Some Convergence Estimates For Algebraic Multilevel Preconditioners

Matthias Bollhöfer* Volker Mehrmann†

Keywords. Sparse approximate inverse, sparse matrix, algebraic multilevel method, preconditioning, conjugate gradient method, Krylov subspace method.

AMS subject classification. Primary 65F05; Secondary 65F10, 65F50, 65Y05.

Abstract

We discuss the construction of algebraic multilevel preconditioners for the conjugate gradient method and derive explicit and sharp bounds for the convergence rates. We present several numerical examples that demonstrate the efficiency of the preconditioner.

1 Introduction

For the solution of large sparse linear systems of the form

$$Ax = b,\ A \in \mathrm{GL}(n, \mathbb{R}),\ b \in \mathbb{R}^n, \tag{1}$$

sparse approximate inverses [17, 16, 7, 12, 3, 18] have been successfully employed as preconditioners for Krylov–subspace methods [9, 19, 11]. The general principle in these methods is to construct a matrix B that is sparse and approximates A^{-1}. Several techniques have been developed, such as minimizing the norm of $\|AB - I\|$ subject to some prescribed pattern [16, 7, 12] or biconjugate techniques that are based on approximate factorizations $Z^\top AW \approx D$, where Z, W are upper triangular matrices and D is diagonal [3, 4].

These techniques work well for large classes of matrices, but there are cases when the sparse approximate inverse needs a large number of nonzero entries to become a suitable approximation to the inverse of A. Also, in the case of approximate inverses based on norm–minimizing techniques, it happens frequently that many singular values of the residual matrix $E = I - AB$ are quite small, while a small number of singular values are big and stay big even when more fill–in for B is allowed.

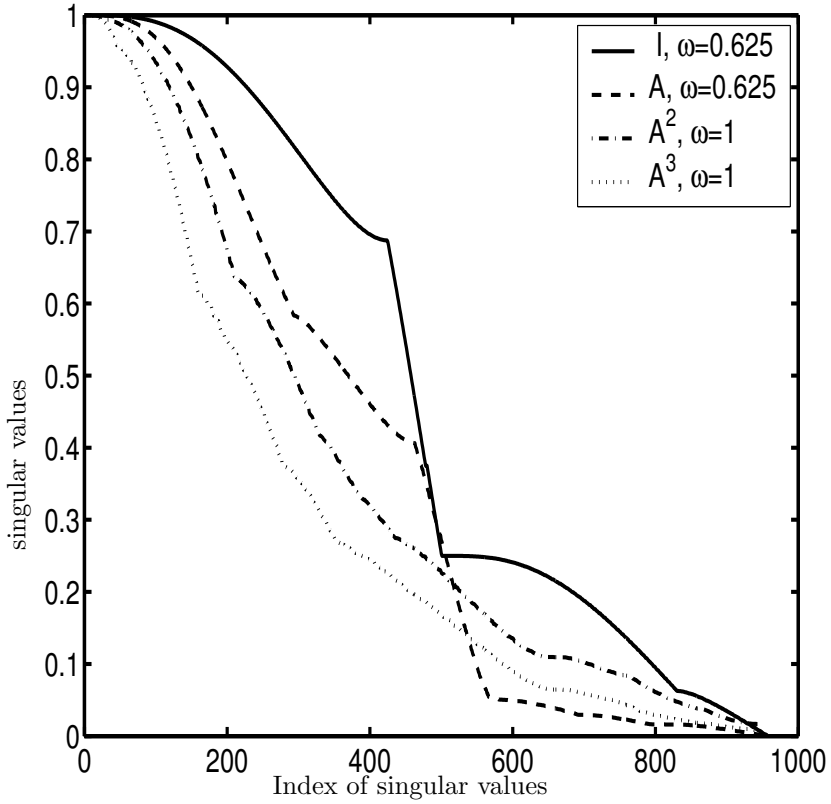
Consider the following example.

