0.63	0.64
$3\pi/8$	0.43	0.44	0.57	0.59
$\pi/2$	0.43	0.44	0.54	0.54
$5\pi/8$	0.43	0.44	0.58	0.59
$3\pi/4$	0.43	0.44	0.63	0.64
$7\pi/8$	0.43	0.44	0.57	0.58
π	0.43	0.44	0.53	0.54
$9\pi/8$	0.43	0.44	0.59	0.59
$5\pi/4$	0.43	0.44	0.63	0.64
$11\pi/8$	0.43	0.44	0.57	0.59
$3\pi/2$	0.43	0.44	0.54	0.54
$13\pi/8$	0.43	0.44	0.57	0.58
$7\pi/4$	0.43	0.44	0.62	0.63
$15\pi/8$	0.43	0.44	0.58	0.58

Table IV. Asymptotic convergence factor for the multilevel scheme (anisotropic diffusion).

β	ν = 1	10^{-2}	10^{-4}	10^{-6}
0	0.48	0.53	0.54	0.54
$\pi/8$	0.47	0.47	0.59	0.59
$\pi/4$	0.48	0.50	0.64	0.64
$3\pi/8$	0.48	0.51	0.58	0.59
$\pi/2$	0.47	0.51	0.53	0.54
$5\pi/8$	0.48	0.50	0.58	0.59
$3\pi/4$	0.48	0.49	0.63	0.64
$7\pi/8$	0.48	0.48	0.57	0.58
π	0.48	0.52	0.54	0.54
$9\pi/8$	0.50	0.48	0.59	0.59
$5\pi/4$	0.52	0.50	0.64	0.64
$11\pi/8$	0.53	0.51	0.57	0.59
$3\pi/2$	0.50	0.51	0.54	0.54
$13\pi/8$	0.52	0.51	0.58	0.58
$7\pi/4$	0.51	0.49	0.62	0.63
$15\pi/8$	0.49	0.48	0.58	0.58

- one multiplication by M that involves one scaling and one saxpy operation on vectors of length n_2 ($3\,n_2$ *flops*), one multiplication by S ($(2\,nz(S) - n_2)$ *flops*) and two applications of the preconditioner on the coarser level with n_2 nodes.

Letting $c(n)$ be the associated cost, one has thus, since $nz(A_{22}) = n_2$,

$$\begin{aligned} c(n) &= 4\,nz(A_{11}) + 2\,(nz(A_{12}) + nz(A_{21}) + n_2 + nz(S)) + 2\,c(n_2) \\ &= 2\,(nz(A_{11}) + nz(A) + nz(S) + c(n_2)) \end{aligned}$$

For five point matrices, $nz(A) \approx 5\,n$, $nz(A_{11}) \approx n_1 + 2\,n \approx 11n/4$, $nz(S) \approx 5\,n_2 \approx 5n/4$ and $n_2 \approx n/4$. Therefore,

$$c(n) \approx 18\,n + 2\,c\left(\frac{n}{4}\right)$$

Applying this formula recursively on j levels one gets

$$c(n) \approx 18\,n \cdot \left(\sum_{k=0}^{j-1} \left(\frac{1}{2}\right)^k\right) \approx 36\,n$$

The global cost for one preconditioner solution plus one multiplication by A is thus only about $45\,n$ *flops*, i.e. roughly only twice the cost that one would have with a mere ILU(0)-like preconditioner.

5. NUMERICAL RESULTS FOR VARIABLE FLOW

We again consider in this section the five-point up-wind finite difference approximation of the convection diffusion Equation (2) in the unit square, but now with variable flow. More specifically, we analyse the convergence of our multilevel preconditioner for two benchmark problems from the literature. In both cases, the flow has a rotating direction.

