Numer. Math. 56, 157–177 (1989)

Numerische Mathematik

Algebraic Multilevel Preconditioning Methods. I

O. Axelsson and P.S. Vassilevski*
Department of Mathematics, Catholic University, Toernooiveld, 6525 ED Nijmegen, The Netherlands

Dedicated to the memory of Peter Henrici

Summary. A recursive way of constructing preconditioning matrices for the stiffness matrix in the discretization of selfadjoint second order elliptic boundary value problems is proposed. It is based on a sequence of nested finite element spaces with the usual nodal basis functions. Using a nodeordering corresponding to the nested meshes, the finite element stiffness matrix is recursively split up into two-level block structures and is factored approximately in such a way that any successive Schur complement is replaced (approximated) by a matrix defined recursively and therefore only implicitely given. To solve a system with this matrix we need to perform a fixed number (v) of iterations on the preceding level using as an iteration matrix the preconditioning matrix already defined on that level. It is shown that by a proper choice of iteration parameters it suffices to use $v>(1-\gamma^2)^{-\frac{1}{2}}$ iterations for the so constructed v-fold V-cycle (where $v=2$ corresponds to a W-cycle) preconditioning matrices to be spectrally equivalent to the stiffness matrix. The conditions involve only the constant γ in the strengthened C.-B.-S. inequality for the corresponding two-level hierarchical basis function spaces and are therefore independent of the regularity of the solution for instance. If we use successive uniform refinements of the meshes the method is of optimal order of computational complexity, if $\gamma^2<\frac{8}{9}$. Under reasonable assumptions of the finite element mesh, the condition numbers turn out to be so small that there are in practice few reasons to use an accelerated iterative method like the conjugate gradient method, for instance.

Subject Classifications: AMS(MOS): 65F10, 65N20, 65N30; CR: G1.3.

1 Introduction

The preconditioned conjugate gradient method is one of the most efficient linear algebraic solver for problems with symmetric positive definite matrices. The

* On leave from Institute of Mathematics and from Center of Informatics and Computer Technology, Bulgarian Academy of Sciences, Sofia, Bulgaria. The research of the second author reported here was supported in part by the Committee of Science, Bulgaria, under Grant No. 55/26.03.87

numerical solution of elliptic selfadjoint second order boundary value problems by the finite element method leads to certain classes of linear algebraic problems with symmetric positive definite matrices, which may be very large but sparse. As is well-known, the rate of convergence of the preconditioned *CG* method depends on the properties of the preconditioning matrix. For example, if the preconditioning matrix is spectrally equivalent (see for instance Axelsson and Barker [1]) to the original matrix, the rate of convergence will be independent of the size of problem (or equivalently, of the discretization parameter h). Thus the task of deriving efficient preconditioning matrices is essential. One way to solve this task is by a multigrid method. In the best cases it even does not require any acceleration by the conjugate gradient method and as is well-known, the resulting computational complexity is of optimal or nearly optimal order. However, this is only true when the differential equation problem is regular enough and the discretization, e.g., the triangulation of the region, is almost uniform, cf. Bank and Dupont [5], Braess and Hackbusch [9]. It is not clear how to choose the parameters in the multigrid methods, such as in smoothing matrices and the prolongation and restriction operators in more general situations, in particular when many singularities due for instance to boundary or interior corners (where the latter may result from discontinuous coefficients in the differential operator) are present. Hence it is of importance to construct methods with a small convergence factor *utilizing purely algebraic means*, i.e. not based on any regularity assumptions.

There are two ways already studied to derive nearly optimal preconditioning matrices both based on nested discretizations. The first one proposed in Yserentant [14] and generalized in Bank et al. [6] uses the so-called hierarchical basis functions, whereas the second one, proposed in Vassilevski [13] uses the standard nodal basis functions and the multilevel ordering of the nodes of the triangulation of the considered polygonal domain and the preconditioning matrix is constructed by approximate block-factorization of the stiffness matrix with respect to this multilevel ordering. Both methods are proven to be nearly optimal. (The condition number for two-dimensional boundary value problems increases as $O(\log h)$ or $O((\log h)^2)$, where h is a meshsize parameter.) The proofs are based on a classical interpolation estimate of the L^∞-norm of the discrete functions (for a recent proof of this, see Yserentant [14]). These methods are extensions of the previously studied two-level iterative methods, see Bank and Dupont [4], Braess [8], Axelsson and Gustafsson [2], and Axelsson [3].

However, the interpolation estimate used for the hierarchical basis function method is true only for problems in two space dimensions. For three-dimensional problems one finds that the condition number increases as $O(h^{-1})$, $h \to 0$. In our paper this estimate is not required because we use standard nodal basis functions and a more accurate approximation of Schur complements. The present method is an extension and modification of the factorization method used in Vassilevski [13], where the successive Schur complements are replaced, i.e. approximated, by the preconditioner of the stiffness matrix at the current level. Here we propose to use more accurate approximations to these Schur complements by matrices, which are implicitly given and recursively defined. This recursion uses an iterative procedure on each level with as iteration matrix the precon-

ditioning matrix already defined on the preceding level. If this iterative procedure consists of only one step we get the multilevel preconditioning matrix already studied in Vassilevski [13]. The preconditioning method derived when two such recursive iterations are used we call W-cycle preconditioning method, because of its similarity to the W-cycle for the classical multigrid iteration. As in the two-level iterative methods referred to above, the only technical means we use is the strengthened Cauchy-Bunyakowski-Schwarz (C.-B.-S.) inequality.

We distinghuish between two versions. In version (i) we approximate the Schur complements by a matrix polynomial of degree v, which depends on the exact Schur complement and the preconditioner on the (next) coarse level. In version (ii) the Schur complements is approximated by a matrix polynomial which depends on the stiffness matrix and the preconditioner on the (next) coarse level.

In both versions we solve the systems with the stiffness matrix block part corresponding to the new mesh nodes exactly. This matrix has a condition number bounded independently on the fineness of the mesh, so it requires only a number of arithmetric operations proportional to the number of nodes to solve this system to machine number precision by a conjugate gradient method, for instance. In a following report, we shall consider the case where also this matrix is approximated. In special cases, this matrix is even diagonal.

We derive sufficient conditions in the form of upper bounds of the constant γ in the strengthened C.-B.-S. inequality for the v-fold ($v \geqq 2$) preconditioning matrices to be spectrally equivalent to the original stiffness matrix. The upper bounds depend on the constant γ and on the choice and degree of matrix polynomials approximating the Schur complements. The most efficient choices are based on Chebyshev polynomials as we shall see.

The computational complexity depends on the polynomial degree v, and on the manner we recursively refine the mesh. For a two-dimensional problem, the computational complexity will be of optimal order, i.e. proportional to the number of node points on the finest mesh, if $v \leqq 3$ (it will be of optimal order, save a factor $O(\log h)$, if $v = 4$), and if we use a uniform refinement of the triangulation, dividing each triangle into four congruent ones. For a three-dimensional problem, it will be of optimal order if $v \leqq 7$ if we use a corresponding uniform refinement.

Some particular choices of finite element spaces are considered such as piecewise linear basis functions on triangular elements, for which these sufficient conditions are verified. The spectral equivalence bounds are independent of the coefficients in the diffusion problem, if these are piecewise constant (constant on each element of the coarse mesh).

The remainder of the paper is organized as follows.

In the second part we give some preliminary facts and derive two versions of constructing v-fold V-cycle preconditioning matrices. In the third section we state and prove the main results about the relative condition numbers of these preconditioning matrices with respect to the original stiffness matrix. In the fourth part we study the space of piecewise linear and piecewise quadratic basis

functions on triangular elements. The techniques presented can be used also for elliptic problems in three space dimensions. However, in the present paper we consider only plane polygonal domains.