Example 1 *For the symmetric positive definite matrix LANPRO/NOS2 from the Harwell–Boeing collection [8], we apply the sparse approximate inverse suggested in [17, 16] with the sparsity pattern of A^k, $k = 0, 1, 2, 3$. This approach constructs an approximate inverse L of the Cholesky factor $\tilde{L}$ of A^{-1}, i.e., $\tilde{L} \approx L$ where $\tilde{L}\tilde{L}^\top = A^{-1}$. To evaluate the quality of the preconditioner, we set $E = I - \omega L^\top AL$ for a suitably chosen parameter ω, i.e., E is our residual, adapted to the symmetric positive case. The scaling parameter ω is chosen, so that as many eigenvalues as possible of $L^\top AL$ are close to 1. In Figure 1 we display the singular values of E in decreasing order. We observe that most of the singular values tend to 0, while a few singular values stay close to 1 even when increasing the number of nonzeros for L.*

The observation that many singular values are small but some stay large, even when the fill-in in the approximate inverse is increased, can be interpreted as follows. The matrix B approximates

*Institut für Mathematik, MA 4–5, Technische Universität Berlin, D–10623 Berlin, Germany

†Institut für Mathematik, MA 4–5, Technische Universität Berlin, D–10623 Berlin, Germany



A well on a subspace of large size, while there is almost no approximation on the complementary subspace. In the numerical treatment of partial differential equations this effect is typically called *smoothing property* [13].

If, after scaling the norm to 1, many of the singular values of the matrix $E = I - AB$ are small, while the others are close to 1, then it follows that

$$E = I - AB = UV^\top + F = E_0 + F, \tag{2}$$

with matrices $U, V^\top \in \mathbb{R}^{n,r}$ and $\|F\| \leqslant \eta < 1$, i.e., the residual $E = I - AB$ can be approximated well by a matrix of rank r much smaller than n. Here realistically $r \leqslant cn$ with $c \approx 0.5$. We want to use such an approximative factorization $UV^\top$ for the preconditioning of large scale problems. If the rank r of the approximation is of this order, then, to keep the storage requirements small, it would be ideal to have a sparse representation of the approximation $UV^\top$. This does not necessary mean that U, V have to be sparse, e.g. we could have $E_0 = PZ^{-1}\hat{P}^\top$ with sparse matrices $P, \hat{P}, Z$, while their product need not be sparse at all.

In order to determine such a sparse factorization, we observe that if $\eta \ll 1$, then the entries of E_0 only slightly differ from those of E. So we may expect that an appropriate selection of columns of E will be a good choice for determining $UV^\top$. This is underscored by the following lemma.

Lemma 1 *[6]. Let $E \in \mathbb{R}^{n,n}$ and let $E = [U_1, U_2]\,\mathrm{diag}\,(\Sigma_1, \Sigma_2)[V_1, V_2]^\top$ be the singular value decomposition of E with U_1, V_1 having r columns and $\|\Sigma_2\|_2 \leqslant \varepsilon$. Then there exist a permutation matrix $\Pi = [\Pi_1, \Pi_2]$ with analogous partitioning such that*

$$\inf_{Z \in \mathbb{R}^{r,r}} \|U_1 Z - E\Pi_1\| \leqslant \varepsilon. \tag{3}$$

By applying Lemma 1 to $E^\top$ instead of E we can analogously approximate V_1 by suitably chosen rows of E.

In [6] it has been discussed how one can use approximate factorizations such as (2) to modify a given preconditioner B. Instead of iterating with the preconditioned matrix AB one uses the

updated preconditioned matrix $AB(I+(AB)^{-1}UV^\top)$. In order to avoid that a small error, between the exact subspace U (associated with large singular values of E) and a set of columns of E, blows up the norm of $UV^\top$, one determines a decomposition of the form

$$E = (AB)PS + F. \tag{4}$$

Here we require that $\|F\| \leqslant \eta$ and $P, S^\top \in \mathbb{R}^{n,r}$ to keep the error $AB(I + (AB)^{-1}UV^\top) - I = AB(I+PS)-I$ small. But since $E = I-AB$, this means that we have to determine a factorization

$$AB(I + PS) = I - F \tag{5}$$

with $\|F\| \leqslant \eta$.

For the analysis of such a preconditioner, we need to study the approximation properties of the correction term $I + PS$ in (5). A detailed analysis and sharp convergence bounds that extend results of [6] are given in Section 3.

