DEPARTMENT OF MATHEMATICS
UNIVERSITY OF NIJMEGEN The Netherlands

OPTIMIZING TWO-LEVEL PRECONDITIONINGS FOR THE CONJUGATE GRADIENT METHOD

Owe Axelsson, Igor Kaporin

Report No. 0116 (August 2001)

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF NIJMEGEN
Toernooiveld
6525 ED Nijmegen
The Netherlands

Optimizing Two-Level Preconditionings for the Conjugate Gradient Method

Owe Axelsson* Igor Kaporin†

Keywords: robust preconditioning, two-level preconditioning, spectral condition number, K-condition number, conjugate gradient method

Abstract

The construction of efficient iterative linear equation solvers for ill-conditioned general symmetric positive definite systems is discussed. Certain known two-level conjugate gradient preconditioning techniques are presented in a uniform way and are further generalized and optimized with respect to the spectral or the K-condition numbers. The resulting constructions have shown to be useful for the solution of large-scale ill-conditioned symmetric positive definite linear systems.

1 Introduction

In the present paper, we address the construction of preconditionings for the Conjugate Gradient algorithm, see, e.g.[1], and the rate of convergence of the method. This method is used for solving linear algebraic systems

$$Ax = b \tag{1.1}$$

with a large, normally sparse, unstructured Symmetric Positive Definite (SPD) matrix A of order n, such as arising in computational mechanics, from symmetrization of unsymmetric problems, etc.

Below we consider a preconditioning which is closely related to both the Generalized Augmented Matrix (GAM) preconditioning [18] and the approximate Schur complement one [4], [1], [3]. We restrict our considerations to two-level schemes based on a 2×2 splitting of the coefficient matrix and present a uniform framework for the analysis of such preconditioners. One of the main results is that the K-condition number of the preconditioned matrix is minimized under a very simple choice of approximation of the Schur complement. The upper bounds obtained for the K-condition and spectral condition numbers of the preconditioned matrix show that one

*Department of Mathematics, University of Nijmegen, The Netherlands, e-mail: axelsson@sci.kun.nl

†Center for Supercomputer and Massively Parallel Applications,Computing Center of Russian Academy of Sciences,Vavilova 40, Moscow 117967, Russia, e-mail: kaporin@ccas.ru

can expect good overall preconditioning quality whenever the preconditioning of the leading block of the matrix has a sufficiently high quality. The latter can be attained by a proper choice of the 2×2 splitting of the matrix, as well as by application of improved preconditioning methods such as Second Order Cholesky type incomplete factorizations [16]. As shown in [8], for finite element applications of second order problems using a certain element based preconditioner, it is also possible to obtain accurate bounds of the condition number which hold uniformly in both problem and discretization parameters.

In the present paper we consider a combined preconditioning strategy for the Conjugate Gradient algorithm intended to achieve fast convergence while providing low iteration costs. The algorithm can be considered as a proper combination of preconditioning strategies described in [4, 7, 18, 12, 13, 15, 16]. In the first stage, the original matrix A is split into a 2 by 2 block form with the leading block as large as possible while still well-conditioned. This is typically accompanied by a certain congruence transformation which is intended to enable a further improvement of the conditioning of the whole matrix or its leading block. In the second stage, an (approximate) block Jacobi preconditioning is used to construct the preconditioner in its final form.

The remainder of the paper is organized as follows. In Section 2 we recall two upper bounds on the number of iterations for the conjugate gradient method, and indicate their usefulness in the construction of preconditioners. In Section 3 a uniform presentation of various two-stage preconditionings is presented with condition number optimality results. In the same framework, a treatment of Schur complement preconditioners is given in Section 4. In Section 5 we describe the Conjugate Gradient Normal Equations algorithm for a guaranteed precision iterative solution of highly unsymmetric systems and discuss the potential use of the above described preconditionings with this method.

2 Two iteration bounds for the Preconditioned CG method

In order to highlight the target functions that should be optimized by the preconditioning, let us recall some known convergence results for the PCG method.

Consider the PCG method for the solution of SPD systems with symmetric positive definite preconditioner H that approximates A^{-1} in some sense. The standard estimation for the PCG iteration number needed for an ε times reduction of the error norm $(r_i^T A^{-1} r_i)^{1/2}$ is (see. e.g., [1])

$$i_\kappa(\varepsilon) \leq \left\lceil \frac{1}{2}\sqrt{\kappa(HA)}\log\frac{2}{\varepsilon} \right\rceil,$$

where, for any symmetrizable matrix M with positive eigenvalues,

$$\kappa(M) = \lambda_{\max}(M)/\lambda_{\min}(M).$$

This bound follows from a well-known estimate, cf.[1], establishing the linear rate of convergence for the PCG iterations. In some (model) cases, the latter estimate is useful for obtaining *a priori* bounds expressed via the parameters of the problem solved. For an important example, see [2]. However, the requirement of "optimal" conditioning does not in general yield a concrete construction of the preconditioning.

Therefore, an alternative approach was developed based on the use of an iteration number estimate via the K-condition number, cf.[15, 1]. Based on the corresponding superlinear convergence rate result, a simplified iteration number estimate of the following form holds (provided that the A-norm of the residual is replaced by the H-norm):

$$i_K(\varepsilon) = \left\lceil \log_2 K(HA) + \log_2 \frac{1}{\varepsilon} \right\rceil,$$

where, by definition,

$$K(M) = \left(\frac{1}{n} \mathrm{trace}(M) \right)^n / \det(M).$$

This bound can be useful in predicting the superlinear rate of convergence for a number of iterations exceeding, but close to, $\log_2 K(HA)$. It follows that

$$i_K(\varepsilon) \leq \left\lceil n \log_2(\frac{a}{g}) + \log_2(\frac{1}{\varepsilon}) \right\rceil$$

where a is the arithmetic average and g is the geometric average of the eigenvalues of HA. Typically, in practice for instance when considering a class of problems of increasing sizes, such as for difference methods for partial differential equations, it holds that $a/g \geq 1 + c$ for some postive c, independent on n. Hence in such cases $i_K(\varepsilon) \leq cn + \log_2 \frac{1}{\varepsilon}$.
Therefore this condition number somestimes gives rather pessimistic *a priori* upper bounds of the number of iterations. However, it may readily be (nearly) minimized in the context of various preconditioning procedures, as shown already in [12, 15, 1]. Thus, the K-condition number can be viewed as a useful tool for the construction of preconditionings. As soon as the preconditioning is specified, one can also try to estimate its standard (spectral) condition number in order to verify its efficiency. Several examples of such investigations are found in the paper.

3 Two-stage preconditionings

In this section, we will consider the following general scheme for preconditioning of SPD matrices. Let A be a result of certain preprocessing of the original matrix A_0, e.g. by preordering, or scaling it to unit diagonal,

$$A = \mathrm{Diag}(A_0)^{-1/2} A_0 \mathrm{Diag}(A_0)^{-1/2}$$

or, sometimes, even by a two-sided preconditioning by Incomplete Cholesky,

$$A = U^{-T} A_0 U^{-1},$$

such as with the use of the IC2 preconditioning of [16].