Problem 5.1. [23]. Here,

$$\overline{v}(x, y) = \begin{pmatrix} \cos\left(\pi\left(x - \frac{1}{3}\right)\right)\sin\left(\pi\left(y - \frac{1}{3}\right)\right) \\ -\cos\left(\pi\left(y - \frac{1}{3}\right)\right)\sin\left(\pi\left(x - \frac{1}{3}\right)\right) \end{pmatrix}$$

inside the circle of centre $(\frac{1}{3}, \frac{1}{3})$ and radius $\frac{1}{4}$, and $\overline{v}(x, y) = 0$ outside. The boundary conditions are $u = 0$ on the left, right and bottom boundaries, and $u = 1$ on the top boundary.

Since we are concerned here with a realistic convergence study (and no more with an assessment of the eigenvalue distribution), it is natural to use some Krylov accelerator. In the present case, we use GMRES with right preconditioning [13]. This method minimizes the residual and is therefore mathematically equivalent to several other generalized conjugate gradient methods [12].

We first considered a uniform grid with $h^{-1} = 128$ and different values for the restart parameter; for completeness, we also included the results for the stationary iterations (17) with $\tau = 1.5$. As the convergence may be exceptionally favourable for some specific right hand sides, besides the true right

Table V. Results for Problem 1 with $h^{-1} = 128$.

Method	ν: 1	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
	True right hand side					
Stationary	0.35	0.35	0.35	0.35	0.51	0.57
GMRES(rest.):						
(1)	0.32	0.32	0.32	0.33	0.44	0.50
(2)	0.23	0.23	0.23	0.26	0.41	0.48
(3)	0.21	0.21	0.21	0.25	0.35	0.40
(5)	0.20	0.20	0.20	0.24	0.34	0.38
(∞)	0.19	0.19	0.19	0.24	0.31	0.35
	Exact solution=random(0,1)					
Stationary	0.41	0.40	0.40	0.45	0.53	0.54
GMRES(rest.):						
(1)	0.33	0.33	0.33	0.38	0.46	0.49
(2)	0.24	0.23	0.24	0.33	0.43	0.46
(3)	0.21	0.21	0.22	0.33	0.39	0.39
(5)	0.20	0.20	0.21	0.32	0.38	0.38
(∞)	0.19	0.19	0.20	0.31	0.36	0.34
Reusken [23]		0.23	0.25	0.32	0.34	0.34

hand side of the discrete problem, we also considered an artificial right hand side generated in such a way that the solution vector is random. This latter choice also allows some comparison with Reusken's method from Reference [23] that have same structure as ours.

The results are given in Table V for our multilevel scheme with six levels (the coarsest grid is again 3×3). The quantity reported is the mean convergence factor for the number of iterations such that the relative residual error has been made less than 10^{-8}. The results for Reusken's method are retrieved from Reference [23]. Note that they correspond to the mean convergence factor for 20 iterations, independently of the actual convergence.

Looking at the results, it turns out that GMRES with very small restart (say 2 or 3) offers already a significant improvement over the stationary version, and delivers performances close to the optimal, non-restarted version. On the other hand, at least with a not too small value of the restart parameter, our method converges as fast as Reusken's method, although it is much less costly, since the latter uses nine point coarse grid matrices and up to six line Jacobi relaxations to approximate A_{11}^{-1}. Estimating the computational complexity as in the preceding section, this means that each application of this preconditioner requires more than four times the number of *flops* needed by ours. Moreover, according to Reference [39, p. 99], the number of Jacobi relaxations has in addition to be increased when the grid is refined, which is by no means surprising in light of the analysis in Reference [37]. For fairness, let us nevertheless mention that the focus in Reusken work [23] is on the coarse grid approximation, and that his method could probably be made more cost effective by integrating our approach to approximate A_{11}^{-1}.