Furthermore, to obtain an efficient linear system solver, in (4) we have to find a representation $PS = P(Z^{-1}\hat{P}^\top)$ with sparse matrices $P, \hat{P}, Z$ and we also have to find an efficient method to determine the desired columns of E as suggested by Lemma 1. Both topics have been discussed in [6]. To determine the columns of E for the construction of the preconditioner, one could use a QR–like decomposition with column pivoting of $E\Pi = QR$ and use as P the first r columns of Q. A more elegant choice would be to use instead of the first r columns of Q the first r columns of $E\Pi$, since their columns span the same space as the associated columns of Q. But the computation of such a QR–like decomposition, i.e., in particular the computation of the pivoting matrix Π, has to be carried out with small memory requirements.

Once a strategy for constructing and updating sparse approximate inverse preconditioners has been established, this approach can be applied in a recursive way, leading to algebraic multigrid type methods. These methods are discussed in Section 2. Essential for the success of these multilevel processes is the coarsening process that we discuss in Section 4. We illustrate our results with several numerical examples.

In the sequel for symmetric matrices A, B we will use the notation $A \geqslant B$, if $A - B$ has nonnegative eigenvalues. When talking about a matrix P we do not strictly distinguish between the matrix P itself and the space $\{Px : x \in \mathbb{R}^r\}$ that is spanned by its columns.

2 Algebraic Multilevel Preconditioners

In this section we briefly review two types of algebraic multilevel preconditioners for symmetric positive definite systems that were introduced in [6]. Since here we concentrate on the symmetric case, we focus on preconditioners $B = LL^\top$ and use the symmetrized version of residual matrix $E = I - L^\top AL$. The symmetrized version of (4) then is as

$$E = (L^\top AL)PS + F.$$

Let $A \in \mathbb{R}^{n,n}$ be symmetric positive definite. Suppose that $LL^\top$ is a symmetric positive definite matrix that is an approximation to A^{-1} and that we use as preconditioner for the conjugate gradient method [10]. Here it is understood that L is a sparse matrix, so that linear systems can be solved with low complexity on the available computer architecture. Suppose that one realizes during the iteration process that the approximation of A^{-1} by $LL^\top$ is not satisfactory, i.e., that the condition number of the preconditioned matrix $M = L^\top AL$ is not small enough to get fast convergence of the conjugate gradient method.

In order to improve the preconditioner one can determine a matrix of the form

$$M^{(1)} = L(I + PZ^{-1}P^\top)L^\top \tag{6}$$

with $P \in \mathbb{R}^{n,p}$, $Z \in \mathbb{R}^{r,r}$ nonsingular, sparse and r smaller than n, so that $M^{(1)}$ is a better approximation to A^{-1} than $LL^\top$. In (6) the sparse approximate inverse $LL^\top$ is augmented as

$$LL^\top \longrightarrow LL^\top + \quad \begin{array}{|c|}\hline n \\ \times \\ r \\ \hline \end{array} \quad \boxed{} \quad \boxed{r \times n}$$

Note that one does not have to assume that P has low rank, but only that $r \leqslant cn$ with $c < 1$, e.g. $c = 0.5$. It is very important, however, that P, Z are sparse.

The particular form (6) resembles that of an algebraic two–level method, where multiplication with $P, P^\top$ corresponds to the mapping between fine and coarse grid and Z represents the coarse grid system. Note further, that in using $L(I + PZ^{-1}P^\top)L^\top$ as a preconditioner for A, only a system with Z has to be solved.

As shown in Lemma 1, skilfully chosen columns/rows of the residual matrix $E = I - M = I - L^\top AL$ can be used to approximate the invariant subspace of E associated with its large eigenvalues. As will be shown below, precisely this invariant subspace has to be approximated by P. In the sense of the underlying undirected graph of E, we refer to those nodes as *coarse grid nodes* whose columns/rows of E will be used to approximate the invariant subspace of E associated with its largest singular values while the remaining nodes are called *fine grid nodes.* The process of detecting a suitable set of coarse grid nodes will be called *coarsening process.* Once one has selected the coarse grid nodes, they are in a natural way embedded in the initial graph. In addition the subset of coarse grid nodes has its own related graph. It has been shown for special case in [6] that among all symmetric positive preconditioners for A of type $M = L(I + PZ^{-1}P^\top)L^\top$ with fixed L and P the choice $Z = P^\top MP$ is (almost) optimal in the sense of quadratic forms. In this case the graph of Z describes a natural graph associated with the coarse grid nodes. We will call it *coarse grid* in analogy to the graph associated with a grid in the numerical solution of partial differential equations.