- **Stage 1.** Let Z be a nonsingular matrix with 2 by 2 block structure, and consider a congruence transformation of A, keeping the same block structure

$$B = Z^T A Z = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

such that $K(B) \leq K(A)$. The main purpose of such a transformation is to reduce as much as possible the quantity

$$\gamma = \|B_{11}^{-1/2} B_{12} B_{22}^{-1/2}\|,$$

which always satisfies $0 < \gamma < 1$. We will see that γ should not be too close to 1 in order for the preconditioning to be efficient.

- **Stage 2.** Let D be a block-diagonal 2 by 2 matrix with diagonal blocks D_1 and D_2 equal to the (approximate) inverses of B_{11} and B_{22}, respectively. We shall refer to such a preconditioning as Approximate Block Jacobi preconditioning. Then, typically, $K(DB) < K(B) \leq K(A)$ and since $\lambda_i(DB) = \lambda_1(HA)$, it holds $K(DB) = K(HA)$, where the resulting preconditioner for A will be

$$H = ZDZ^T.$$

The effect of approximate Block Jacobi preconditioning on the spectal condition number and on the K-condition number were first studied in [4] and [15], respectively.

We consider first two illuminating examples of congruence transformations which can be related to the first stage of such preconditionings.

Example 1. Let

$$Z = \begin{bmatrix} I_1 & -A_{11}^{-1} A_{12} \\ 0 & I_2 \end{bmatrix}$$

where $I_i, \ \ i = 1, 2$ denote the identity matrices of consistent orders. Then an elementary computation shows that

$$Z^T A Z = \begin{bmatrix} A_{11} & 0 \\ 0 & S \end{bmatrix},$$

where $S = A_{22} - A_{21} A_{11}^{-1} A_{12}$ is the Schur complement. Hence, in this case, $\gamma = 0$. However, this is not a viable choice as it requires exact solutions of systems with A_{11} and S, and S is in general a full matrix.

It is anyhow interesting to note that the above transformation, when applied to the indefinite matrix

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & 0 \end{bmatrix},$$

leads to

$$Z^T A Z = \begin{bmatrix} A_{11} & 0 \\ 0 & -A_{21}A_{11}^{-1}A_{12} \end{bmatrix}.$$

In certain applications (such as for certain finite element approximation of finite element Stokes problem for incompressible fluids), it turns out that the matrix $A_{21}A_{11}^{-1}A_{12}$ is well-conditioned and the transformation can hence be of interest in this case.

Example 2. We recall now (see e.g. [9] and [1]) another well known example of a congruence transformation showing the relation between the standard and hierarchical (nodal) basis function matrices. Let J_{12} be the interpolation matrix between the sets of standard finite element basis functions and hierarchical basis functions, i.e., it holds $v_{SB} = J_{12}v_{HB}$ for corresponding elements in the two sets. The matrix J_{12} is typically very sparse. Let n_2 be the number of degrees of freedom of the coarse space, let n_1 be that of the added basis functions, and let

$$Z = \begin{bmatrix} I_1 & J_{12} \\ 0 & I_2 \end{bmatrix}.$$

Thus if A is the standard basis function matrix, it holds

$$B = Z^T A Z = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

where

$$B_{11} = A_{11}, \quad B_{12} = A_{12} + A_{11}J_{12}, \quad B_{21} = B_{12}^T,$$

and

$$B_{22} = A_{22} + A_{21}J_{12} + J_{12}^T A_{12} + J_{12}^T A_{11} J_{12}.$$

Here B is the hierarchical basis function matrix, see, e.g.[9]. While (e.g. in the case when a discretization of a 2-nd order elliptic equation is considered)

$$1 - \|A_{11}^{-1/2} A_{12} A_{22}^{-1/2}\| = O(h^2),$$

where h is a meshsize parameter, it holds that

$$\gamma = \|B_{11}^{-1/2} B_{12} B_{22}^{-1/2}\| = 1 - c, \quad 0 < c < 1,$$

for some c which does not depend on size, nor shape of elements and also not on jumps of coefficients if they occur only at the coarse mesh edges, for further details, see [4, 9].

While the hierarchical basis function submatrices B_{12}, B_{21}, are less sparse than the corresponding matrices for the standard basis function matrix (A), the congruence transformation $Z^T A Z$ allows one to work with A in computing actions of the iteration matrix.

Next we consider certain examples of Block Jacobi preconditionings, quite similar to those which were already described, e.g,, in [4], [1], [3].

3.1 Estimating the K-condition number for the Exact, Full and Partial 2 by 2 Block Jacobi preconditioning

Let us first consider the simplest case of the exact Block Jacobi method with preconditioner

$$H = \begin{bmatrix} B_{11}^{-1} & 0 \\ 0 & B_{22}^{-1} \end{bmatrix}$$

to the matrix B. It is well known that this preconditioning (up to an arbitrary positive scalar factor) is optimum over all 2 by 2 preconditionings with respect to the spectral condition number. Furthermore, the condition number of $H^{-1}B$ is $\kappa(HB) = (1+\gamma)/(1-\gamma)$, where $\gamma = \|B_{11}^{-\frac{1}{2}} B_{12} B_{22}^{-\frac{1}{2}}\|$. The following result [15] (see also [1]) shows that such optimality holds also in the sense of the K-conditioning.

Theorem 3.1 *Let an SPD $n \times n$ matrix B be split into 2×2 block form as above and let D_1 and D_2 be arbitrary SPD matrices of the same orders n_1 and n_2 as B_{11} and B_{22}, respectively, so that the block diagonal matrix*

$$D = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}$$

is also SPD. Then the minimum of the matrix functional $K(DB)$ is attained at $D_1 = B_{11}^{-1}$ and $D_2 = B_{22}^{-1}$ and is equal to

$$\min_{D_1, D_2 \text{ are SPD}} K(DB) = K(D_B^{-1}B) = \frac{\det(B_{11})\det(B_{22})}{\det(B)},$$

where

$$D_B = \begin{bmatrix} B_{11} & 0 \\ 0 & B_{22} \end{bmatrix}$$

is the block diagonal part of B.

Proof. (See Section A1 of [15].) Since $\mathrm{trace}(DB) = \mathrm{trace}(DD_B)$, $n^{-1}\mathrm{trace}(D_B^{-1}B) = 1$ and $\det(DB) = \det(DD_B)\det(D_B^{-1}B)$ it follows that the identity

$$K(DB) = K(DD_B)K(D_B^{-1}B)$$

holds. As follows from the arithmetic-geometric mean inequality, cf. [1, 15], the minimum of $K(DD_B)$ is equal to 1 and is attained if and only if $DD_B = \alpha I$ for some $\alpha > 0$. Hence, $D = \alpha D_B^{-1}$ and the required result readily follows for $\alpha = 1$.
Q.E.D.