2 Derivation of the ν-Fold V-Cycle Preconditioning Matrices

We consider the variational formulation of a second order selfadjoint elliptic boundary value problem: Find $u \in \mathring{W}_2^1(\Omega)$, such that

$$A(u, v) = (f, v), \quad \text{all } v \in \mathring{W}_2^1(\Omega). \tag{2.1}$$

We consider Ω a plane polygonal domain. Further

$$A(u, v) := \int_\Omega \sum_{i,j} a_{i,j}(x) \frac{\partial}{\partial x_i} u \frac{\partial}{\partial x_i} v \, dx \tag{2.1a}$$

is assumed to be symmetric, bounded and $\mathring{W}_2^1(\Omega)$-elliptic and

$$(f, v) := \int_\Omega f v \, dx.$$

Let τ_1 be some (coarse) initial triangulation of Ω. By a refining procedure we obtain a sequence of triangulations τ_k, $k = 1, 2, \ldots, l$. For triangular elements local refinement is allowed, for example, by connecting the vertices of a triangle with the midpoints of the opposite edges of that triangle. Finally with any triangulation τ_k we associate the corresponding finite element space V_k of piecewise polynomial functions, that are continuous in $\bar{\Omega}$. In any V_k we choose the nodal basis functions used most frequently in practial computations, $\{\phi_i^{(k)}\}_{i=1}^{n_k}$, where n_k is the number of nodepoints in τ_k. If N_k is the set of the nodes in τ_k and x_j runs over all these nodes then

$$\phi_i^{(k)}(x_j) = \delta_{i,j}$$

(the Kronecker function). Once having a basis the following sequence of stiffness matrices can be computed

$$A^{(k)} = \{A(\phi_i^{(k)}, \phi_j^{(k)})\}_{i,j=1}^{n_k}, \quad k = 1, 2, \ldots, l.$$

By, construction we have

$$N_{k+1} \supset N_k.$$

Hence at the $k+1$'th level the partitioning $N_{k+1} \setminus N_k$ and N_k of the nodes in N_{k+1} can be used. Corresponding to this ordering, $A^{(k+1)}$ takes the following two by two block structure,

$$A^{(k+1)} = \begin{bmatrix} A_{11}^{(k+1)} & A_{12}^{(k+1)} \\ A_{21}^{(k+1)} & A_{22}^{(k+1)} \end{bmatrix},$$

where

$$A_{11}^{(k+1)}=\{A(\phi_i^{(k+1)},\phi_j^{(k+1)})\}_{i,j:x_i,x_j\in N_{k+1}\setminus N_k},$$
$$A_{12}^{(k+1)}=\{A(\phi_j^{(k+1)},\phi_i^{(k+1)})\}_{i,j\ x_i\in N_{k+1}\setminus N_k,x_j\in N_k},$$
$$A_{22}^{(k+1)}=\{A(\phi_i^{(k+1)},\phi_j^{(k+1)})\}_{i,j:x_i,x_j\in N_k}.$$

As $V_k\subset V_{k+1}$ by construction we may alternatively use in V_{k+1} the so-called two-level basis functions (see [4, 2, 3 and 7]),

$$\{\phi_i^{(k+1)},\ x_i\in N_{k+1}\setminus N_k \text{ and } \phi_i^{(k)},\ x_i\in N_k\}.$$

Then any function $v\in V_{k+1}$ can be expanded by using these two bases i.e. we have

$$v(x)=\sum_{i=1}^{n_{k+1}} v_i\,\phi_i^{(k+1)} \qquad (v_i=v(x_i))$$
$$=\sum_{i:x_i\in N_{k+1}\setminus N_k}\hat{v}_i\,\phi_i^{(k+1)}+\sum_{i:x_i\in N_k} v_i\,\phi_i^{(k)}.$$

This expression defines a mapping $J(=J_{k+1})$, which transforms any coefficient vector $\underline{\hat{v}}=\begin{bmatrix}\underline{v}_1\\ \underline{v}_2\end{bmatrix}$ of the representation of a function $v\in V_{k+1}$ with respect to the two-level basis to the coefficient vector $\underline{v}=(v_i)$ of the representation of v in the nodal basis of V_{k+1}.

In matrix notations J has the form

$$J=\begin{bmatrix}I & J_{12}\\ 0 & I\end{bmatrix}.$$

We assume here that the nodes from $N_{k+1}\setminus N_k$ are ordered first and then the nodes from N_k.

We may compute also the two-level stiffness matrices $\bar{A}^{(k+1)}$ for the two-level basis functions. If $\bar{A}^{(k+1)}$ is partitioned in the same manner as $A^{(k+1)}$ into the two by two block structure, $(\bar{A}_{i,j}^{(k+1)})_{i,j=1}^2$, then we have the following factorizations

$$\bar{A}^{(k+1)}=\begin{bmatrix}\bar{A}_{11}^{(k+1)} & 0\\ \bar{A}_{21}^{(k+1)} & \bar{S}^{(k+1)}\end{bmatrix}\begin{bmatrix}I & \bar{A}_{11}^{(k+1)^{-1}}\bar{A}_{12}^{(k+1)}\\ 0 & I\end{bmatrix},$$
$$A^{(k+1)}=\begin{bmatrix}A_{11}^{(k+1)} & 0\\ A_{21}^{(k+1)} & S^{(k+1)}\end{bmatrix}\begin{bmatrix}I & A_{11}^{(k+1)^{-1}}A_{12}^{(k+1)}\\ 0 & I\end{bmatrix},$$

where $\bar{S}^{(k+1)}$ and $S^{(k+1)}$ are the Schur complements

$$\bar{S}^{(k+1)}=\bar{A}_{22}^{(k+1)}-\bar{A}_{21}^{(k+1)}\bar{A}_{11}^{(k+1)^{-1}}\bar{A}_{12}^{(k+1)},$$
$$S^{(k+1)}=A_{22}^{(k+1)}-A_{21}^{(k+1)}A_{11}^{(k+1)^{-1}}A_{12}^{(k+1)}.$$

Note that these are positive definite. Based on the identity

$$\bar{A}^{(k+1)} = J^t A^{(k+1)} J$$

a straightforward computation (see Vassilevski [13]) shows that

$$\bar{S}^{(k+1)} = S^{(k+1)}. \tag{2.2'}$$

Note further that $\bar{A}_{22}^{(k+1)} = A^{(k)}$ and that $\bar{A}_{11}^{(k+1)} = A_{11}^{(k+1)}$. Then the following result from Axelsson [3], Axelsson and Gustafsson [2] can be reformulated as follows:

Lemma 2.1. *There exists a constant* γ, $0 < \gamma < 1$, *such that*

$$1 \leq \frac{\underline{v}_2^t A^{(k)} \underline{v}_2}{\underline{v}_2^t S^{(k+1)} \underline{v}_2} \leq (1-\gamma^2)^{-1} \quad \text{all } \underline{v}_2 \in \mathbf{R}^{n_k}.$$

Here γ is the constant in the strengthened C.-B.-S. inequality used in the two-level methods studied in Bank and Dupont [4], Axelsson and Gustafsson [2], Braess [8], Axelsson [3] and Maitre and Musy [12]. Note that the inequalities are sharp. In [4] and [2] it has further been shown that the spectral condition number of $A_{11}^{(k+1)}$ is bounded above by a constant independent of the fineness of the mesh. The constant γ can be computed locally from the local element matrices. Therefore γ is independent of the diffusion coefficients $a_{i,j}$ in (2.1a) if these are constant on the elements of the coarsest triangulation. In our particular case, the strengthended C.-B.-S. inequality reads

$$\begin{aligned} |\underline{v}_2^t A_{21}^{(k+1)} \underline{v}_1| = |\underline{v}_1^t A_{12}^{(k+1)} \underline{v}_2| &= \left| (\underline{v}_1^t, 0) A^{(k+1)} \begin{bmatrix} 0 \\ \underline{v}_2 \end{bmatrix} \right| \\ &\leq \gamma (\underline{v}_1^t A_{11}^{(k+1)} \underline{v}_1)^{\frac{1}{2}} (\underline{v}_2^t \bar{A}_{22}^{(k+1)} \underline{v}_2)^{\frac{1}{2}} \\ &= \gamma (\underline{v}_1^t A_{11}^{(k+1)} \underline{v}_1)^{\frac{1}{2}} (\underline{v}_2^t A^{(k)} \underline{v}_2)^{\frac{1}{2}} \quad \text{all } \underline{v}_1 \in \mathbf{R}^{n_{k+1}-n_k},\ \underline{v}_2 \in \mathbf{R}^{n_k},\ \underline{v}^t = [\underline{v}_1^t, \underline{v}_2^t]. \end{aligned} \tag{2.2}$$