Recalling the well-known techniques of constructing good preconditioners for the conjugate gradient method applied to symmetric positive definite systems, e.g. [10, 14, 20], we should choose P and Z such that

$$\mu A \leqslant \left(M^{(1)}\right)^{-1} \leqslant \mu\kappa^{(1)} A \tag{7}$$

with $\kappa^{(1)}$ as small as possible and $\mu > 0$. Here $\kappa^{(1)} \geqslant 1$ is the condition number of $M^{(1)}A$, i.e., the ratio of the largest by the smallest eigenvalue of $M^{(1)}A$ and thus $\kappa^{(1)} = 1$ would be optimal. It is well-known that a small condition number is essential for fast convergence of the conjugate gradient method [10].

For a discretized elliptic partial differential equation one can construct optimal preconditioners (with a condition number that does not depend on the fineness of the discretization) using multigrid methods [14]. In order to obtain a similar preconditioner from sparse approximate inverses, consider the use of $LL^\top$ in a linear iteration scheme with initial guess $x^{(0)} \in \mathbb{R}^n$. The iteration scheme [22] for the solution of $Ax = b$ has the form

$$x^{(k+1)} = x^{(k)} + LL^\top(b - Ax^{(k)}), k = 0, 1, 2, \ldots.$$

The error propagation matrix $I - LL^\top A$ satisfies $x - x^{(k+1)} = (I - LL^\top A)(x - x^{(k)})$. In multilevel techniques [13] one uses this iteration for pre and post smoothing and in addition one has to add a coarse grid correction. In terms of the error propagation matrix this means that instead of $I - LL^\top A$ we have $(I - LL^\top A)(I - \hat{P}Z^{-1}\hat{P}^\top A)(I - LL^\top A)$ as error propagation matrix, where $\hat{P} = LP$. This product can be rewritten as $I - M^{(2)}A$ with

$$\begin{aligned} M^{(2)} &= 2LL^\top - LL^\top ALL^\top + (I - LL^\top A)\hat{P}Z^{-1}\hat{P}^\top(I - ALL^\top) \\ &= L\left(2I - M + EPZ^{-1}P^\top E\right)L^\top, \end{aligned} \tag{8}$$

where $M = L^\top AL$ and $E = I - M$. Note that when applying $M^{(2)}$ to a vector x, the matrices A (or M) and $LL^\top$ have to be applied only twice, i.e., the application of this operator is less

expensive than it looks. Again one is interested in choosing P, Z such that

$$\mu A \leqslant \left(M^{(2)}\right)^{-1} \leqslant \mu\kappa^{(2)} A \tag{9}$$

with $\kappa^{(2)}$ as small as possible.

To derive the approximation properties of $M^{(1)}, M^{(2)}$ we first discuss the optimal factors P, Z for given A, L. We use the spectral decomposition of the residual matrix

$$E = I - L^\top AL = I - M = V\Lambda V^\top, \tag{10}$$

where $\Lambda = \operatorname{diag}(\lambda_1 \ldots, \lambda_n)$ with diagonal elements in decreasing order and $V = [v_1, \ldots, v_n]$ is orthogonal.

For a given fixed rank r, it has been shown in [2, 6] that

$$P := [v_1, \ldots, v_r], \; Z := P^\top MP \tag{11}$$

gives an (almost) optimal approximation, i.e., we obtain the minimal $\kappa^{(1)}$ in (7). But this observation is more of theoretical interest, since in practice the spectral decomposition is not available and even if it were available, then it would be very expensive to apply, since the matrix P would be a full matrix. Instead we would like to determine P, Z that are inexpensive to apply and still produce good approximation properties in $M^{(1)}$ and $M^{(2)}$. The specific choice $Z = P^\top MP$ then corresponds to an exact two–level method. A more general choice of Z should satisfy the relation

$$\gamma Z \leqslant P^\top MP \leqslant \Gamma Z$$

to fit into the framework which we will present in Section 3. One can read Z as a perturbed coarse grid system or a perturbed 2–level method, when $P^\top MP$ is not solved exactly. The common choice to perturb $P^\top MP$ would be to substitute $(P^\top MP)^{-1}$ recursively by further multilevel steps. To do this we replace the term Z^{-1} in