Remark 3.1 Clearly, the 2 by 2 splitting should be chosen such that $\det(B_{11}B_{22})$ is as small as possible to obtain a better K-conditioning. Further, one can readily see that the attained value of K is

$$K(D_B^{-1}B) = 1/\det(I_2 - C^T C), \qquad C = B_{11}^{-1/2} B_{12} B_{22}^{-1/2},$$

where I_2 is the identity matrix of order n_2, which stresses again the importance of making the norm of the matrix C as small as possible.

In practice, it appears to be an important case when the matrix D_1 is prescribed, and only D_2 can be optimized. The corresponding preconditioning can be regarded as the Partial Block Jacobi one. The following result, which generalize that of Theorem 3.1, holds in this case.

Theorem 3.2 *Let an SPD $n \times n$ matrix B be split into 2×2 block form*

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}.$$

with the orders of the diagonal blocks B_{11} and B_{22} being n_1 and n_2, respectively. Let D be the block diagonal matrix

$$D = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix},$$

with the $n_1 \times n_1$ SPD block D_1 fixed and the D_2 block being an arbitrary $n_2 \times n_2$ SPD matrix. Then the minimum of the matrix functional $K(DB)$ with respect to D_2 is attained for

$$D_2 = \sigma B_{22}^{-1}$$

and is equal to

$$\min_{D_2 \text{ is SPD}} K(DB) = \frac{\sigma^{n_1} \det(B_{22})}{\det(B)\det(D_1)} = K(D_1 B_{11}) K(D_B^{-1} B),$$

where

$$\sigma = \frac{1}{n_1} \text{trace}(D_1 B_{11}).$$

Proof. Using the inequality $K(X) \geq 1$ with $X = D_2 B_{22}$ (which holds for any diagonalizable matrix X with positive eigenvalues), one has

$$\text{trace}(D_2 B_{22}) \geq n_2 (\det(D_2 B_{22}))^{1/n_2}.$$

Therefore, we obtain the following lower bound for $K(DB)$:

$$K(DB) = \frac{\left(\frac{1}{n}(\text{trace}(D_1 B_{11}) + \text{trace}(D_2 B_{22}))\right)^n}{\det(D_1)\det(D_2)\det(B)} \geq \frac{\left(\frac{1}{n}(n_1 \sigma + n_2 (\det(D_2 B_{22}))^{1/n_2})\right)^n}{\det(D_1)\det(D_2)\det(B)}$$

$$= \left(\frac{\frac{n_1}{n}\sigma + \frac{n_2}{n}(\det(D_2 B_{22}))^{1/n_2}}{(\det(D_2 B_{22}))^{1/n}} \right)^n \frac{\det(B_{22})}{\det(B)\det(D_1)} \geq \left(\min_{\tau > 0} \varphi(\tau) \right)^n \frac{\det(B_{22})}{\det(B)\det(D_1)},$$

where we denoted $\tau = \det(D_2 B_{22})$ and

$$\varphi(\tau) = \frac{n_1}{n} \sigma \tau^{-\frac{1}{n}} + \frac{n_2}{n} \tau^{\frac{1}{n_2} - \frac{1}{n}}.$$

An elementary computation shows now that the minimum of φ equals $\sigma^{n_1/n}$ and is attained for

$$\tau = \sigma^{n_2},$$

Hence the expression for the optimum value of the K-condition number is proved. Further, it follows from the proof that this lower bound is attained when $K(X) = K(D_2 B_{22}) = 1$, which yields $D_2 B_{22} = \alpha I_2$ with $\alpha > 0$. The above formula for τ now readily yields the required result by letting $\alpha = \sigma$.
Q.E.D.

Remark 3.2 Theorem 3.2 shows that by first minimizing the K-condition number with respect to D_2 (for D_1 fixed) and then minimizing the resulting condition number with respect to D_1 the same optimality result holds as when the condition number is minimized by simultaneously varying D_1 and D_2.

Remark 3.3 In the case when both D_1 and D_2 are only approximations to the inverses of the diagonal blocks of B but the following scaling property holds,

$$\mathrm{trace}(D_1 B_{11})/n_1 = \mathrm{trace}(D_2 B_{22})/n_2 = 1,$$

a simple exact formula holds for the resulting K-condition number of the preconditioned matrix:

$$K(DB) = K(D_1 B_{11}) K(D_2 B_{22}) K(D_B^{-1} B). \tag{3.1}$$

3.2 Estimating the spectral condition number for Approximate 2 by 2 Block Jacobi preconditioning

Let us now consider the estimates for the standard (spectral) condition number $\kappa(DA)$ obtained when applying an Approximate Block Jacobi preconditioning.

These results are found in [4, 1] but are presented here for completeness. In particular, we follow [1], pp. 378-380. They hold also for singular (i.e. positive semidefinite) matrices with B_{11} nonsingular, for which $Bv = 0$ implies $B_{22} v_2 = 0$.

We consider then the extreme eigenvalues of the generalized eigenvalue problem $\lambda D x = B x$. Note that in this subsection, we let $D = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}$ where D_2 will be singular if B_{22} is singular. Further, γ is the constant in the strengthened Cauchy-Bunyakowski-Schwarz (CBS) inequality,

$$x_1^T B_{12} x_2 \leq \gamma \{x_1^T B_{11} x_1 x_2^T B_{22} x_2\}^{\frac{1}{2}}.$$

If both B_{11} and B_{22} are positive definite, then, as we have seen, $\gamma = \|B_{11}^{-\frac{1}{2}} B_{12} B_{22}^{-\frac{1}{2}}\|$. Clearly, $\gamma < 1$.
Below, the notation $A \geq B$ means that $A - B$ is positive semidefinite.

Theorem 3.3 *Let B be symmetric and positive semidefinite and split in a two by two block form such that if $Bv = 0$ then $B_{22}v_2 = 0$ when $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ is split correspondingly. Let γ be the constant in the corresponding strengthened CBS inequality and assume that*

$$\alpha_1 B_{11} \le D_1 \le \beta_1 B_{11}$$
$$\alpha_2 B_{22} \le D_2 \le \beta_2 B_{22}$$

for some $0 < \alpha_1 \le \beta_1$, $0 < \alpha_2 \le \beta_2$. Then with $D = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}$,

a) $\lambda_{\max} \le \frac{1}{\alpha_1}\left\{\frac{1}{2}\left(1+\frac{\alpha_1}{\alpha_2}\right) + \left[\left(\frac{1}{2}\left(1-\frac{\alpha_1}{\alpha_2}\right)\right)^2 + \frac{\alpha_1}{\alpha_2}\gamma^2\right]^{\frac{1}{2}}\right\}$,

$\lambda_{\min} \ge \frac{1-\gamma^2}{\beta_2}\left\{\frac{1}{2}\left(1+\frac{\beta_1}{\beta_2}\right) + \left[\left(\frac{1}{2}\left(1-\frac{\beta_1}{\alpha_2}\right)\right)^2 + \frac{\beta_1}{\beta_2}\gamma^2\right]^{\frac{1}{2}}\right\}^{-1}$

and the condition number of the preconditioned matrix $D^{-1}B$ is $\kappa \le \lambda_{\max}/\lambda_{\min}$.

b) If we scale the blocks so that $\frac{\alpha_1}{\alpha_2} \le \frac{\beta_1}{\beta_2} = 1$, then $\kappa \le \frac{\beta_1}{\beta_2}\,\frac{1+\gamma}{1-\gamma}$.

c) The following simplified upper bound holds,

$\kappa \le \frac{1}{1-\gamma^2}\left(\frac{1}{\alpha_1}+\frac{1}{\alpha_2}\right)(\beta_1+\beta_2)$.