We present now the W-cycle, or more generally the ν-fold V-cycle preconditioning matrices, where $\nu = 2$ corresponds to a W-cycle. In version (i) it takes the following form. Let $M^{(1)} = A^{(1)}$ and for $k = 1$ to $l-1$ define

$$M^{(k+1)} = \begin{bmatrix} A_{11}^{(k+1)} & 0 \\ A_{21}^{(k+1)} & \tilde{A}^{(k)} \end{bmatrix} \begin{bmatrix} I & A_{11}^{(k+1)^{-1}} A_{12}^{(k+1)} \\ 0 & I \end{bmatrix} \tag{2.3}$$

where, with $S = S^{(k+1)}$,

$$\tilde{A}^{(k)^{-1}} = (I - P_\nu(M^{(k)^{-1}} S)) S^{-1}. \tag{2.4}$$

P_ν is a polynomial of degree $\nu \geqq 1$ and satisfies the following conditions:

$$0 \leqq P_\nu(t) < 1, \quad 0 < t \leqq 1, \quad P_\nu(0) = 1. \tag{2.5}$$

We let

$$Q_{\nu-1}(t) = (1 - P_\nu(t))/t. \tag{2.6}$$

Note that $Q_{\nu-1}$ is a polynomial of degree $\nu-1$,

$$Q_{\nu-1}(t)>0, \quad 0<t<1 \quad \text{and} \quad Q_{\nu-1}(0)=-P_\nu'(0).$$

$M^{(k+1)}$ shall be used as a preconditioner of $A^{(k+1)}$. For $\nu=1$ and $P_\nu(t)=1-t$ we get the preconditioner

$$M^{(k+1)}=\begin{bmatrix} A_{11}^{(k+1)} & 0 \\ A_{21}^{(k+1)} & M^{(k)} \end{bmatrix}\begin{bmatrix} I & A_{11}^{(k+1)-1}A_{12}^{(k+1)} \\ 0 & I \end{bmatrix}$$

studied in Vassilevski [13].

This method is also similar to the method used by Braess [7] for a multilevel method for a special difference scheme and to the method of Kuznetsov [11].

Examples of polynomials satisfying (2.5) are:

Example 2.1. $P_\nu(t)=(1-t)^\nu$. Here $Q_{\nu-1}(t)=1+(1-t)+\ldots+(1-t)^{\nu-1}$.

Example 2.2. P_ν satisfies (2.5) and has the smallest local maximum in the interval $[\alpha, 1]$, $0<\alpha<1$. This gives

$$P_\nu(t)=\left[T_\nu\left(\frac{1+\alpha-2t}{1-\alpha}\right)+1\right]\Big/\left[T_\nu\left(\frac{1+\alpha}{1-\alpha}\right)+1\right],$$

where T_ν is the Chebyshev polynomial,

$$T_\nu(x)=2xT_{\nu-1}(x)-T_{\nu-2}(x), \qquad \nu=2, 3, \ldots, T_0(x)=1, \; T_1(x)=x.$$

If $\alpha=1$ we get the polynomial in Example 2.1.

Note that $P_\nu(\alpha)=2\Big/\left[T_\nu\left(\frac{1+\alpha}{1-\alpha}\right)+1\right]$ and $P_\nu(1)=[(-1)^\nu+1]\Big/\left(T_\nu\left(\frac{1+\alpha}{1-\alpha}\right)+1\right)$.

The following example is similar but we require in addition that $P_\nu'(1)=-1$ and ν is odd. This extra condition will determine $\alpha=\alpha_\nu$.

Example 2.3. ν odd, P_ν satisfies (2.5), $P_\nu'(1)=-1$ and, for $\nu\geq 3$, P_ν has the smallest local maximum at its local extreme point(s) in (0, 1). This leads us to

$$P_\nu(t)=\left[T_\nu\left(\frac{1+\alpha_\nu-2t}{1-\alpha_\nu}\right)+1\right]\Big/\left[T_\nu\left(\frac{1+\alpha_\nu}{1-\alpha_\nu}\right)+1\right],$$

where $0<\alpha_\nu<1$ will be determined by the condition $P_\nu'(1)=-1$. Since

$$P_\nu'(t)=-\frac{2}{1-\alpha_\nu}T_\nu'\left(\frac{1+\alpha_\nu-2t}{1-\alpha_\nu}\right)\Big/\left[T_\nu\left(\frac{1+\alpha_\nu}{1-\alpha_\nu}\right)+1\right]$$

and $T_\nu'(-1)=\nu^2$ (ν odd), we get

$$\frac{2\nu^2}{1-\alpha_\nu}\left[T_\nu\left(\frac{1+a_\nu}{1-\alpha_\nu}\right)+1\right]^{-1}=1$$

or, after a simplification,

$$2v(1-\alpha_v)^{\frac{v-1}{2}}/[(1+\alpha_v^{\frac{1}{2}})^v+(1-\alpha_v^{\frac{1}{2}})^v]=1.$$

For $v=3$ and $v=5$ we get $\alpha_v=v^{-1}$ and

$$P_3(t)=1-5t+8t^2-4t^3=(1-t)(1-2(1-t))^2,$$
$$P_5(t)=1-11t+45t^2-85t^3+75t^4-25t^5$$
$$=(1-t)(1-10(1-t)+35(1-t)^2-50(1-t)^3+25(1-t)^4).$$

As $v\to\infty$ (v odd) one finds $\alpha_v\sim\dfrac{(\ln 4v^2)^2}{4v^2}$.

Implementational Details

In an iterative method with $M^{(l)}$ as a preconditioner, we need to solve linear systems with $M^{(k+1)}$. This requires two solutions with matrix $A_{11}^{(k+1)}$ and one with matrix $\tilde{A}^{(k)}$. Since the condition number of $A_{11}^{(k+1)}$ is $O(1)$, independent of k (and hence of the fineness of the mesh), we can solve $A_{11}^{(k+1)}$ by an iterative method in an optimal order of computational complexity to machine number precision, if we want.

To solve linear systems with $\tilde{A}^{(k)}$, i.e. to compute $\tilde{A}^{(k)^{-1}}\underline{v}_2$ for some vectors $\underline{v}_2\in\mathbf{R}^{n_k}$, we note that by (2.4) and (2.6)

$$\underline{y}_2=\tilde{A}^{(k)^{-1}}\underline{v}_2=Q_{v-1}(M^{(k)^{-1}}S)\,M^{(k)^{-1}}\underline{v}_2. \tag{2.7}$$

Let $Q_{v-1}(t)=q_0+q_1t+\ldots+q_{v-1}t^{v-1}$. Then the matrix-vector product (2.7) can be evaluated by performing v iterations of the following procedure,

$$\underline{y}_2^{(0)}=0,\qquad \text{for } r=1 \text{ until } v \text{ solve}$$
$$M^{(k)}\underline{y}_2^{(r)}=q_{v-r}\underline{v}_2+S\underline{y}_2^{(r-1)}. \tag{2.8}$$

Then $\underline{y}_2=\underline{y}_2^{(v)}$.

Hence, to compute $\underline{y}_2$ we need to solve v linear systems with $M^{(k)}$, the preconditioning matrix of the preceding level and in addition to compute matrix-vector multiplications with matrix $S^{(k+1)}$, $v-1$ times, i.e. in particular to solve additional linear systems with $A_{11}^{(k+1)}$, $v-1$ times.

We consider now version (ii), where in (2.4) we replace S with $A^{(k)}$, and we let

$$\tilde{A}^{(k)^{-1}}=(I-P_v(M^{(k)^{-1}}A^{(k)}))\,A^{(k)^{-1}}. \tag{2.9}$$

This method can be implemented as (2.8) but with S replaced with $A^{(k)}$, i.e.

$$y_2^{(0)}=0, \quad \text{for } r=1 \text{ until } \nu \text{ solve}$$
$$M^{(k)}\, y_2^{(r)}=q_{\nu-r}\, \underline{v}_2+A^{(k)}\, \underline{y}_2^{(r-1)}. \tag{2.10}$$

Accordingly, in version (ii) we do not have to solve the additional linear systems with $A_{11}^{(k+1)}$ as in the matrix recursion (2.8).