$$L\left(I + PZ^{-1}P^\top\right)L^\top, \text{ where } Z = P^\top MP, \tag{12}$$

by a further additive approximation $L_1\left(I + P_1 A_{H,1}^{-1} P_1^\top\right) L_1^\top$, where $L_1 L_1^\top$ is a factored sparse approximate inverse of the exact coarse grid system $P^\top MP$, P_1 is used to define the next coarsening step and $A_{H,1} = (P_1^\top L_1^\top Z L_1 P_1)$. A three–level preconditioner then has the form

$$\begin{aligned} M_2^{(1)} &= L\left(I + P\left(L_1\left(I + P_1 A_{H,1}^{-1} P_1^\top\right) L_1^\top\right) P^\top\right) L^\top \\ &= LL^\top + LPL_1 L_1^\top P^\top L^\top + LPL_1 P_1 A_{H,1}^{-1} P_1^\top L_1^\top P^\top L. \end{aligned}$$

For the construction of higher levels, the procedure is analogous. This leads to the following algebraic multilevel schemes.

Definition 1 *Let $A \in \mathbb{R}^{n,n}$ be symmetric positive definite and let $n = n_l > n_{l-1} > \cdots > n_0 > 0$ be integers. For chosen full rank matrices $\hat{P}_k \in \mathbb{R}^{n_k, n_{k-1}}$, $k = l, l-1, \ldots, 1$, define A_k via*

$$A_k = \begin{cases} A & k = l \\ \hat{P}_{k+1}^\top A_{k+1} \hat{P}_{k+1} & k = l-1, l-2, \ldots, 0. \end{cases}$$

Choose nonsingular $L_k \in \mathbb{R}^{n_k, n_k}$ such that $L_k L_k^\top \approx A_k^{-1}$, $k = 0, \ldots, l$. Then multilevel sparse approximate preconditioners $M_l^{(1)}, M_l^{(2)}$ are recursively defined for $k = 0, 1, 2, \ldots, l$ via

$$M_k^{(1)} = \begin{cases} A_0^{-1} & k = 0 \\ L_k L_k^\top + \hat{P}_k M_{k-1}^{(1)} \hat{P}_k^\top & k > 0 \end{cases} \tag{13}$$

and

$$M_k^{(2)} = \begin{cases} A_0^{-1} & k = 0 \\ L_k(2I - L_k^\top A_k L_k) L_k^\top + (I - L_k L_k^\top A_k)\hat{P}_k M_{k-1}^{(2)} \hat{P}_k^\top (I - A_k L_k L_k^\top) & k > 0 \end{cases} \tag{14}$$

For $l = 1$ we obviously obtain the operators $M^{(1)}$ and $M^{(2)}$. If one exactly decomposes $A_0^{-1} = L_0 L_0^\top$, e.g. by Cholesky decomposition and sets $\Pi_k = \hat{P}_l \hat{P}_{l-1} \cdots \hat{P}_{k+1}$ then one obtains [6]

$$M_l^{(1)} = \sum_{k=0}^{l} \Pi_k L_k L_k^\top \Pi_k^\top \tag{15}$$

and

$$I - M_l^{(2)} A = (I - \Pi_l L_l L_l^\top \Pi_l^\top A) \cdots (I - \Pi_0 L_0 L_0^\top \Pi_0^\top A) \cdots (I - \Pi_l L_l L_l^\top \Pi_l^\top A). \tag{16}$$

One sees from (15) and (16) that $M_l^{(1)}$ can be viewed as *additive multilevel method*, since all the projections Π_k are formally performed simultaneously, while $M_l^{(2)}$ can be viewed as *multiplicative multilevel method*, since the projections Π_k are performed successively.

Operator $M_l^{(2)}$ is immediately derived from the V–cycle method in partial differential equations [13]. For operator $M_l^{(1)}$ a special case occurs when $L_k L_k^\top = \frac{1}{\alpha_k} I$. In this case $E_k = I - \alpha_k A_k$ and choosing columns of E_k can be expressed as applying a permutation $\Phi_k \in \mathbb{R}^{n_k, n_{k-1}}$ to E_k, i.e. $P_k = (I - \alpha_k A_k)\Phi_k$. In this case $M_l^{(1)}$ reduces to

$$M_l^{(1)} = \frac{1}{\alpha_l}(I + \alpha_l P_l M_{l-1} P_l^\top) = \frac{1}{\alpha_l}\left(I + \frac{\alpha_l}{\alpha_{l-1}} P_l \left(I + \alpha_{l-1} P_{l-1} M_{l-2} P_{l-1}^\top\right) P_l^\top\right).$$

For this type of operator in [15] optimal choices for α_k have been discussed with respect to an a priori chosen permutation matrix Φ_k. This kind of operator has also been studied in [1, 2].