Table VI. Results for Problem 1 with GMRES(2).

h^{-1}	ν = 1	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
	Mean convergence factors					
128	0.23	0.23	0.23	0.26	0.41	0.48
256	0.23	0.23	0.23	0.24	0.34	0.43
512	0.23	0.23	0.23	0.24	0.34	0.45
1024	0.23	0.23	0.23	0.23	0.33	0.37
	Number of iterations					
128	13	13	13	14	21	25
256	13	13	13	13	18	22
512	13	13	13	13	17	24
1024	13	13	13	13	17	19

We next checked if our method has grid independence convergence by considering uniform grids with decreasing mesh sizes. In each case, we used our multilevel method with a number of levels such that the coarsest grid is 3×3. For this experiment, we selected GMRES(2). We could as well have considered GMRES(3), which is probably slightly more efficient, or even the non-restarted version, which is feasible since the number of iterations never exceeds 25. However, GMRES(2) is maybe slightly more attractive as it is fairly cheap and nearly as easy to implement as stationary iterations (the minimization of the residual becomes subtle to implement with the need to allow arbitrary, possibly large, values of the restart parameter). Note nevertheless that with public domain software (e.g. [40]), Krylov subspace methods become quite user friendly.

The mean convergence factors and the number of iterations are reported in Table VI. Here again, the stopping criterion was $\|\mathbf{r}_k\| \leq 10^{-8}\|\mathbf{r}\|_0$. One may check that the convergence is grid independent or, even better, improves as the grid is refined. On the other hand, we observe a moderate deterioration for small ν, especially on not too fine grids. Taking into account the computational cost of each iteration, this deterioration does nevertheless not really spoil the cost efficiency of the method.

Problem 5.2. [41]. Here,

$$\overline{v}(x, y) = \begin{pmatrix} x(1-x)(2y-1) \\ -y(1-y)(2x-1) \end{pmatrix}$$

with the boundary conditions: $u = 0$ on the left, right and bottom boundaries, and $u = 1$ on the top boundary. The direction and magnitude of the convective flow are illustrated on Figure 2.

For this problem, a potential additional difficulty comes with the observation that discretization errors are likely to be reduced with a non-uniform grid. This is illustrated on Figure 3, where we have plotted the solution for $\nu = 10^{-5}$ together with a stretched grid. In the neighbourhood of the boundaries, this grid is refined in such a way that the ratio of maximum mesh size to minimum mesh size is equal to 200 and the ratio of subsequent mesh sizes is kept constant. One sees that this allows us to have several

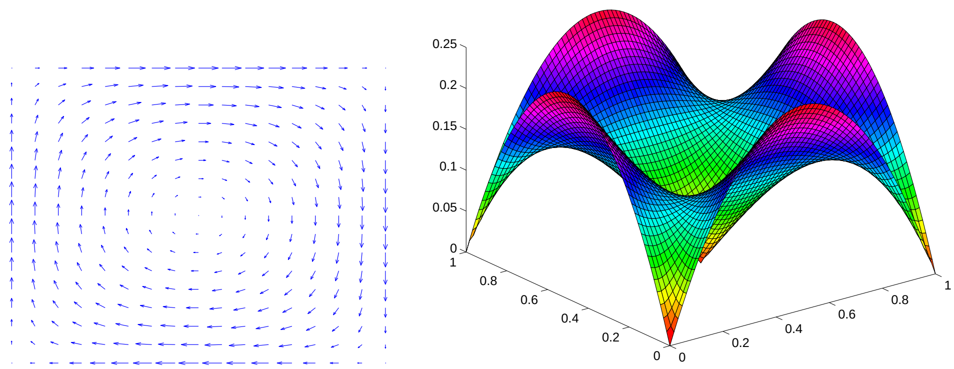

Figure 2. Direction and magnitude of the flow for Problem 2.

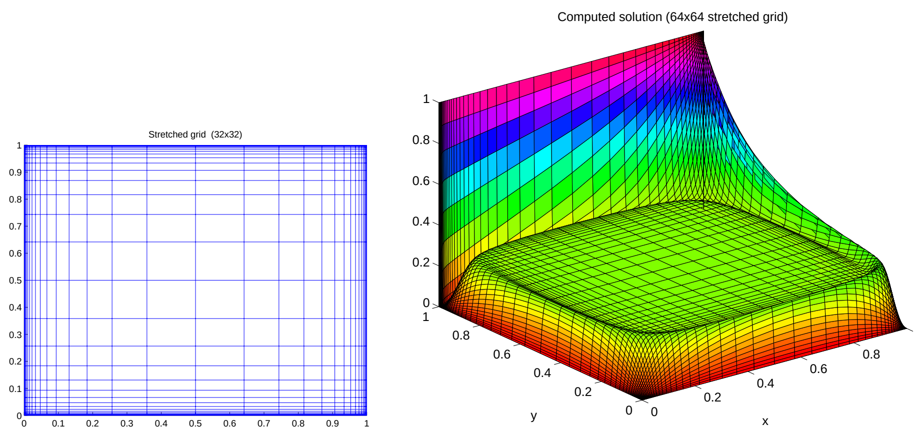

Figure 3. Stretched grid and computed solution for Problem 2.