3 The Relative Condition Number of $M=M^{(l)}$ with Respect to $A=A^{(l)}$

We shall derive a recurrence relation between certain upper bounds of the relative condition numbers of $M^{(k)}$ with respect to $A^{(k)}$, $k=2, 3, \ldots, l$. Consider first version (i):

Note then that by (2.3),

$$M^{(k+1)}=\begin{bmatrix} A_{11}^{(k+1)} & A_{12}^{(k+1)} \\ A_{21}^{(k+1)} & (\tilde{A}^{(k)}+A_{21}^{(k+1)}\, A_{11}^{(k+1)^{-1}}\, A_{12}^{(k+1)}) \end{bmatrix}=A^{(k+1)}+\begin{bmatrix} 0 & 0 \\ 0 & (\tilde{A}^{(k)}-S^{(k+1)}) \end{bmatrix}. \tag{3.1}$$

Hence we have

$$\underline{v}^t(M^{(k+1)}-A^{(k+1)})\, \underline{v}=\underline{v}_2^t(\tilde{A}^{(k)}-S^{(k+1)})\, \underline{v}_2. \tag{3.2}$$

Lemma 3.1. *Let $\tilde{A}^{(k)}$ be defined by* (2.4) *and $Q_{\nu-1}$ by* (2.6). *Then with $S=S^{(k+1)}$,*

$$\begin{aligned} S^{\frac{1}{2}}\, \tilde{A}^{(k)^{-1}}\, S^{\frac{1}{2}} &= I-P_\nu(S^{\frac{1}{2}}\, M^{(k)^{-1}}\, S^{\frac{1}{2}}) \\ &= S^{\frac{1}{2}}\, M^{(k)^{-1}}\, S^{\frac{1}{2}}\, Q_{\nu-1}\,(S^{\frac{1}{2}}\, M^{(k)^{-1}}\, S^{\frac{1}{2}}). \end{aligned}$$

Proof. Since S is a Schur complement of a positive definite matrix, it is itself positive definite, so $S^{\frac{1}{2}}$ exists. Lemma 3.1 follows now readily by pre- and post-multiplication of (2.4) by $S^{\frac{1}{2}}$ and using (2.6). □

Lemma 3.2. a) *Let $T^{(k)}=S^{\frac{1}{2}}\, M^{(k)^{-1}}\, S^{\frac{1}{2}}$, where $S=S^{(k+1)}$, let*

$$\lambda_{k+1}=\sup_{\underline{v}} \frac{\underline{v}^t M^{(k+1)}\, \underline{v}}{\underline{v}^t A^{(k+1)}\, \underline{v}} \qquad \forall \underline{v}=\begin{bmatrix} \underline{v}_1 \\ \underline{v}_2 \end{bmatrix} \in \mathbf{R}^{n_{k+1}},$$

and let

$$t_k=\inf_{\underline{v}_2}\{\underline{v}_2^t\, T^{(k)}\, \underline{v}_2/\underline{v}_2^t\, \underline{v}_2\}, \qquad \underline{v}_2 \in \mathbf{R}^{n_k}.$$

Then

$$(1-\gamma^2)/\lambda_k \leqq t_k \leqq \underline{v}_2^t\, T^{(k)}\, \underline{v}_2/\underline{v}_2^t\, \underline{v}_2 \leqq 1 \qquad \forall \underline{v}_2 \in \mathbf{R}^{n_k};$$

b) *With λ_k and t_k defined as in part* a) *we have*

$$\lambda_{k+1}=\sup_{\underline{v}_2}\{\underline{v}_2^t\, \tilde{A}^{(k)}\, \underline{v}_2/\underline{v}_2^t\, S\underline{v}_2\}=(1-P_\nu(t_k^*))^{-1}=(t_k^*\, Q_{\nu-1}(t_k^*))^{-1},$$

where $P_v(t_k^*)=\max P_v(t_k^*)=\max P_v(t)$, $t_k \leqq t \leqq 1$.

Proof. To prove part a) note first that by (2.5) and Lemma 3.1 we have

$$0 \leqq \underline{v}_2^t S^{(k+1)} \underline{v}_2 \leqq \underline{v}_2^t \tilde{A}^{(k)} \underline{v}_2 . \tag{3.3}$$

In particular, by (3.1) it now follows

$$\underline{v}^t M^{(k+1)} \underline{v} \geqq \underline{v}^t A^{(k+1)} \underline{v} \qquad \forall \underline{v} \in \mathbf{R}^{n_{k+1}}, \tag{3.3'}$$

so $\lambda_{k+1} \geqq 1$.

Further, by Lemma 2.1 it follows,

$$t_k = \inf_{\underline{v}_2} \underline{v}_2^t S \underline{v}_2 / \underline{v}_2^t M^{(k)} \underline{v}_2 \geqq \inf_{\underline{v}_2} \frac{\underline{v}_2^t S \underline{v}_2}{\underline{v}_2^t A^{(k)} \underline{v}_2} \inf_{\underline{v}_2} \frac{\underline{v}_2^t A^{(k)} \underline{v}_2}{\underline{v}_2^t M^{(k)} \underline{v}_2} \geqq (1-\gamma^2) \lambda_k^{-1} .$$

Finally (2.2') and (3.3') show that

$$\frac{\underline{v}_2^t S \underline{v}_2}{\underline{v}_2^t M^{(k)} \underline{v}_2} \leqq \frac{\underline{v}_2^t S \underline{v}_2}{\underline{v}_2^t A^{(k)} \underline{v}_2} \frac{\underline{v}_2^t A^{(k)} \underline{v}_2}{\underline{v}_2^t M^{(k)} \underline{v}_2} = \frac{\underline{v}_2^t \bar{S} \underline{v}_2}{\underline{v}_2^t A^{(k)} \underline{v}_2} \frac{\underline{v}_2^t A^{(k)} \underline{v}_2}{\underline{v}_2^t M^{(k)} \underline{v}_2} \leqq 1 \qquad \forall \underline{v}_2 \in \mathbf{R}^n ,$$

so

$$\sup_{\underline{v}_2} \underline{v}_2^t T^{(k)} \underline{v}_2 / \underline{v}_2^t \underline{v}_2 \leqq 1 . \quad \square$$

To prove part b) we show first that

$$\lambda_{k+1} = \sup_{\underline{v}} \{ \underline{v}^t M^{(k+1)} \underline{v} / \underline{v}^t A^{(k+1)} \underline{v} \}$$

is taken on the subspace

$$H = \{ \underline{v} : A_{11}^{(k+1)} \underline{v}_1 + A_{12}^{(k+1)} \underline{v}_2 = 0 \}.$$

To this end we note that a well-known property of Schur complements states that

$$\inf_{\underline{v}} \underline{v}^t A \underline{v} = \inf_{\underline{v}_2} \underline{v}_2^t S \underline{v}_2 .$$

Further it follows from (3.1) that

$$\begin{aligned} \underline{v}^t M^{(k+1)} \underline{v} &= \underline{v}^t (M^{(k+1)} - A^{(k+1)}) \underline{v} + \underline{v}^t A^{(k+1)} \underline{v} \\ &= \underline{v}_2^t (\tilde{A}^{(k)} - S) \underline{v}_2 + \underline{v}^t A^{(k+1)} \underline{v} . \end{aligned}$$

Hence

$$\begin{aligned} \lambda_{k+1} &= \sup_{\underline{v}} \frac{\underline{v}^t M^{(k+1)} \underline{v}}{\underline{v}^t A^{(k+1)} \underline{v}} = 1 + \sup_{\underline{v}} \frac{\underline{v}_2^t (\tilde{A}^{(k)} - S) \underline{v}_2}{\underline{v}^t A^{(k+1)} \underline{v}} \\ &= 1 + \sup_{\underline{v}_2} \frac{\underline{v}_2^t (\tilde{A}^{(k)} - S) \underline{v}_2}{\underline{v}_2^t S \underline{v}_2} \end{aligned} \tag{3.4}$$

which shows the first part of Lemma 3.2b.

The last part follows by (2.4), (3.2) and Lemma 3.2:

$$\lambda_{k+1} = \sup_{\underline{v}_2} \underline{v}_2^t S((I - P_\nu(M^{(k)^{-1}} S))^{-1} \underline{v}_2 / \underline{v}_2^t S \underline{v}_2$$

$$= \sup_{\underline{v}_2} \{(S^{\frac{1}{2}} \underline{v}_2)^t (I - P_\nu(T^{(k)}))^{-1} S^{\frac{1}{2}} \underline{v}_2 / \underline{v}_2^t S \underline{v}_2$$

$$= (1 - P_\nu(t_k^*))^{-1} = (t_k^* Q_{\nu-1}(t_k^*))^{-1}. \quad \square$$

Majorizing Sequence

Let $\hat{\lambda}_k, \hat{t}_k$ be defined by $\hat{t}_k = (1-\gamma^2)/\hat{\lambda}_k$, $\hat{\lambda}_{k+1} = (1 - P_\nu(\hat{t}_k^*))^{-1}$ where

$$P_\nu(\hat{t}_k^*) = \max P_\nu(t), \qquad \hat{t}_k \leqq t \leqq 1.$$

Then, it follows from Lemma 3.2b that $\lambda_{k+1} \leqq \hat{\lambda}_{k+1}$, i.e. $\{\hat{\lambda}_k\}$ is a majorizing sequence of $\{\lambda_k\}$.