In summary, the construction of updated preconditioners can be interpreted as a multilevel scheme. In view of this interpretation we need an analysis of the approximation properties of the resulting operators. In the following section we extend the analysis that was introduced in [6] and sharpen the bounds derived there.

3 Approximation Properties

In this section we discuss the approximation properties of $M^{(1)}$ and $M^{(2)}$ from (6), (8) for the case $l = 1$ and an approximate coarse grid system $Z \approx \hat{P}^\top A \hat{P}$. To simplify notation we will drop the index k from Definition 1.

We first recall the following result:

Theorem 2 *[6] Let $A \in \mathbb{R}^{n,n}$ be symmetric positive definite and let $L \in \mathbb{R}^{n,n}$ be nonsingular such that $M = L^\top A L \leqslant I$. Let $P \in \mathbb{R}^{n,r}$ with $\operatorname{rank}(P) = r$, $\hat{P} = LP$ and let $W \in \mathbb{R}^{n,n-r}$ have $n - r$ be such that $W^\top M P = 0$. Let, furthermore, $Z \in \mathbb{R}^{r,r}$ be symmetric positive definite such that*

$$\gamma P^\top M P \leqslant Z \leqslant \Gamma P^\top M P \tag{17}$$

with positive constants γ, Γ.

i) If

$$W^\top W \leqslant \Delta W^\top M W, \tag{18}$$

then for $M^{(1)}$ as in (13) we have

$$\frac{\gamma}{\gamma + 1} A \leqslant \left(M^{(1)}\right)^{-1} \leqslant \max\{\Gamma, \Delta\} A. \tag{19}$$

ii) If in (17) $\gamma \geqslant 1$ and

$$\begin{bmatrix} 0 & 0 \\ 0 & W^\top M W \end{bmatrix} \leqslant \Delta \left[P, W\right]^\top (M - EME) \left[P, W\right], \tag{20}$$

then for $M^{(2)}$ as in (14) we have

$$A \leqslant \left(M^{(2)}\right)^{-1} \leqslant \max\{\Gamma, \Delta\} A, \tag{21}$$

where $E = I - M$.

For the operator $M^{(1)}$ one can also estimate the condition number of $M^{(1)}A$ in terms of the angle between the invariant subspace associated with the r smallest eigenvalues and P, see [2]. Note that in Theorem 2 we have $\Delta \geqslant 1$, since $M \leqslant I$. Thus, if we set $Z = P^\top MP$ then $\gamma = \Gamma = 1$ and the bounds for $M^{(1)}$ are determined by Δ only. Via (18) we see that the inequality for M is only needed on the subspace W which is the M–orthogonal complement of span P.

In the following result we sharpen the bounds for $M^{(1)}, M^{(2)}$ in Theorem 2.

Theorem 3 *Under the assumptions of Theorem 2 the following holds.*
i) If $\Delta, \hat{\Delta}$ are constants satisfying

$$W^\top W \leqslant \Delta W^\top MW, \;\; W^\top MW \leqslant \hat{\Delta} W^\top M^2 W, \tag{22}$$

then for $M^{(1)}$ as in (13) we have

$$\frac{\gamma}{\gamma+1} A \leqslant (M^{(1)})^{-1} \leqslant \max\{\Gamma, 2\frac{(\Gamma+1)\Delta\hat{\Delta}}{\Delta+\Gamma\hat{\Delta}}\} A. \tag{23}$$

ii) If in (17) $\gamma \geqslant 1$ and $\hat{\Delta}$ is a constant satisfying

$$W^\top MW \leqslant \hat{\Delta} W^\top (M - EME) W, \tag{24}$$

then for $M^{(2)}$ as in (14) we have

$$A \leqslant (M^{(2)})^{-1} \leqslant \Gamma\hat{\Delta} A. \tag{25}$$

Proof. For the proof of i) set $\Delta = \max\{\Gamma, 2\frac{(\Gamma+1)\Delta\hat{\Delta}}{\Delta+\Gamma\hat{\Delta}}\}$. It suffices to show that