Proof. The extreme eigenvalues are the extreme values of

$$\frac{x^TBx}{x^TDx} = \frac{x_1^TB_{11}x_1 + 2x_1^TB_{12}x_2 + x_2^TB_{22}x_2}{x_1^TD_1x_1 + x_2^TD_2x_2}.$$

Using the strengthened CBS-inequality and the arithmetic-geometric inequality $\sqrt{ab} \le \frac{1}{2}(\zeta a + \zeta^{-1}b)$, where $\zeta > 0$, we find

$$2|x_1^TB_{12}x_2| \le \gamma\zeta x_1^TB_{11}x_1 + \gamma\zeta^{-1}x_2^TB_{22}x_2.$$

This shows that

$$\lambda_{\max} \le \min_{\zeta>0}\max_{x_1,x_2}\frac{(1+\gamma\zeta)x_1^TB_{11}x_1 + (1+\gamma\zeta^{-1})x_2^TB_{22}x_2}{x_1^TD_1x_1 + x_2^TD_2x_2}$$

and using the given spectral relations between B_{11} and D_1 and B_{22} and D_2 we obtain

$$\lambda_{\max} \le \min_{\zeta>0}\max\left\{\frac{1+\gamma\zeta}{\alpha_1}, \frac{1+\gamma\zeta^{-1}}{\alpha_2}\right\}$$

where the optimal value of ζ is found from the equation

$$(1+\gamma\zeta)/\alpha_1 = (1+\gamma\zeta^{-1})/\alpha_2.$$

The lower eigenvalue bound is found in a similar way. Here

$$\begin{aligned}\lambda_{\min} &\geq \max_{\gamma<\zeta<\gamma^{-1}} \min_{x_1,x_2} \frac{(1-\gamma\zeta)x_1^T B_{11}x_1+(1-\gamma\zeta^{-1})x_2^T B_{22}x_2}{x_1^T D_1 x_1+x_2^T D_2 x_2} \\ &\geq \max_{\gamma<\zeta<\gamma^{-1}} \min\left\{\frac{1-\gamma\zeta}{\beta_1}, \frac{1-\gamma\zeta^{-1}}{\beta_2}\right\}\end{aligned}$$

where (if $\beta_1 \leq \beta_2$, otherwise exchange ζ with ζ^{-1} above) the optimal value of ζ satisfies

$$\gamma\zeta = \frac{1}{2}\left(1-\frac{\beta_1}{\beta_2}\right) + \left[\left(\frac{1}{2}\left(1-\frac{\beta_1}{\beta_2}\right)\right)^2 + \frac{\beta_1}{\beta_2}\gamma^2\right]^{\frac{1}{2}}$$

and we note that $\gamma^2 < \gamma\zeta < 1$, i.e. $\gamma < \zeta < \gamma^{-1}$, which gives the lower bound of $\lambda_{\min}$. Part b) follows by direct computation and to prove part c) we let $\gamma = 1$ in the square root expressions for $\lambda_{\max}$ and $\lambda_{\min}$. *Q.E.D.*

3.3 Deriving the n_2-Rank Modification K-optimum preconditioning

Let us now consider preconditioners of the form

$$H = I + VSV^T$$

for the matrix A, where the matrix V is a fixed $n \times n_2$ matrix with $n_2 \ll n$, and S is a symmetric $n_2 \times n_2$ matrix depending properly on V and A. This construction was investigated, e.g. in [12, 7, 18].

Following [12], let us choose S as the solution of the following optimization problem:

$$S = \arg\min_{S=S^T} K((I + VSV^T)A) \tag{3.2}$$

The resulting preconditioning was referred to as Low Rank Modification (LRM) in [12], in view of the requirement of limiting of the size of S when it is supposed to be calculated explicitly.

Assume that the columns of V are orthogonalized and let

$$Z_2 = V(V^TV)^{-1/2},$$

so that

$$Z_2^T Z_2 = I_2,$$

and introduce the $n \times n_1$ matrix Z_1 such that

$$Z_1^T Z_2 = 0$$

and

$$Z_1^T Z_1 = I_1.$$

In this case, the matrix

$$Z = [Z_1 Z_2]$$

will be orthogonal, and, by $ZZ^T = I$, one has

$$Z_1 Z_1^T + Z_2 Z_2^T = I.$$

One has then

$$K((I + VSV^T)A) = K(Z^T(I + VSV^T)ZZ^T AZ) = K((I + Z^T VSV^T Z)(Z^T AZ))$$

$$= K((I + Z^T Z_2 (V^T V)^{1/2} S (V^T V)^{1/2} Z_2^T Z)(Z^T AZ))$$

$$= K\left(\begin{bmatrix} I_1 & 0 \\ 0 & I_2 + (V^T V)^{1/2} S (V^T V)^{1/2} \end{bmatrix} \begin{bmatrix} Z_1^T AZ_1 & Z_1^T AZ_2 \\ Z_1^T AZ_2 & Z_2^T AZ_2 \end{bmatrix} \right).$$

Thus we have the same problem, the solution of which was given by Theorem 3.2. Therefore, setting

$$B_{i,j} = Z_i^T AZ_j, \quad i,j = 1,2, \qquad \sigma = \text{trace}(Z_1^T AZ_1)/n_1,$$

$$D_1 = I_1, \quad D_2 = I_2 + (V^T V)^{1/2} S (V^T V)^{1/2},$$

one has

$$I_2 + (V^T V)^{1/2} S (V^T V)^{1/2} = \sigma (Z_2^T AZ_2)^{-1}$$

which gives

$$S = -(V^T V)^{-1} + \sigma (V^T V)^{-1/2} (Z_2^T AZ_2)^{-1} (V^T V)^{-1/2}$$

$$= -(V^T V)^{-1} + \sigma \left((V^T V)^{1/2} Z_2^T AZ_2 (V^T V)^{1/2} \right)^{-1},$$

$$= -(V^T V)^{-1} + \sigma \left(V^T AV \right)^{-1},$$

where

$$\sigma = \frac{\text{trace}(A) - \text{trace}((V^T V)^{-1} V^T AV)}{n - n_2}.$$

This is the same formula as obtained in [12]:

$$H = \left(I - V(V^T V)^{-1} V^T \right) + \sigma V \left(V^T AV \right)^{-1} V^T. \tag{3.3}$$

In [7, 18] a somewhat different formula for S (namely, without the term $-(V^T V)^{-1}$ and with a different choice of σ) was used.