Example 2.1. Consider now Example 2.1, where $P_\nu(t) = (1-t)^\nu$ and $Q_{\nu-1} = 1 + (1-t) + \ldots (1-t)^{\nu-1}$. Then $1 - P_\nu(t)$ is monotonically increasing in the interval (0, 1) and $\hat{t}_k^* = \hat{t}_k$. Lemma 3.2 shows that

$$\hat{t}_{k+1} = (1-\gamma^2)\, \hat{t}_k\, Q_{\nu-1}(\hat{t}_k). \tag{3.5}$$

Since $tQ_{\nu-1}(t)$ is concave and $\frac{d}{dt}(tQ_{\nu-1}(t))|_{t=0} = Q_{\nu-1}(0) = \nu$, the recursion in (3.5) has a fix-point $\hat{t} = \hat{t}(\nu)$ in (0, 1), if $(1-\gamma^2)\,\nu > 1$, i.e. if $\gamma^2 < 1 - \nu^{-1}$. The fix-point satisfies $Q_{\nu-1}(\hat{t}(\nu)) = (1-\gamma^2)^{-1}$ and for $\nu = 2$ we get $\hat{t}(2) = \dfrac{1-2\gamma^2}{1-\gamma^2}$, if $\gamma^2 < \frac{1}{2}$. Since $\hat{t}_1 = 1-\gamma^2$ and $\hat{t}_1 > \hat{t}(2)$, the sequence (3.5) is monotonically decreasing and $\hat{\lambda} \equiv (1-\gamma^2)/\hat{t}(2)$ is an upper bound of λ_k, i.e. $\lambda_k \leqq \dfrac{(1-\gamma^2)^2}{1-2\gamma^2}$, if $\gamma^2 < \frac{1}{2}$. For $\nu = 3$ we get similarly $\hat{t}(3) = \dfrac{2}{1-\gamma^2}(2-3\gamma^2) \Big/ \left(3 + \left(1 + \dfrac{4\gamma^2}{1-\gamma^2}\right)^{\frac{1}{2}}\right)$ and

$$\lambda_k \leqq \hat{\lambda} = \frac{1}{2}(1-\gamma^2)\left(3 + \left(\frac{1+3\gamma^2}{1-\gamma^2}\right)^{\frac{1}{2}}\right) \Big/ (2-3\gamma^2), \qquad \text{if } \gamma^2 < \frac{2}{3}.$$

Note that for $\gamma^2 = \frac{1}{2}$ (a case occuring frequently in practice) we have in this case $\hat{\lambda} = (3+\sqrt{5})/4 \simeq 1.31$, i.e. a very small condition number.

Example 2.2. Here it follows from (3.4) that

$$\lambda_{k+1} = [1 - \max_{\alpha \leqq t \leqq 1} P_\nu(t)]^{-1}$$

if

$$\alpha \leqq t_k \leqq (1-\gamma^2)/\lambda_k.$$

Lemma 3.3. a) $T_\nu\left(\frac{1+\alpha}{1-\alpha}\right)=\frac{1}{2}\left[\left(\frac{1+\sqrt{\alpha}}{1-\sqrt{\alpha}}\right)^\nu+\left(\frac{1-\sqrt{\alpha}}{1+\sqrt{\alpha}}\right)^\nu\right]$, $\alpha \neq 1$.

b) $\lambda_{k+1}=\left(\frac{(1+\sqrt{\alpha})^\nu+(1-\sqrt{\alpha})^\nu}{(1+\sqrt{\alpha})^\nu-(1-\sqrt{\alpha})^\nu}\right)^2$, $0<\alpha<1$, if

$$\gamma^2 \leqq 1-\left(\frac{(1+\sqrt{\alpha})^\nu+(1-\sqrt{\alpha})^\nu}{2\sum_{s=1}^{\nu}(1+\sqrt{\alpha})^{\nu-s}(1-\sqrt{\alpha})^{s-1}}\right)^2. \tag{3.6}$$

c) *The asymptotic reduction rate of the preconditioned conjugate gradient method is*

$$\rho=\frac{\sqrt{\lambda_{k+1}}-1}{\sqrt{\lambda_{k+1}}+1}=\left(\frac{1-\sqrt{\alpha}}{1+\sqrt{\alpha}}\right)^\nu.$$

Proof. Part a) follows by a straightforward computation. Since

$$\max_{\alpha \leqq t \leqq 1} P_\nu(t)=2\Big/\left[T_\nu\left(\frac{1+\alpha}{1-\alpha}\right)+1\right],$$

we have

$$\lambda_{k+1}=\left[T_\nu\left(\frac{1+\alpha}{1-\alpha}\right)+1\right]\Big/\left[T_\nu\left(\frac{1+\alpha}{1-\alpha}\right)-1\right],$$

so the first part of b) follows by a).

The final part follows by $1-\gamma^2 \geqq \alpha\lambda_k$ and a straightforward computation. Part c) follows from a classical result (see for instance [1]) for conjugate gradient methods. □.

Remark 3.1. The same reduction rate is valid for the preconditioned Chebyshev acceleration method, for Chebyshev polynomials on the interval $[\hat{\lambda}, 1]$.

Theorem 3.1. a) *Let γ be the constant in the strengthened* C.-B.-S. *inequality. Then the relative condition number of $M^{(k)}$ derived by the method* (2.3) *and* (2.4), *with respect to $A^{(k)}$, is bounded by $\lambda_k \leqq (1-\gamma^2)/\hat{t}(\nu)$, if the polynomial P_ν in* (2.4) *satisfies $P_\nu(t)=(1-t)^\nu$, and if $\gamma^2<1-\nu^{-1}$, where $\hat{t}(\nu)$ is the fix-point of* (3.5).

b) *If* $P_\nu(t)=\left[T_\nu\left(\frac{1+\alpha-2t}{1-\alpha}\right)+1\right]\Big/\left[T_\nu\left(\frac{1+\alpha}{1-\alpha}\right)+1\right]$, *and if γ^2 satisfies* (3.6), *then*

$$\lambda_k=\left(\frac{1+\rho}{1-\rho}\right)^2 \qquad k \geqq 2,$$

where $\rho=((1-\sqrt{\alpha})/(1+\sqrt{\alpha}))^{\nu}$. *In particular, if* $\gamma^2<1-\nu^{-2}$, *then there exists* $\alpha>0$ *such that* (3.6) *is satisfied.*

c) *The maximal* α, $0<\alpha<1$ *for which* (3.6) *is satisfied can be computed from the smallest positive reduction rate* ρ *such that*

$$\frac{1+\rho}{1-\rho}\,\frac{1-\rho^{1/\nu}}{1+\rho^{1/\nu}}=(1-\gamma^2)^{\frac{1}{2}}.$$

If $\nu=2$ *and* $\gamma^2<3/4$, *then*

$$\alpha=(3-4\gamma^2)/[2(1-\gamma^2)^{\frac{1}{2}}+1]$$

and

$$\lambda_k\leqq\frac{(1+\alpha)^2}{4\alpha}=(1-\gamma^2)\,[2(1-\gamma^2)^{\frac{1}{2}}+1]/(3-4\gamma^2).$$

Proof. Part a) follows from (3.5) and the computations done for Example 2.1. Part b) and c) are readily established from Lemma 3.3. □

For $\gamma^2=\frac{1}{2}$ and $\nu=2$ we get $\lambda=\frac{1}{2}(\sqrt{2}+1)\simeq 1.21$ and $\rho\simeq 0.048$.

Example 2.3. Example 2.3 is close to Example 2.2, when ν is odd. We consider only the cases $\nu=3$ and $\nu=5$.