$$\left(M^2 + MPZ^{-1}P^\top M\right)^{-1} \leqslant \Delta M^{-1}.$$

Multiplying with $[W, P]^{-1}$ from the left and its transpose from the right and using the fact that $W^\top MP = 0$, we obtain

$$\begin{pmatrix} W^\top M^2 W & W^\top M^2 P \\ P^\top M^2 W & P^\top M^2 P + P^\top MPZ^{-1}P^\top MP \end{pmatrix}^{-1} \leqslant \Delta \begin{pmatrix} W^\top MW & 0 \\ 0 & P^\top MP \end{pmatrix}^{-1}.$$

The diagonal blocks of the left hand side matrix are the inverses M_{11}^{-1}, M_{22}^{-1} of the Schur–complements M_{11}, M_{22}, where

$$M_{22} = P^\top MPZ^{-1}P^\top MP + P^\top M\left(I - MW(W^\top M^2 W)^{-1}W^\top M\right)MP \geqslant \frac{1}{\Gamma} P^\top MP,$$

and

$$\begin{aligned}
M_{11} &= W^\top M^2 W - W^\top M^2 P \left(P^\top M^2 P + P^\top MPZ^{-1}P^\top MP\right)^{-1} P^\top M^2 W \\
&\geqslant W^\top M^2 W - W^\top M^2 P \left(P^\top M^2 P + \frac{1}{\Gamma} P^\top MP\right)^{-1} P^\top M^2 W \\
&\geqslant W^\top M^2 W - \frac{\Gamma}{\Gamma+1} W^\top M^2 P \left(P^\top M^2 P\right)^{-1} P^\top M^2 W \\
&= \frac{1}{\Gamma+1} W^\top M^2 W + \frac{\Gamma}{\Gamma+1} W^\top M \left(I - MP\left(P^\top M^2 P\right)^{-1} P^\top M\right) MW \\
&= \frac{1}{\Gamma+1} W^\top M^2 W + \frac{\Gamma}{\Gamma+1} W^\top M \left(W\left(W^\top W\right)^{-1} W^\top\right) MW \\
&\geqslant \left(\frac{1}{(\Gamma+1)\hat{\Delta}} + \frac{\Gamma}{(\Gamma+1)\Delta}\right) W^\top MW.
\end{aligned}$$

Since for all symmetric positive definite matrices

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{12}^\top & A_{22} \end{pmatrix} \leqslant 2 \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix},$$

inequality (23) follows.

For ii), we set $T = I - M^{1/2}P(P^\top MP)^{-1}P^\top M^{1/2}$, $\tilde{T} = I - M^{1/2}PZ^{-1}P^\top M^{1/2}$ and $\Delta = \Gamma\hat{\Delta}$. T is the exact orthogonal projection to $M^{1/2}W$, which is the orthogonal complement of $M^{1/2}P$ and $\tilde{T}$ can be interpreted as a special perturbation of T. We can see easily that

$$I - M^{1/2}L^{-1}M^{(2)}L^{-\top}M^{1/2} = I - (2M - M^2) - EM^{1/2}PZ^{-1}P^\top M^{1/2}E = E\tilde{T}E.$$

In order to show (25), we have to find $\Delta > 0$ such that

$$\Delta(I - E\tilde{T}E) \geqslant M^{1/2}L^{-1}A^{-1}L^{-\top}M^{1/2} = I,$$

or equivalently

$$E\tilde{T}E \leqslant (1 - \frac{1}{\Delta})I.$$

Note that (24) is equivalent to

$$TE^2T \leqslant (1 - \frac{1}{\hat{\Delta}})I.$$

This can be seen by multiplying with $M^{1/2}$ on both sides and with $[P, W]$ from the right and its transpose from the left. This yields

$$TE^2T \leqslant (1 - \frac{1}{\hat{\Delta}})I$$

or equivalently

$$[P, W]^\top M^{1/2}TE^2TM^{1/2}[P, W] \leqslant (1 - \frac{1}{\hat{\Delta}})[P, W]^\top M[P, W].$$