Thereby, the following more general preconditioner was considered,

$$H = M^{-1} + \sigma V \widehat{C}^{-1} V^T \tag{3.4}$$

where M and $\widehat{C}$ are positive definite preconditioners for A and $A_V = V^T AV$, respectively. Here M is typically a smoother used to damp the higher eigenvalue modes of A while $\widehat{C}$ can be chosen as a much simpler operator than B_V. The positive parameter σ is chosen to move the set of smallest eigenvalues of $M^{-1}A$ to a cluster of bigger eigenvalues, in this way improving the conditioning significantly for ill-conditioned

problems where, typically, there exist several small eigenvalues of A. A good choice is $\sigma = \lambda_{\max}(M^{-1}A)/\lambda_{\max}(\widehat{C}^{-1}A_V)$, which number can normally be estimated with little expense (see [18]).

As we have seen, the projection operator from (3.3) is based on choosing S to minimize the K-condition number in (3.2). However, as follows from the discussion in Section 2, this may not be the best choice in actually minimizing the number of iterations.

The following estimate of the extreme eigenvalues of HA holds showing a significant reduction in the condition number when the vector space spanned by the column vectors of V is sufficiently close to the eigenvector space for the smallest eigenvalues.

Theorem 3.4 *([18]) Let $H = M^{-1} + \sigma V\widehat{C}^{-1}V^T$ and assume that $\{\lambda_i, \underline{v}_i\}_{i=1}^n$ is an ordered set of eigenpairs of $M^{-1}A$ such that $\lambda_1 \leq \ldots \leq \lambda_n$. Let the matrix $V_e = [\underline{v}_1, \ldots, \underline{v}_m]$. If V is such that the subspace $\mathcal{W} = (Im\ A^{\frac{1}{2}}V)^{\perp}$ and $\mathcal{V}_e = Im(A^{\frac{1}{2}}V_e)$ satisfy*

$$\gamma = \cos(\mathcal{W}, \mathcal{V}_e) = \sup_{\substack{x \in W \\ y \in \mathcal{V}_e}} \frac{x^T y}{\{x^T x\ y^T y\}^{\frac{1}{2}}}$$

then the minimal eigenvalue of CB is bounded as

$$\lambda_{\min}(HA) \geq \max\left\{\lambda_1, (1-\gamma)\min\left\{\frac{\lambda_{\max}(M^{-1}A)}{\kappa(\widehat{C}^{-1}A_V)}, \lambda_{m+1}\right\}\right\}$$

and the maximal eigenvalue of HA is bounded as

$$\lambda_{\max}(HA) \leq 2\lambda_{\max}(M^{-1}A)$$

for any choce of V and $\widehat{C}$.

As shown in [18], the theorem can be generalized to include eigenspaces of $\widehat{M}^{-1}\widehat{A}$ for nearby matrices $\widehat{M}$ and $\widehat{A}$ satisfying $\widehat{M} \geq M$ and $\widehat{A} \leq A$.

The preconditioning method (3.4) has been called approximate subspace projection (ASP) method. Note that in the next section we will consider a similar preconditioning with M having rank n_1 and therefore presenting a generalization of the ASP and LRM preconditionings.

The numerical experiments presented in [12] showed that even with a relatively weak explicit preconditioning IIC the LRM techniques provides essential reduction of the total arithmetic costs of the method. (However, being used alone, LRM may even sometimes lead to a somewhat slower convergence.) Therefore, one may expect even greater improvements in the case of IC2-LRM preconditioning. Similarly, the extensive numerical experiments in [18] showed how the ASP method can be implemented in practice and gave several examples of significant reductions of the condition numers of several orders of magnitude.

As was pointed out in [7, 18] the subspace spanned by the columns of V should well approximate the subspace corresponding to the eigenvectors of A with the smallest eigenvalues. When there are few very small isolated eigenvalues of A, then such

a matrix V can be computed via the Lanczos method, and the ASP preconditioning will give a substantial reduction of the iteration number even with small n_2. However, such matrices V are not easily found in a general case, especially when n_2 may not be small. Some techniques are demonstrated in [18] to find a proper V for discretizations of second order elliptic problems. In more general cases we will consider a somewhat different approach which can be related to approximate Schur complement type preconditioners.

Remark 3.4 It is interesting to note that the theory developed above in Section 3.2 can be applied here in order to estimate the spectral condition number of the LRM preconditioned matrix as defined above. One can readily see that in this case all the conditions of Theorem 3.3 hold with

$$\xi_1 = \lambda_{\min}(Z_1^T A Z_1), \qquad \eta_1 = \lambda_{\max}(Z_1^T A Z_1), \qquad \xi_2 = \eta_2 = \sigma,$$

and

$$\gamma = \|(Z_1^T A Z_1)^{-1/2} Z_1^T A Z_2 (Z_2^T A Z_2)^{-1/2}\|.$$

Thus, the condition number estimate appears to be expressed essentially in the same terms as those of [7, 18] for a similar preconditioning.

Remark 3.5 Replacing this "exact" expression of S satisfying (3.2) with a certain approximation

$$S = -(V^T V)^{-1} + B_V^{-1},$$

one should require, by Theorem 3.3, that

$$\xi_2 \left(I_2 + (V^T V)^{1/2} S (V^T V)^{1/2}\right)^{-1} \le Z_2^T A Z_2 \le \eta_2 \left(I_2 + (V^T V)^{1/2} S (V^T V)^{1/2}\right)^{-1},$$

which is eqiuvalent to

$$\xi_2 B_V \le V^T A V \le \eta_2 B_V.$$

Hence, if the latter spectral bounds holds, Theorem 3.3 gives an estimate for the spectral condition number attained with the Approximate LRM preconditioning

$$H = \left(I - V(V^T V)^{-1} V^T\right) + V B_V^{-1} V^T.$$

In particular, one can see that as soon as $\eta_2/\xi_2 \le \eta_1/\xi_1$, such an approximation to $V^T A V$ can be regarded as quite acceptable.

4 Approximate Schur Complement type preconditionings

In this section, we present a common framework for two-level preconditionings using approximate Schur complements.