Consider first example (2.3) with $\nu=3$ where

$$P_\nu(t)=(1-t)\,(2t-1)^2,\qquad Q_{\nu-1}(t)=1+4(1-t)^2. \tag{3.7}$$

We have $\max\limits_{\frac{1}{3}\leqq t\leqq 1} P_\nu(t)=2/\left[T_\nu\left(\dfrac{1+\alpha_\nu}{1-\alpha_\nu}\right)+1\right]=\dfrac{1-\alpha_\nu}{\nu^2}=\dfrac{2}{27}$, and if $(1-t_k)^3\leqq\frac{2}{27}$ i.e. if $t_k\leqq 1-\frac{1}{3}\,2^{\frac{2}{3}}\simeq 0.58$, then method (2.3) with the polynomial (3.7) will give smaller relative condition numbers than with $P_3(t)=(1-t)^3$ as in Example 2.1. Since $P_\nu(t)=\frac{2}{27}$ for $t=\frac{1}{3}$ (and $t=\frac{5}{6}$) we have for $\frac{1}{3}\leqq t_k\leqq 1-\frac{1}{3}\,2^{\frac{2}{3}}$, that $t_k^*=\frac{5}{6}$ in (3.4) and

$$\hat{\lambda}_{k+1}=(1-\tfrac{2}{27})^{-1}=\tfrac{27}{25}=1.08.$$

This corresponds to $t_1=1-\gamma^2<1-\frac{1}{3}\,2^{\frac{1}{3}}$ and $\frac{1}{3}\leqq(1-\gamma^2)\,(1-P_\nu(t_k^*))=(1-\gamma^2)\,\frac{25}{27}$ i.e.

$$0.42\simeq\tfrac{1}{3}\,2^{\frac{1}{3}}<\gamma^2\leqq\tfrac{16}{25}.$$

If $\gamma^2>\frac{16}{25}$, then t_k^* in (3.4) equals t_k and it follows from (3.5) that

$$\hat{t}_{k+1}=(1-\gamma^2)\,\hat{t}_k\,Q_{\nu-1}(\hat{t}_k)$$

or

$$\hat{t}_{k+1}=(1-\gamma^2)\,\hat{t}_k[1+4(1-\hat{t}_k)^2],\qquad k=1,2,\ldots. \tag{3.8}$$

The recursion (3.8) has a fix-point if $(1-\gamma^2)\,Q_{\nu-1}(0)>1$, i.e. if $\gamma^2<\frac{4}{5}$, and then

$$\hat{t}=\left(1-\frac{1}{4}\,\frac{\gamma^2}{1-\gamma^2}\right)\Big/\left[1-\frac{1}{2}\left(\frac{\gamma^2}{1-\gamma^2}\right)^{\frac{1}{2}}\right]$$

and

$$\lambda_{k+1}\leqq\hat{\lambda}_{k+1}\leqq(1-\gamma^2)/\hat{t}=(1-\gamma^2)^2\left[1+\frac{1}{2}\left(\frac{\gamma^2}{1-\gamma^2}\right)^{\frac{1}{2}}\right]\Big/\left(1-\frac{5}{4}\gamma^2\right). \tag{3.9}$$

For the example (2.3) with $\nu=5$ we get similarly that the local maximum satisfies

$$\max P_\nu(t)=(1-\alpha_\nu)/\nu^2=\tfrac{4}{125}$$

whence

$$\hat{\lambda}_{k+1}\leqq 1+\tfrac{4}{125}=1.032,$$

if $\frac{1}{5}\leqq(1-\gamma^2)\,(1-P_\nu(t^*))$, i.e. if $\gamma^2\leqq\frac{96}{121}$.

If $\gamma^2\geqq\frac{96}{121}$ then

$$\hat{t}_{k+1}=(1-\gamma^2)\,\hat{t}_k\,Q_{\nu-1}(\hat{t}_k),\qquad k=1,2,\ldots$$

has a fix point if $(1-\gamma^2)\,Q_{\nu-1}(0)>1$, i.e. if $\gamma^2<\frac{10}{11}$.

We collect some of the above results in the following theorem.

Theorem 3.2. *Let γ be the constant in the strengthened* C.-B.-S. *inequality. Then the relative condition number of $M^{(k)}$ derived by the method* (2.3) *and* (2.4), *with respect to $A^{(k)}$ with $P_\nu(t)=(1-t)\,(2t-1)^2$ $(\nu=3)$, is bounded by $\hat{\lambda}_{k+1}\leqq 1.08$ if $\gamma^2\leqq\frac{16}{25}$ and $\hat{\lambda}_{k+1}$ is bounded by the expression in* (3.9) *if $\frac{16}{25}\leqq\gamma^2<\frac{4}{5}$. If $P_\nu(t)=(1-t)\,(1-10(1-t)+35(1-t)^2-50(1-t)^3+25(1-t)^4)$ $(\nu=5)$, then $\hat{\lambda}_{k+1}\leqq 1.032$ if $\gamma^2\leqq\frac{96}{121}$.*

Proof. It follows from (3.3′) and Lemma 3.2 that the relative condition number of $M^{(k)}$ w.r.t. $A^{(k)}$ is bounded above by λ_k. The rest of the results have already been derived. □

Consider now version (ii). Here Lemmata 3.1 and 3.2 take the form

Lemma 3.4. *Let $\tilde{A}^{(k)}$ be defined by* (2.9) *and $Q_{\nu-1}$ by* (2.6). *Then*

$$\begin{aligned}A^{(k)\frac{1}{2}}\,\tilde{A}^{(k)-1}\,A^{(k)\frac{1}{2}}&=I-P_\nu(A^{(k)\frac{1}{2}}\,M^{(k)-1}\,A^{(k)\frac{1}{2}})\\&=A^{(k)\frac{1}{2}}\,M^{(k)-1}\,A^{(k)\frac{1}{2}}\,Q_{\nu-1}(A^{(k)\frac{1}{2}}\,M^{(k)-1}\,A^{(k)\frac{1}{2}}).\end{aligned}$$

Lemma 3.5. *Let $T^{(k)}=A^{(k)\frac{1}{2}}\,M^{(k)-1}\,A^{(k)\frac{1}{2}}$ and let $\lambda_k=\sup\dfrac{\underline{v}_2^t\,M^{(k)}\underline{v}_2}{\underline{v}_2^t\,A^{(k)}\underline{v}_2}$, $\underline{v}_2\in\mathbf{R}^{n_k}$ and*

$$t_k=\inf_{\underline{v}_2}\{\underline{v}_2^t\,T^{(k)}\,\underline{v}_2/\underline{v}_2^t\,\underline{v}_2\},\qquad \underline{v}_2\in\mathbf{R}^{n_k}.$$

Then

$$\lambda_k^{-1}=t_k\leqq\underline{v}_2^t\,T^{(k)}\,\underline{v}_2/\underline{v}_2^t\,\underline{v}_2\leqq 1\qquad\forall\underline{v}_2\in\mathbf{R}^{n_k}. \tag{3.10}$$

Proof. Lemma 3.4 follows since P_ν is a polynomial and by the definition of $Q_{\nu-1}$. To prove Lemma 3.5, note that by Lemma 3.4 and (2.5),

$$0 \leqq \underline{v}_2^t A^{(k)} \underline{v}_2 \leqq \underline{v}_2^t \tilde{A}^{(k)} \underline{v}_2.$$

Hence we have

$$\underline{v}_2^t \tilde{A}^{(k)} \underline{v}_2 / \underline{v}_2^t S \underline{v}_2 \geqq \underline{v}_2^t A^{(k)} \underline{v}_2 / \underline{v}_2^t \bar{S} \underline{v}_2 \geqq 1 \qquad \forall \underline{v}_2 \in \mathbf{R}^{n_k},$$

so by (3.1),

$$\underline{v}^t M^{(k+1)} \underline{v} > \underline{v}^t A^{(k+1)} \underline{v} \qquad \forall \underline{v} \in \mathbf{R}^{n_{k+1}}.$$

This shows the upper bound of (3.10). The lower bound follows by the definition of t_k. □

To find an upper bound of λ_{k+1} we note that it follows from (3.1) and (3.2) that

$$\lambda_{k+1} = \sup_{\underline{v}_2} \{\underline{v}_2^t \tilde{A}^{(k)} \underline{v}_2 / \underline{v}_2^t S \underline{v}_2\}$$

so by Lemma 2.1 and (2.4),

$$\begin{aligned} \lambda_{k+1} &= \sup_{\underline{v}_2} \{\underline{v}_2^t \tilde{A}^{(k)} \underline{v}_2 / \underline{v}_2^t A^{(k)} \underline{v}_2\} \sup_{\underline{v}_2} \{\underline{v}_2^t A^{(k)} \underline{v}_2 / \underline{v}_2^t S \underline{v}_2\} \\ &\leqq (1-\gamma^2)^{-1} (1 - P_\nu(t_k^*))^{-1} = (1-\gamma^2)^{-1} (t_k^* Q_{\nu-1}(t_k^*))^{-1}, \end{aligned} \tag{3.11}$$

when $P_\nu(t_k^*) = \max P_\nu(t)$, $\lambda_k \leqq t \leqq 1$.

As for version (i), we see that the recursion (3.11) has a fix-point if $(1-\gamma^2) Q_{\nu-1}(0) > 1$. Therefore we have the same results for version (ii) as in Theorems 3.1 and 3.2 for version (i), if we multiply the upper bounds of $\tilde{\lambda}_{k+1}$ in those theorems with $(1-\gamma^2)^{-1}$, which result we state in a theorem.