grid points in the region where the solution varies greatly, which would not be the case with a uniform grid (the ratio of 200 was chosen so as to minimize the discretization error). Note that this stretching makes the discrete diffusion operator highly anisotropic in part of the domain. From the point of view of the iterative solution, this is the main difficulty introduced by non-uniform grids. We therefore hope that the results obtained with this example are representative of those one would obtain with specially (*a priori* or *a posteriori*) adapted meshes [42,6].

The results are reported in Table VI for the multilevel method with a number of levels such that the coarsest grid is 3×3. Here again, we used GMRES(2) and the stopping criterion was $\|\mathbf{r}_k\| \leq 10^{-8}\|\mathbf{r}\|_0$. One may see that the convergence is grid independent. In some cases, it is better for the stretched grid than for the corresponding uniform grid. More precisely, the convergence is more uniform with respect

Table VII. Results for Problem 2.

Grid	ν					
	1	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
	Uniform grid: mean convergence factors					
128×128	0.23	0.23	0.23	0.30	0.51	0.59
256×256	0.23	0.23	0.23	0.26	0.46	0.59
512×512	0.23	0.23	0.23	0.25	0.40	0.58
1024×1024	0.23	0.23	0.23	0.24	0.35	0.55
	Uniform grid: number of iterations					
128×128	13	13	13	16	28	35
256×256	13	13	13	14	24	36
512×512	13	13	13	14	20	34
1024×1024	13	13	13	14	18	31
	Stretched grid: mean convergence factors					
128×128	0.26	0.26	0.26	0.31	0.41	0.51
256×256	0.28	0.28	0.28	0.31	0.41	0.52
512×512	0.29	0.29	0.29	0.31	0.41	0.51
1024×1024	0.30	0.29	0.29	0.31	0.42	0.49
	Stretched grid: number of iterations					
128×128	14	14	14	16	21	28
256×256	15	15	15	16	21	28
512×512	15	15	15	16	21	27
1024×1024	16	15	15	16	22	26

to ν : slightly worse for diffusion dominated problems and better for convection dominated problems.

Now, from a mathematical point of view, the robustness requires the existence of a bound on the convergence rate that is independent of ν . We therefore further tested the method for smaller, maybe unphysical, values of this parameter. The results are reported in Table VIII for the uniform grid with $h^{-1} = 256$, which corresponds to the most difficult case according to Table VII (the picture for other mesh sizes is similar). For both GMRES and the stationary iterations (17) with $\tau = 1.5$, one may check that the convergence factor has indeed a useful limit for $\nu \to 0$. In fact, this behavior is in agreement with our theoretical results in Section 2, despite the limitations inherent in Fourier analysis: on the one hand, a bound exists that holds uniformly in ν , on the other hand (see Figure 1), this analysis is clearly too pessimistic for diffusion dominated problems, allowing a faster convergence in this case. In particular, the eigenvalues are then clustered around the real axis, which tends to make the Krylov accelerator more efficient, as is also seen from Table VIII.

Comparison with other preconditioners

Finally, we compared our multilevel method with the simpler SSOR, incomplete factorization (ILU) and modified incomplete factorization (MILU) preconditioners [12,13,33]. For the latter, we tested several level of fill-in, using the algorithm from Reference [13, p. 280] to determine which entries to accept.

Table VIII. Results for Problem 2, using a uniform grid with $h^{-1} = 256$.