Let us suppose that the SPD matrix A is preordered in a proper way and consider its 2×2 splitting as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

Let the orders of the diagonal blocks A_{11} and A_{22} be n_1 and n_2, respectively. For practical reasons, we assume that $n_1 \gg n_2 \gg 1$ and that the matrix A_{11} is considerably better conditioned than A.
It is a well known fact that the following exact formula for the inverse matrix holds:

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}S^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}S^{-1} \\ -S^{-1}A_{21}A_{11}^{-1} & S^{-1} \end{bmatrix},$$

where

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

is the corresponding Schur complement. It is clarifying for the presentation to note that the above formula can be rewritten as a n_2-rank modification of a n_1-rank symmetric nonnegative definite matrix:

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -A_{11}^{-1}A_{12} \\ I_2 \end{bmatrix} S^{-1} \begin{bmatrix} -A_{21}A_{11}^{-1} & I_2 \end{bmatrix}.$$

Both of the above formulas are readily obtained from the following simple block matrix $L^T DL$-factorization:

$$A^{-1} = \begin{bmatrix} I_1 & -A_{11}^{-1}A_{12} \\ 0 & I_2 \end{bmatrix} \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & S^{-1} \end{bmatrix} \begin{bmatrix} I_1 & 0 \\ -A_{21}A_{11}^{-1} & I_2 \end{bmatrix}.$$

As above, we denote by I_1 and I_2 the identity matrix of the order n_1 and n_2, respectively.
Another useful relation is $\det(A) = \det(A_{11}) \det(S)$.
The statement of the problem is rather simple: let us replace the matrix A_{11}^{-1} by certain approximate inverses, symmetric D_1, or even unsymmetric H_1, that is,

$$A_{11} \approx D_1^{-1}, \qquad A_{11}H_1 \approx I_1,$$

and determine the matrix D_2 (to be used instead of S^{-1}) in order to obtain the preconditioner

$$H = \begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -H_1A_{12} \\ I_2 \end{bmatrix} D_2 \begin{bmatrix} -A_{21}H_1^T & I_2 \end{bmatrix} \tag{4.1}$$

which is as close to A^{-1} as possible, e.g. in the sense of minimization of $\kappa(HA)$ or $K(HA)$.

Note that if D_2 is dense, and $D_1 = U_1^{-1}U_1^{-T}$ as is the case when an incomplete Cholesky decomposition for A_{11} is used, and H_1 is taken as a sparse approximate inverse, then only about of

$$2\mathrm{nz}(U_1) + 2\mathrm{nz}(H_1) + 2\mathrm{nz}(A_{12}) + n_2^2$$

floating point operations (*flops*) are needed to multiply such a preconditioner H by a vector.

As follows from Theorem 3.2, if the additional scaling condition $\mathrm{trace}(D_1 A_{11}) = n_1$ holds, then the solution to the above problem is given by

$$D_2 = (A_{22} - A_{21}(H_1 + H_1^T - H_1^T A_{11} H_1)A_{12})^{-1},$$

in the sense that this matrix gives the minimum value of both the spectral and the K-condition numbers, cf.[1, 15].
This can be easily demonstrated if one considers

$$Z = \begin{bmatrix} I_1 & -H_1 A_{12} \\ 0 & I_2 \end{bmatrix}$$

and writes the preconditioner as previously,

$$H = ZDZ^T.$$

One has then

$$K(HA) = K(ZDZ^T A) = K(DZ^T AZ) = K(DB)$$

with

$$B = Z^T AZ = \begin{bmatrix} A_{11} & (I_1 - A_{11}H_1)A_{12} \\ A_{21}(I_1 - H_1^T A_{11}) & A_{22} - A_{21}(H_1 + H_1^T - H_1^T A_{11} H_1)A_{12} \end{bmatrix}.$$

Note that if a sparse approximate inverse H_1 is used, then the block B_{22} is also sparse (or at least its rows are easily computable) which gives a possibility to use a relatively large n_2 and apply an approximate inversion also for the block B_{22}, e.g., using the IC2 factorization. Another possibility is to recursively apply the method, e.g. as in [9, 3].

Remark 4.1 Note that the same preconditioning (but with $D_1 = H_1$) was cited in [17], formulas (2.9), (2.10), (2.12); the references therein go back to [22, 20]. In [17] the above formula for H_2 was found too complicated to be implemented in a multilevel method.

A similar construction was also used and analyzed in [21] (again with $D_1 = H_1$, cf. formulas (2.8)-(2.10) there) with a reference to [19].

Remark 4.2 The obtained preconditioning appears to be rather similar to the approximate subspace projection method, ASP, or generalized augmented matrix method, GAM, see, e.g. [18] and references therein. Indeed, let us denote the $n \times n_2$ block by

$$V = \begin{bmatrix} -H_1 A_{12} \\ I_2 \end{bmatrix};$$

then one has

$$H = \begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix} + V(V^T AV)^{-1} V^T.$$

However, the first term in GAM preconditioning was chosen to be of full rank n rather than n_1 in our case. Such a restriction may likely impair the resulting preconditioning quality.

4.1 Improving K-conditioning by the Approximate Schur Complement

Let us consider an approach to the construction of the matrix H_1 approximating A_{11} by the minimization of $K(Z^T AZ)$. Since $\det(Z) = 1$, such a setting is actually reduced to

$$\min_{H_1 \text{ is sparse}} \operatorname{trace}(A_{21}(I_1 - A_{11}H_1)^T A_{11}^{-1}(I_1 - A_{11}H_1)A_{12}),$$

or, even simpler,

$$\min_{H_1 \text{ is sparse}} \operatorname{trace}(-2A_{21}H_1A_{12} + A_{21}H_1^T A_{11}^{-1}H_1A_{12}),$$

which obviously presents an unconstrained quadratic optimization problem. The latter appears to be rather (structurally) complicated for general sparsity patterns of H_1; hence, let us consider certain special cases. Incidently, nearly the same minimization problem and similar constructions for the matrix H_1 were considered in [10].

In the case of a diagonal matrix $H_1 = \operatorname{Diag}(h)$ the above optimization problem appears to be rather easily solvable (at least, approximately). One can find that the vector h representing the diagonal of the matrix H_1, can be found as the solution of the system

$$(A_{11} \circ (A_{12}A_{21}))h = \operatorname{diag}(A_{12}A_{21}),$$

where "$\circ$" stands for the Hadamard (componentwise) product of matrices. The matrix of this system typically has strong diagonal dominance, which makes it possible to determine an approximation to h by a simple iterative method. It turns out that the case when $A_{12}A_{21}$ has zero diagonal entries yields virtually no essential complications.

In the case, when the matrix H_1 is chosen as a polynomial in A_{11},

$$H_1 = q_{k-1}(A_{11}) = \sum_{i=1}^{k} \gamma_i A_{11}^{i-1},$$

the polynomial coefficients can be found as the solution of the Hankel type system

$$\begin{bmatrix} \mu_1 & \mu_2 & \dots & \mu_k \\ \mu_2 & \mu_3 & \dots & \mu_{k+1} \\ \dots & \dots & \dots & \dots \\ \mu_k & \mu_{k+1} & \dots & \mu_{2k-1} \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \dots \\ \gamma_k \end{bmatrix} = \begin{bmatrix} \mu_0 \\ \mu_1 \\ \dots \\ \mu_{k-1} \end{bmatrix}$$

where

$$\mu_i = \operatorname{trace}(A_{21}A_{11}^i A_{12}).$$

Of course, k should not be large in order to make the values of μ_i and the entries of B_{22} easily computable.