Theorem 3.3. a) *The relative condition number of* $M^{(k)}$ *derived by the method* (2.3) *and* (2.9), *with respect to* $A^{(k)}$ *is bounded by*

$$\hat{\lambda}_{k+1} \leqq 1/\hat{t}(\nu),$$

if $P_\nu(t) = (1-t)^\nu$ *and* $\gamma^2 < 1 - \nu^{-1}$, *where* $\hat{t}(\nu)$ *is the fix-point of* (3.5).

b) *If* $P_\nu(t) = \left[T_\nu\left(\frac{1+\alpha-2t}{1-\alpha}\right)+1\right] \Big/ \left[T_\nu\left(\frac{1+\alpha}{1-\alpha}\right)+1\right]$, *and if* γ^2 *satisfies* (3.6), *then*

$$\lambda_{k+1} \leqq \frac{1}{1-\gamma^2} \left\{ \left[1 + \left(\frac{1-\sqrt{\alpha}}{1+\sqrt{\alpha}}\right)^\nu\right] \Big/ \left[1 - \left(\frac{1-\sqrt{\alpha}}{1+\sqrt{\alpha}}\right)^\nu\right] \right\}^2$$

where α *is the smallest positive root of*

$$(1-\gamma^2) t^{-1} = \left\{ \left[1 + \left(\frac{1-\sqrt{t}}{1+\sqrt{t}}\right)^\nu\right] \Big/ \left[1 - \left(\frac{1-\sqrt{t}}{1+\sqrt{t}}\right)^\nu\right] \right\}^2.$$

In particular, if $\gamma^2 < 1 - \nu^{-2}$, *then there exists such a root* $\alpha < 0$.

If $v=2$ *and* $\gamma^2<\frac{3}{4}$, *then*

$$\alpha=(3-4\gamma^2)/[2(1-\gamma^2)^{\frac{1}{2}}+1]$$

and

$$\lambda_k\leqq[2(1-\gamma^2)^{\frac{1}{2}}+1]/(3-4\gamma^2).$$

c) *If* $P_v(t)=(1-t)\,(2t-1)^2$ $(v=3)$ *then*

$$\lambda_{k+1}\leqq 1.08/(1-\gamma^2) \quad \text{if } \gamma^2\leqq\tfrac{16}{25}$$

and

$$\lambda_{k+1}\leqq(1-\gamma^2)\left[1+\frac{1}{2}\left(\frac{\gamma^2}{1-\gamma^2}\right)^{\frac{1}{2}}\Big/\left(1-\frac{5}{4}\gamma^2\right)\right]$$

if $\frac{16}{25}\leqq\gamma^2\leqq\frac{5}{4}$.

If $P_v(t)=(1-t)\,(1-10(1-t)+35(1-t)^2-50(1-t)^3+25(1-t)^4)$ $(v=5)$ *then*

$$\lambda_{k+1}\leqq 1.032/(1-\gamma^2), \quad \textit{if } \gamma^2\leqq\tfrac{96}{121}.$$

4 Applications and Conclusions

Consider first a nested sequence of triangular meshes for a polygonal domain, where we use piecewise linear finite element basis functions for the diffusion problem

$$-\underline{\nabla}\, a\, \underline{\nabla} u=f, \quad \underline{x}\in\Omega\subset\mathbf{R}^2,\ u=0 \text{ on } \partial\Omega. \tag{4.1}$$

The coefficients can be discontinuous, but we assume that they are constant on each element of the coarsest triangulation. (If they vary smoothly in each element, we get about the same results numerically.) For this case, it has been shown in Maitre and Musy [12], that $\gamma^2=\frac{3}{8}+\frac{1}{4}(d-\frac{3}{4})^{\frac{1}{2}}$, where $d=\max_r d_r$, $d_r=\sum_{i=1}^{3}\cos^2\theta_i^{(r)}$ and $\theta_i^{(r)}$ are the angles in the r'th triangle. For simplicity we assume that the recursive mesh refinement is done uniformly by dividing each triangle of the present level in four congruent triangles. Since $d<3$ (the limit case $d=3$ occurs only for a degenerative triangle), we have

$$\gamma^2=\frac{3}{4}-\frac{3-d}{2}\Big/[(4d-3)^{\frac{1}{2}}+3]<\tfrac{3}{4}.$$

Hence, if we use $v=2$ and

$$P_v(t)=\left[T_v\left(\frac{1+\alpha-2t}{1-\alpha}\right)+1\right]\Big/\left[T_v\left(\frac{1+\alpha}{1-\alpha}\right)+1\right],$$

where $\alpha=(3-4\gamma^2)/[2(1-\gamma^2)^{\frac{1}{2}}+1]$, Theorem 3.1 shows that

$$\lambda_k \leqq (1-\gamma^2)\,[2(1-\gamma^2)^{\frac{1}{2}}+1]/(3-4\gamma^2)$$

for version (i) of our method.

For $\nu=3$ and $P_\nu(t)=(1-t)(2t-1)^2$ we get $\lambda_k \leqq 1.08$ if $\gamma^2 \leqq \frac{16}{25}$ and λ_k is bounded by the expression in (3.9) if $\frac{16}{25} \leqq \gamma^2 \leqq \frac{4}{5}$.

For the special, but common, case of isosceles triangles $\left(\text{where } \theta_1=\frac{\pi}{2},\ \theta_2=\theta_3=\frac{\pi}{4}\right)$, we have $d=1$ and $\gamma^2=\frac{1}{2}$ so $\lambda_k \leqq \frac{1}{2}(\sqrt{2}+1) \simeq 1.21$ $(\nu=2)$ and $\lambda_k \leqq 1.08$ $(\nu=3)$. (Using Example 2.2 with α chosen optimally for this particular value of γ^2, we get an even smaller condition number for the case $\nu=3$.)

Hence a basic iterative method of the form

$$\underline{x}^{n+1}=\underline{x}^n-\frac{1}{\tau}M^{(l)^{-1}}(A^{(l)}\underline{x}^n-\underline{f}), \qquad n=0,1,\ldots,$$

where $\tau=2/(1+\lambda^{-1})$, to solve $A\underline{x}=\underline{f}$, converges with a reduction rate

$$\frac{\lambda-1}{\lambda+1} \simeq 0.094 \qquad (\nu=2, \gamma^2=\tfrac{1}{2}),$$

$$\frac{\lambda-1}{\lambda+1} \simeq 0.039 \qquad (\nu=3, \gamma^2 \leqq \tfrac{16}{25})$$

and

$$\frac{\lambda-1}{\lambda+1} \simeq 0.065 \qquad (\nu=3, \gamma^2=\tfrac{2}{3})$$

respectively.

The asymptotic reduction rate per step of a preconditioned Chebyshev or conjugate gradient method is

$$\frac{\sqrt{\lambda}-1}{\sqrt{\lambda}+1} \simeq 0.047 \qquad (\nu=2, \gamma^2=\tfrac{1}{2}),$$

$$\frac{\sqrt{\lambda}-1}{\sqrt{\lambda}+1} \simeq 0.019 \qquad (\nu=3, \gamma^2 \leqq \tfrac{16}{25})$$

and

$$\frac{\sqrt{\lambda}-1}{\sqrt{\lambda}+1} \simeq 0.033 \qquad (\nu=3, \gamma^2 \leqq \tfrac{2}{3})$$

respectively.

For version (ii) we get corresponding estimates of the reduction rates but with λ amplified with $(1-\gamma^2)^{-1}$.

We consider now the asymptotic work estimate of the method applied for the Example (4.1).

Let $w(k)$ be the amount of arithmetic work performed on level k. Then we have per iterative step,

$$w(l) \leqq CN + \nu w(l-1) \leqq CN\left(1 + \frac{\nu}{4}(1 + \nu w(l-2))\right),$$

etc., where N is the number of nodepoints on the finest mesh and C is a constant, which depends on the choice of the basis functions (i.e. the sparsity of the matrix) and the iterative method which is used, but does not depend on l or ν.

By recursion we get

$$w(l) \leqq CN\left(1 + \frac{\nu}{4} + \left(\frac{\nu}{4}\right)^2 + \ldots + \left(\frac{\nu}{4}\right)^{l-1}\right) + C_0 \nu^{l-1},$$

for some constant C_0, where the last term is the work done in total on the coarsest mesh, which level is visited ν^{l-1} times and where the linear system at this level is assumed to be solved exactly, by a direct method, for instance.