Method	ν							
	1	10^{-2}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
	Mean convergence factor							
Stationary	0.35	0.35	0.59	0.62	0.65	0.66	0.66	0.66
GMRES(2)	0.23	0.23	0.46	0.59	0.62	0.63	0.63	0.63
GMRES(∞)	0.19	0.19	0.44	0.57	0.60	0.61	0.61	0.61
	Number of iterations							
Stationary	18	18	29	37	41	42	42	42
GMRES(2)	13	13	24	36	39	40	40	40
GMRES(∞)	12	12	23	33	37	37	37	37

For the five-point matrices considered here, this means that an (M)ILU(m) factorization generates about $2(m-1)$ fill entries (for $m \geq 3$) per row in each triangular factor; such a preconditioner requires therefore less storage than our multilevel scheme only for $m \leq 2$.

The timing results for Problem 2 with a uniform grid are reported in Table IX. The stopping criterion was $\|\mathbf{r}_k\| \leq 10^{-6}\|\mathbf{r}\|_0$, and the reported CPU time includes in all cases the time to compute the preconditioner; 'n.c.' (no convergence) means that the stopping criterion was still not satisfied after 500 iterations. This experiment was conducted on an Origin 2000 SGI system with 4 Gb RAM (enough to run all solvers in core). The program was written in mixed Fortran 77 & Fortran 90; no particular optimization effort was made, but we used the same program (in which the preconditioner is an option) to run all methods, i.e. the comparison is expected to be fair.

Clearly, ILU methods cannot compete with the multilevel preconditioner, even with generous fill-in. There are also less robust, since MILU strongly improves ILU for diffusion dominated problems (as expected from the general theory in the symmetric case [12,42]), but fails to converge when the convection dominates. One may think that the situation is likely to be improved by using more sophisticated schemes based on drop tolerance (see e.g. Reference [13]), but these methods require in general more preprocessing, and we observed that, for both mesh sizes considered, the multilevel algorithm solves the problem with $\nu = 10^{-5}$ in less time than is needed just to set up an ILU(10) factorization.

From a more general point of view, these results offer one more illustration that rotating convection–diffusion equations give rise to problems that are very difficult for many solvers (see References [39,44] for recent discussions of multigrid). It is therefore remarkable that our multilevel method performs nicely on such problems using exactly the same basic variant that was initially designed to solve symmetric positive definite systems. Looking further at Table IX, one may even see that GMRES acceleration is in fact not needed to this purpose, and that stationary iterations (17) performs nearly as well, if not better. In this case, apart from the need to select an appropriate relaxation parameter τ , the multilevel method is in fact used as a stand alone solver. The main merit of GMRES (or similar) acceleration is therefore to offer some additional robustness, since no parameter has to be supplied. Note however that in *all* numerical experiments, we used $\tau = 1.5$ without further optimization.

Table IX. CPU times to solve Problem 2 with various preconditioners.

	$h^{-1} = 128$			$h^{-1} = 512$		
ν:	10^{-1}	10^{-3}	10^{-5}	10^{-1}	10^{-3}	10^{-5}
Multilevel, Stationary	0.562	0.595	0.936	13.1	13.1	29.9
Multilevel, GMRES(2)	0.580	0.695	1.47	12.7	13.5	28.2
GMRES(10):						
SSOR	13.3	n.c.	n.c	n.c.	n.c.	n.c.
ILU(0)	9.46	n.c.	n.c.	n.c.	n.c.	n.c.
ILU(5)	1.90	5.22	4.10	170.	n.c.	n.c
ILU(10)	2.45	2.96	3.66	103.	414	578.
ILU(15)	4.89	5.55	5.99	126.	231.	390.
MILU(0)	1.47	4.45	n.c.	69.4	n.c.	n.c.
MILU(5)	1.12	1.87	n.c.	33.9	86.7	n.c.
MILU(10)	2.36	3.03	n.c.	51.8	78.3	n.c
MILU(15)	4.88	5.78	n.c.	98.6	120.9	n.c.
GMRES(20):						
SSOR	12.3	n.c.	n.c	n.c.	n.c.	n.c.
ILU(0)	8.77	n.c.	n.c	n.c.	n.c.	n.c.
ILU(5)	1.92	2.92	3.23	147.	n.c.	n.c.
ILU(10)	2.50	2.76	3.15	102.	314.	480.
ILU(15)	4.92	5.55	5.63	144.	172.	357.
MILU(0)	1.92	5.27	n.c.	76.8	n.c.	n.c.
MILU(5)	1.28	1.85	n.c.	38.2	95.4	n.c
MILU(10)	2.36	2.89	n.c.	56.0	80.2	n.c.
MILU(15)	4.86	5.68	n.c.	98.9	115.	n.c.