An important property of this Approximate Schur Preconditioning is that it tends to improve the value of $K(D_B^{-1}B)$ as compared to $K(D_A^{-1}A)$, which seems important in view of the relation (3.1). Indeed, one has

$$K(D_B^{-1}B) = 1/\det(D_B^{-1}B) = 1/\det(I_1 - B_{11}^{-1/2}B_{12}B_{22}^{-1}B_{21}B_{11}^{-1/2}),$$

where

$$B_{11}^{-1/2}B_{12}B_{22}^{-1}B_{21}B_{11}^{-1/2}$$

$$= A_{11}^{-1/2}(I_1 - A_{11}H_1)A_{12}(A_{22} - A_{21}(H_1 + H_1^T - H_1^T A_{11}H_1)A_{12})^{-1}A_{21}(I_1 - H_1^T A_{11})A_{11}^{-1/2}$$

$$= A_{11}^{-1/2}(I_1 - A_{11}H_1)A_{12}(S + A_{21}(I_1 - H_1^T A_{11})A_{11}^{-1}(I_1 - A_{11}H_1)A_{12})^{-1}A_{21}(I_1 - H_1^T A_{11})A_{11}^{-1/2}.$$

Then, using the equality

$$\det(I_1 - X(S + X^T X)^{-1}X^T) = \det(S)/\det(S + X^T X)$$

with

$$X = A_{11}^{-1/2}(I_1 - A_{11}H_1)A_{12},$$

one gets

$$K(D_B^{-1}B) = \det(I_2 + S^{-1/2}A_{21}(I_1 - H_1^T A_{11})A_{11}^{-1}(I_1 - A_{11}H_1)A_{12}S^{-1/2})$$

$$\leq \left(1 + \frac{\|S^{-1}\|}{n_2}\mathrm{trace}\left(A_{21}(I_1 - H_1^T A_{11})A_{11}^{-1}(I_1 - A_{11}H_1)A_{12}\right)\right)^{n_2}.$$

Note that the latter estimate actually presents an upper bound for $K(D_B^{-1}B)$ in terms of $K(B)$, the minimization of which with respect to H_1 has been discussed in this subsection.

4.2 Improving spectral conditioning by the Approximate Schur Complement

With respect to the estimation of the spectral condition number, one may expect a considerable reduction of γ as compared to the original matrix A. For instance, it can be shown that if

$$A_{21}(A_{11}^{-1} - H_1 - H_1^T + H_1^T A_{11}H_1)A_{12} \leq \rho^2 A_{21}A_{11}^{-1}A_{12}, \qquad \rho < 1,$$

then

$$\frac{\gamma_B^2}{1 - \gamma_B^2} \leq \rho^2 \frac{\gamma_A^2}{1 - \gamma_A^2}$$

where

$$\gamma_A = \|A_{11}^{-1/2}A_{12}A_{22}^{-1/2}\|, \qquad \gamma_B = \|B_{11}^{-1/2}B_{12}B_{22}^{-1/2}\|.$$

The proof can easily be constructed using the same formulas as in the end of the preceeding subsection. In particular, one can see that

$$\gamma_B^2 = \max_{y \neq 0} \frac{y^T X^T X y}{y^T S y + y^T X^T X y}$$

with the same X as above, and therefore, by $X^T X \leq \rho^2 A_{21} A_{11}^{-1} A_{12}$,

$$\gamma_B^2 \leq \frac{\omega}{1+\omega}, \qquad \omega = \lambda_{max}(S^{-1} A_{21} A_{11}^{-1} A_{12}) = \frac{\gamma_A^2}{1-\gamma_A^2}.$$

Hence, the required estimate readily follows.

4.3 The choice of 2 by 2 splitting of the coefficient matrix

As was demonstrated above, it is advantageous to have the block A_{11} not only of large size n_1 but as well-conditioned as possible, since the latter requirement makes it easier to find a good approximate inverse for it. Also, it is advantageous when the columns of A_{21} are pairwise orthogonal, or nearly orthogonal. The latter condition can easily be satisfied if A is a sparse matrix, e.g. of the type arising when solving boundary value problems for elliptic PDE's using FD or FE discretizations. Hence, the splitting can be based on the extraction of the block A_{22} corresponding to an "independent set" of grid nodes. Otherwise, when A is not sparse or its sparsity is not regular enough, one can base the splitting of A using a certain "threshold pivot" Incomplete Cholesky factorization. For instance, supposing that A is symmetrically scaled to unit diagonal, one can set a certain threshold parameter ($\theta < 1$ and, in the course of an incomplete factorization (e.g. IC2 algorithm [16]) at the k-th step, one sets the whole current column of the right Cholesky factor U equal to the k-th column of the identity matrix whenever it appears that the actually computed value is $u_{ii} \leq \theta$. Such an algorithm returns the IC2 factorization of certain submatrix A_{11} of A such that all diagonal elements of its IC factor are sufficiently close to 1. The value of n_1 can be adjusted by a proper choice of the threshold θ.

5 The Preconditioned CGNR method with two-sided 2-norm error bounds

A possible application of high quality preconditionings for general SPD matrices may be the iterative solution of highly unsymmetric (sparse) linear systems.

In order to find the solution of unsymmetric linear system

$$A_0 x = b,$$

let us use the substitution $x = A_0^T y$; the system then takes the form

$$Ay = b$$

with

$$A = A_0 A_0^T.$$

Using now the PCG algorithm presented in [6] and using there the substitutions $x_k = A_0 y_k$, $q_k = A_0^T p_k$, one readily gets the solution method in the following form:

$$r_0 = b - A_0 x_0, \quad q_0 = A_0^T H r_0; \quad \sigma_{-1} = 0,$$

$$
\begin{aligned}
\textbf{for } i &= 0, 1, \ldots : \\
\sigma_i &= \sigma_{i-1} + 1/r_i^T H r_i, \\
&\textbf{if } (i > 0 \textbf{ and } i = 0 \bmod d) \textbf{ then} \\
&\quad \textbf{if } (2\|r_i\| \le \|b - A_0 x_i\|) \textbf{ then } QUIT\_1 \\
&\quad \mu_1^{(i)} = \lambda_{\min}(T_i) \\
&\quad \textbf{if } \left(\mu_1^{(i)} \approx \mu_1 \textbf{ and } \sum_{j=i-d}^{i-1} \omega_j + \frac{1}{\sigma_i \mu_1^{(i)}} \le \frac{\varepsilon^2}{1-\varepsilon^2} \sum_{j=0}^{i-1} \omega_j \right) \textbf{ then } QUIT\_2 \\
&\textbf{endif} \\
\alpha_i &= r_i^T H r_i / q_i^T q_i, \\
\omega_i &= (r_i^T H r_i)^2 / q_i^T q_i, \\
x_{i+1} &= x_i + q_i \alpha_i, \\
r_{i+1} &= r_i - A_0 q_i \alpha_i, \\
\beta_i &= r_{i+1}^T H r_{i+1} / r_i^T H r_i, \\
q_{i+1} &= H r_{i+1} + q_i \beta_i,
\end{aligned}
$$