Hence

$$w(l) \leqq CN \frac{1 - \left(\frac{\nu}{4}\right)^l}{1 - \frac{\nu}{4}} + C_0 \nu^{l-1}, \quad \text{if } \nu \neq 4, \tag{4.2}$$

and

$$w(l) \leqq CNl + C_0 \nu^{l-1}, \quad \text{if } \nu = 4.$$

It follows from Theorems 3.1, 3.2 and 3.3 that the total work to iterate to a certain fixed relative precision ε (relative to the initial error) is of optimal order i.e. $O(N \log \varepsilon^{-1})$, of computational complexity, i.e. proportional to N, if γ is sufficiently small and $2 \leqq \nu \leqq 3$.

If $\nu = 4$, it grows as $O(N \log N \log \varepsilon^{-1})$ and if $\nu > 4$ it grows as $O(N^{1+\log_2(\nu/4)} \cdot \log \varepsilon^{-1})$, $N \to \infty$.

To match discretization and iteration error we choose $\varepsilon^{-1} = O(N)$ or $\varepsilon^{-1} = O(N^{\mu})$, for some positive μ. As has been described in Axelsson and Gustafsson [2], for instance, we can get rid of the extra factor $O(\log \varepsilon^{-1}) = O(\log N)$ if we start on the coarsest mesh and compute the numerical solution on each of the meshes, in increasing order.

It is interesting to note that the work estimate (4.2) essentially doubles when we choose the Chebyshev polynomial of degree $\nu = 3$ compared to the one of degree $\nu = 2$, but that the reduction rate becomes less than half of the reduction rate for $\nu = 2$ (at least if $\gamma^2 \geqq \frac{1}{2}$). *Hence it is most efficient to use the third degree polynomial.* Other choices of polynomial degrees will not give a method of optimal order of computational complexity, when we use a recursive uniform subdivision of the mesh.

As a conclusion we note that any finite element approximation of (4.1) for which the constant γ in the strengthened C.-B.-S. inequality satisfies $\gamma^2 < \frac{8}{9}$ can be solved with either version (i) or version (ii) of the method with an optimal

order of computational complexity if we let the polynomial P_ν in (2.4) respectively (2.9) be equal to a Chebyshev polynomial of degree 3, as described in Theorems 3.1 and 3.3, respectively.

When we compare version (i) to version (ii) we note that for $\nu=3$ we must solve four systems with matrix $A_{11}^{(k+1)}$ at every level of version (i) but only two such systems in version (ii). Even taking other computations into account a rough estimate is therefore that each iteration in version (i) costs about a double computational effort compared with version (ii). Since the condition number is $(1-\gamma^2)^{-1}$ times larger for the latter version and since the number of iterations in the PCG method is approximately proportional to the square root of the condition number, we see that version (ii) requires about the double number of iterations compared to version (i) if $\gamma^2 \geqq \frac{3}{4}$ and version (i) would be more efficient. However (again roughly) if $\gamma^2 < \frac{3}{4}$ version (ii) can be the most efficient version. Actually, the estimate used above for the number of conjugate gradient iterations is accurate only when the condition numbers are large, which they are not in the present case. Therefore, a closer look reveals that version (i) can be computationally more efficient even for smaller values of γ^2.

In the above discussion we have assumed that the systems with matrix $A_{11}^{(k+1)}$ can be solved with optimal order of computational complexity. If we solve them by an iterative method there will be some iteration errors left in the solutions of these systems. In a subsequent paper we shall show that this can indeed be done approximately without destroying the overall computation efficiency of the whole method.

Consider now problem (4.1) for an L-shaped domain with a reentrant corner. Starting from an initial uniform triangulation of the domain into isosceles triangles we recursively refine this near the singular point in the usual way (see Axelsson and Gustafsson [2], Fig. 2.3, for instance). This means that the angles of the initial triangulation are preserved and with piecewise linear finite element approximations we have therefore $\gamma^2 = \frac{1}{2}$.

After l such local refinements the number of meshnodes is $N_l = N_1 + 3(l-1)$, where N_1 is the number of meshnodes in the initial triangulation.

In this case the work is

$$\begin{aligned} w(l) &= C(N_l + \nu N_{l-1} + \ldots + \nu^{l-1} N_1) \\ &= CN_1(1+\nu+\ldots+\nu^{l-1}) + 3C(l-1+\nu(l-2)+\ldots+\nu^{l-1}) \\ &= CN_1 \frac{\nu^l - 1}{\nu - 1} + 3C\left(\frac{\nu}{\nu-1}(\nu^l-1) - l\right)\Big/(\nu-1) \end{aligned}$$

and the work relative to the number of meshnodes at level l is

$$\frac{w(l)}{N_1+3(l-1)} \leqq C\left(N_1 + \frac{3\nu}{\nu-1}\right) \cdot \frac{\nu^l}{N_1+3(l-1)}.$$

Hence, in this case the method does not have an optimal order of computational complexity, because the last factor grows (nearly geometrically) with l. However, in practice when solving (2.8) or (2.10) we can frequently let the components corresponding to some coarse level be unchanged during many iterations, be-

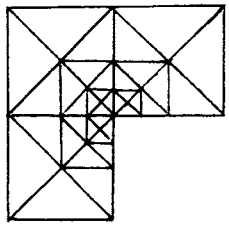

Fig. 4.1. Initial triangulation for a problem with a reentrant corner

cause one finds that they change very little anyway from one iteration to the next. This can decrease the computational complexity significantly. We shall not discuss this further in the present paper. The solution of the systems $A_{11}^{(k+1)}$ for this problem causes no difficulty because of the low order (three) of the systems.

An alternative method of solution for this problem is to first construct a refined mesh about the singularity (see Fig. 4.1) with a fixed number of refinements (5 in Fig. 4.1) to form an initial mesh. This mesh is subsequently refined recursively by dividing all triangles on the previous level into four congruent triangles. Then the previous results for isosceles triangulations apply with $\gamma^2 = \frac{1}{2}$. When solving the system corresponding to the initial triangulation with a direct solution method it can be advisable to use a frontal solution method letting the singular point be the root in the corresponding ordering of the matrix graph (see George and Liu [10] and Axelsson and Barker [1], for instance).

Numerical tests will be reported in a subsequent paper.

References

1. Axelsson, O., Barker, V.A.: Finite element solution of boundary value problems, 1st Ed. New York: Academic Press 1984
2. Axelsson, O., Gustafsson, I.: Preconditioning and two-level multigrid methods of arbitrary degree of approximation. Math. Comput. **40**, 219–242 (1983)
3. Axelsson, O.: On multigrid methods of the two-level type. In: Hackbusch, W., Trottenberg, U. (eds.) Multigrid methods, Proceedings, Köln-Porz 1981, pp. 352–367. LNM 960, Berlin Heidelberg New York: Springer 1982
4. Bank, R., Dupont, T.: Analysis of a two-level scheme for solving finite element equations, Report CNA-159, Center for Numerical Analysis, The University of Texas at Austin 1980
5. Bank, R., Dupont, T.: An optimal order process for solving finite element equations. Math. Comput. **36**, 35–51 (1981)
6. Bank, R., Dupont, T., Yserentant, H.: The hierarchical basis multigrid method. Numer. Math. **52**, 427–458 (1988)
7. Braess, D.: The contraction number of a multigrid method for solving the Poisson equation. Numer. Math. **37**, 387–404 (1981)
8. Braess, D.: The convergence rate of a multigrid method with Gauss-Seidel relaxation for the Poisson equation. In: Hackbusch, W., Trottenberg, U. (eds.) Multigrid methods, Proceedings, Köln-Porz 1981, pp. 368–386. LNM 960, Berlin Heidelberg New York: Springer 1982
9. Braess, D., Hackbusch, W.: A new convergence proof for the multigrid method including the *V*-cycle. SIAM J. Numer. Anal **20**, 967–975 (1983)
10. George, A., Liu, J.W.: Computer solution of large sparse positive definite systems, 1st Ed. Englewood Cliffs NJ: Prentice-Hall 1984

11. Kuznetsov, Y.A.: Multigrid domain decomposition methods for elliptic problems, talk presented at the conference on Numerical Linear Algebra and Parallel Computation, Mathematisches Forschungsinstitut, Oberwolfach, Febr. 28–March 5 (1988)
12. Maitre, J.F., Musy, F.: The contraction number of a class of two-level methods; an exact evaluation for some finite element subspaces and model problems. In: Hackbusch, W., Trottenberg, U. (eds.) Multigrid methods, Proceedings, Köln-Porz 1981, pp. 535–544. LNM 960, Springer, 1982
13. Vassilevski, P.: Nearly optimal iterative methods for solving finite element equations based on multilevel splitting of the matrix 1987 (submitted)
14. Yserentant, H.: On the multilevel splitting of finite element spaces. Numer. Math. **49**, 379–412 (1986)

Received May 2, 1988 / March 11, 1989