ACKNOWLEDGEMENTS

The numerical experiments with GMRES have been made thanks to CERFACS public domain software [40].

REFERENCES

1. Elman H, Silvester D, Wathen A. Iterative methods for problems in computational fluid dynamics. In *Iterative Methods in Scientific Computing* Chan R, Chan T, Golub G (eds). Springer-Verlag: Singapore, 1997; 271–327.
2. Berman A, Plemmons RJ. *Nonnegative Matrices in the Mathematical Sciences*. Academic Press: New York, 1979.
3. Hackbusch W. *Multi-grid Methods and Applications*. Springer: Berlin, 1985.
4. Axelsson O, Nikolova M. Avoiding slave points in adaptive refinement procedure for convection-diffusion problems in 2D. *Computing* 1998; **61**:331–357.
5. Hemker P. Mixed defect correction iteration for the solution of a singular perturbation problem. In *Defect Correction Methods: Theory and Applications*, Bohmer K, Stetter H (eds). Springer-Verlag: Berlin, 1984; 123–145.
6. Nikolova M, Axelsson O. Uniform in ε convergence of defect-correction method for convection-diffusion problems. Technical Report 9905, Department of Mathematics, Catholic University, Nijmegen, The Netherlands, 1999.
7. Elman H, Golub G. Line iterative methods for cyclically reduced convection-diffusion problems. *SIAM Journal on Scientific and Statistical Computing* 1992; **153**:339–363.