where

$$
T_i = \begin{bmatrix}
\frac{1}{\alpha_0} & -\frac{\beta_0}{\alpha_0} & 0 & \ldots \\
-\frac{1}{\alpha_0} & \frac{1}{\alpha_1} + \frac{\beta_0}{\alpha_0} & -\frac{\beta_1}{\alpha_1} & \ldots \\
\ldots & \ldots & \ldots & \ldots \\
\ldots & 0 & -\frac{1}{\alpha_{i-2}} & \frac{1}{\alpha_{i-1}} + \frac{\beta_{i-2}}{\alpha_{i-2}}
\end{bmatrix}
$$

and it is assumed that the minimum eigenvalue of the latter tridiagonal matrix is sufficiently close to

$$\mu_1 = \lambda_{\min}(HA).$$

The latter situation typically takes place for sufficiently good preconditionings and small values of ε as the CG iterations enter the final stage of superlinear convergence.

Using the results of [6] one can see, from

$$\|y - y_i\|_A = \|x - x_i\|,$$

that the following two-sided estimate holds:

$$\sum_{j=i-d}^{i-1} \omega_j \le \|x - x_{i-d}\|^2 \le \sum_{j=i-d}^{i-1} \omega_j + \frac{\mu_1^{-1}}{\sum_{j=0}^{i} \gamma_j^{-1}}, \qquad d \le i \le n.$$

In practice, a value such as $d = 10$ proved to be satisfactory. When returning by $QUIT\_2$, one can see that the above upper bound, the equality

$$\|x - x_k\|^2 = \sum_{j=k}^{n-1} \omega_j, \quad k = 0, \ldots, n-1,$$

used with $k = 0$ and $k = i - d$, and the well-known inequality $\|x - x_i\| \leq \|x - x_{i-d}\|$ guarantee that the iterations are terminated with

$$\|x - x_i\| \leq \varepsilon \|x - x_0\|.$$

The return by $QUIT\_1$ corresponds to an inacceptably large discrepancy between the iterated residual and the "exact" residual. In this case, the restart of CGNE with $x_0 := x_i$ should be performed in an attempt to achieve the required precision.

Acknowledgement

This work has been supported by The NWO-NAVO grant NB 61-488, which is gratefully appreciated.

References

[1] Axelsson O. *Iterative Solution Methods.* Cambridge University Press: Cambridge, 1994.

[2] Axelsson, O., On iterative solvers in structural mechanics, separate displacement ordering and mixed variable methods, *Mathematics and Computers in Simulation*, **50** (1999), 11-30.

[3] Axelsson O. Stabilization of algebraic multilevel iteration methods; additive methods *Numerical Algorithms* 1999; **21**: 23–47.

[4] Axelsson O, Gustafsson I. Preconditioning and Two-Level Multigrid Methods of Arbitrary Degree of Approximation *Mathematics of Computation* 1983; **40**: 219–242.

[5] Axelsson O, Kaporin I, Konshin I, Kucherov A, Neytcheva M, Polman B, Yeremin A. Comparison of algebraic solution methods on a set of benchmark problems in linear elasticity. *Tech. Report* of Department of Mathematics, University of Nijmegen, The Netherlands, 2000, 89p.

[6] Axelsson O, Kaporin I. Error norm estimation and stopping criteria in preconditioned conjugate gradient iterations. *Numerical Linear Algebra with Applications* 2001; **8**: 265–286.

[7] Axelsson O, Neytcheva M, Polman B. The bordering method as a preconditioning method. *Vestnik Moscow Univ., Ser. 15: Vychisl. Mat. Cybern.* 1995, 3–24.

[8] Axelsson, O., Padiy, A., On the additive version of the algebraic multilevel iteration method for anisotropic elliptic problems, *SIAM J. Sci. Comp.*, **20** (1999), 1807-1830.

[9] Axelsson O, Vassilevski P. A survey of multilevel preconditioned iterative methods *BIT* 1989, **29**: 769–793.

[10] Eremin A, Kaporin I. Spectral optimization of explicit iterative methods.I. *J. Soviet Mathematics* 1987; **36**: 207–214.

[11] Kaporin I. On preconditioned conjugate-gradient method for solving discrete analogs of differential problems. *Differential Equations* 1990; **26**(7):897–906 (In Russian).

[12] Kaporin I. Two-level explicit preconditionings for the conjugate-gradient method. *Differential Equations* 1992; **28**(2):280–289 (In Russian).

[13] Kaporin I. Explicitly preconditioned conjugate gradient method for the solution of unsymmetric linear systems. *Int. J. Computer Math.* 1992; **40**: 169–187.

[14] Kaporin I. Spectrum boundary estimation for two-sided explicit preconditioning. *Vestnik Mosk. Univ., ser. 15, Vychisl. Matem. Kibern.* 1993; **2**:28–42 (in Russian).

[15] Kaporin I. New convergence results and preconditioning strategies for the conjugate gradient method. *Numerical Linear Algebra with Applications* 1994; **1**(2): 179–210.

[16] Kaporin I. High quality preconditioning of a general symmetric positive definite matrix based on its $U^TU+U^TR+R^TU$-decomposition. *Numerical Linear Algebra with Applications* 1998; **5**(6):483–509.

[17] Notay Y. Optimal V-cycle Algebraic Multilevel Preconditioning. *Numerical Linear Algebra with Applications* 1998; **5**(5):441–459.

[18] Padiy A, Axelsson O, Polman B. Generalized augmented matrix preconditioning approach and its application to iterative solution of ill-conditioned algebraic systems. *SIAM J. Matrix Anal. Appl.* 2000; **22**(3): 793–818.

[19] Reusken A. A multigrid method based on incomplete Gaussian elimination. *Numer. Linear Algebra Appl.* 1996; **3**(8):369–390.

[20] Ruge JW, Stüben K. Algebraic multigrid (AMG). In: S.F. McCormick, ed., Multigrid Methods, Vol.3 of Frontiers in Applied Math., 73–130. SIAM, Philadelphia, PA, 1987

[21] Shapira Y. Model case analysis of an algebraic multilevel method. *Numerical Linear Algebra with Applications* 1999; **6**(8):655–685.

[22] Stüben K. Algebraic multigrid (AMG): experiences and comparisons. *Appl. Math. Comput.* 1983; **13**: 419–452.