8. Bastian P, Hackbusch W, Wittum G. Additive and multiplicative multi-grid methods—a comparison. *Computing* 1998; **60**:345–364.
9. Hackbusch W, Probst T. Downwind Gauss–Seidel smoothing for convection dominated problems. *Numerical Linear Algebra with Applications* 1997; **4**:85–102.
10. de Zeeuw PM. Matrix-dependent prolongations and restrictions in a blackbox multigrid solver. *Journal of Computational and Applied Mathematics* 1990; **33**:1–27.
11. Stüben K. Algebraic multigrid (AMG): an introduction with applications. In *Multigrid*, Trottenberg U, Oosterlee C, Scüller A (eds). Academic Press, 1999; to be published.
12. Axelsson O. *Iterative Solution Methods*. Cambridge University Press: Cambridge, 1994.
13. Saad Y. *Iterative Methods for Sparse Linear Systems*. PWS Publishing: New York, 1996.
14. Axelsson O. The method of diagonal compensation of reduced matrix entries and multilevel iteration. *Journal of Computational and Applied Mathematics* 1991; **38**:31–43.
15. Axelsson O, Eijkhout V. Analysis of recursive 5-point/9-point factorization method. In *Preconditioned Conjugate Gradient Methods*, Axelsson O, Kolotilina L, (eds). Lectures Notes in Mathematics No. 1457, Springer-Verlag: Berlin 1990; 154–173.
16. 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.
17. Axelsson O, Neytcheva M. Algebraic multilevel iterations for Stieltjes matrices. *Numerical Linear Algebra with Applications* 1994; **1**:213–236.
18. Axelsson O, Vassilevski PS. Algebraic multilevel preconditioning methods—Part I. *Numerische Mathematik* 1989; **56**:157–177.
19. Axelsson O, Vassilevski PS. Algebraic multilevel preconditioning methods—Part II. *SIAM Journal on Numerical Analysis* 1990; **27**:1569–1590.
20. Notay Y. Optimal V cycle algebraic multilevel preconditioning. *Numerical Linear Algebra with Applications* 1998; **5**:441–459.
21. Notay Y. Optimal order preconditioning of finite difference matrices. *SIAM Journal on Scientific Computing* 1999; to be published.
22. Reusken A. Fourier analysis of a robust multigrid method for convection-diffusion equations. *Numerische Mathematik* 1995; **71**:365–398.
23. Reusken A. A multigrid method based on incomplete Gaussian elimination. *Numerical Linear Algebra with Applications* 1996; **3**:369–390.
24. Reusken A. On a robust multigrid solver. *Computing* 1996; **56**:303–322.
25. van der Ploeg A, Botta E, Wubs F. Nested grids ILU-decomposition (NGILU). *Journal of Computational and Applied Mathematics* 1996; **66**:515–526.
26. Vassilevski P. 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.
27. Vassilevski P. Hybrid V-cycle algebraic multilevel preconditioners. *Mathematics of Computation* 1992; **58**:489–512.
28. Vassilevski P. On two ways of stabilizing the hierarchical basis multilevel methods. *SIAM Review* 1997; **39**:18–53.
29. Wagner C, Kinzelbach W, Wittum G. Schur-complement multigrid. *Numerische Mathematik* 1997; **75**:523–545.
30. Axelsson O. On multigrid methods of the two-level type. In *Multigrid Methods*, Hackbusch W, Trottenberg U (eds), Lectures Notes in Mathematics No. 960. Springer-Verlag: Berlin, 1981; 352–367.
31. Axelsson O, Gustafsson I. Preconditioning and two-level multigrid methods of arbitrary degree of approximation. *Mathematics of Computation* 1983; **40**:214–242.
32. 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, USA, 1980.
33. Bank RE. Hierarchical bases and the finite element method. *Acta Numerica* 1996; **5**:1–43.
34. Bank RE, Dupont TF, Yserentant H. The hierarchical basis multigrid method. *Numerische Mathematik* 1988; **52**:427–458.
35. Yserentant H. On the multi-level splitting of finite element spaces. *Numerische Mathematik* 1986; **49**:379–412.
36. Notay Y. Polynomial acceleration of iterative schemes associated with subproper splittings. *Journal of Computational and Applied Mathematics* 1988; **24**:153–167.
37. Notay Y. Using approximate inverses in algebraic multilevel methods. *Numerische Mathematik* 1998; **80**:397–417.
38. Barrett R, Berry MW, Chan TF, Demmel J, Donato J, Dongarra J, Eijkhout V, Pozo R, Romine C and van der Vorst H. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. SIAM: Philadelphia, 1993.
39. Oosterlee CW, Washio T. An evaluation of parallel multigrid as a solver and a preconditioner for singularly perturbed problems. *SIAM Journal on Scientific Computing* 1998; **19**:87–110.
40. Frayssé V, Giraud L, Gratton S. A set of GMRES routines for real and complex arithmetic. Technical Report TR-PA-97-49, CERFACS, Toulouse, France, 1997. *http://www.cerfacs.fr/algor/Softs/GMRES/*

41. Elman H, Silvester D. Fast nonsymmetric iterations and preconditioning for Navier-Stokes equations. *SIAM Journal on Scientific Computing* 1996; **17**:33–46.
42. Miller J, O'Riordan E, Shishkin G. *Fitted Numerical Methods for Singular Perturbation Problems*. World Scientific: Singapore, 1996.
43. Notay Y. Upper eigenvalue bounds and related modified incomplete factorization strategies. In *Iterative Methods in Linear Algebra*, Beauwens R, de Groen P (eds). North-Holland: Amsterdam, 1992.
44. Yavneh I. Coarse-grid correction for nonelliptic and singular perturbation problems. *SIAM Journal on Scientific Computing* 1998; **19**:1682–1699.
45. Georgiev K. On an iterative solution of the convection diffuion problems. In *Iterative Methods in Linear Algebra II*, Margenov S, Vassilevski P (eds). Series in Computational and Applied Mathematics, vol. 3. IMACS 1996; 80–89.