Again this is equivalent to

$$\begin{bmatrix} 0 & 0 \\ 0 & W^\top EMEW \end{bmatrix} \leqslant (1 - \frac{1}{\hat{\Delta}}) \begin{bmatrix} P^\top MP & 0 \\ 0 & W^\top MW \end{bmatrix}$$

and it follows that

$$ETE = ET^2E \leqslant (1 - \frac{1}{\hat{\Delta}})I.$$

Since $\tilde{T} \leqslant (1 - \frac{1}{\Gamma})I + \frac{1}{\Gamma}T$ it suffices to choose $\Delta > 0$ such that

$$(1 - \frac{1}{\Gamma})EME + \frac{1}{\Gamma}EM^{1/2}TM^{1/2}E \leqslant \left(1 - \frac{1}{\Gamma} + \frac{1}{\Gamma}(1 - \frac{1}{\hat{\Delta}})\right) M \leqslant (1 - \frac{1}{\Delta})M$$

to obtain (25). □

Note that in Theorem 3, if $W^\top W \leqslant \Delta W^\top MW$, then we already have $W^\top MW \leqslant \Delta W^\top M^2W$. Thus $\hat{\Delta} \leqslant \Delta$ and

$$\frac{(\Gamma + 1)\Delta\hat{\Delta}}{\Delta + \Gamma\hat{\Delta}} \leqslant \Delta.$$

In this sense the bounds of Theorem 3 are sharper than the bounds of Theorem 2. But if $\hat{\Delta} \ll \Delta$ then we have

$$\frac{(\Gamma + 1)\hat{\Delta}}{1 + \frac{\Gamma\hat{\Delta}}{\Delta}} \leqslant (\Gamma + 1)\hat{\Delta}$$

and this is almost an equality. So if Z is scaled such that $\Gamma = 1$, then we obtain the sharper bound

$$(M^{(1)})^{-1} \leqslant 4\hat{\Delta} A.$$

In other words, the inequality $W^\top M W \leqslant \hat{\Delta} W^\top M^2 W$ gives much better information on the approximation properties than the inequality $W^\top W \leqslant \Delta W^\top M W$.

We have a similar situation for $M^{(2)}$. Clearly $\hat{\Delta} \leqslant \Delta$. But $\hat{\Delta} \ll \Delta$ is possible.

We will demonstrate the difference between the bounds in Theorems 2 and 3 in the following examples.

Example 2 *Consider the problem*

$$-(au')' = f \text{ in } [0,1],\ u(0) = u(1) = 0,$$

where

$$a(x) = \begin{cases} 10^0 & x \in [0, 1/4) \\ 10^2 & x \in (1/4, 1/2) \\ 10^3 & x \in (1/2, 3/4) \\ 10^4 & x \in (3/4, 1] \end{cases}$$

and on the interfaces $1/4, 1/2, 3/4$ the mean value is taken. We discretize this problem using piecewise linear finite elements and a uniform mesh size $h = 1/512$. We compare the condition number of the preconditioned system matrix for the two–level preconditioners

$$M_i^{(1)} = D(I + P_i A_{H,i}^{-1} P_i^\top)D,\ M_i^{(2)} = D(2I - M + E P_i A_{H,i}^{-1} P_i^\top E)D,$$

where D is the positive square root of the diagonal of A and where P_i is the matrix of columns $2, 4, 6, \ldots, 2i$ of $E = I - \frac{1}{2}M$. Here M is defined via $M = DAD$. We have that $A_{H,i} = P_i^\top M P_i$ is the coarse grid system. Note that except for the nodes located near $1/4, 1/2, 3/4$, the matrix E essentially corresponds to the matrix $\mathrm{tridiag}([\frac{1}{4},\ \frac{1}{2},\ \frac{1}{4}])$. *In Figure 2 we depict the exact condition number of the preconditioned system for preconditioner $M^{(1)}$ as well as the bounds on the condition numbers obtained by Theorems 2 and 3. The horizontal axis shows the number of columns used in P_i.*

Figure 2 shows that for $M^{(1)}$ the bound obtained by Theorem 2 is more pessimistic, especially when more columns are used in P_i. The estimate from Theorem 3 is sharper.

Figure 3 shows the bounds for $M^{(2)}$. In this case one can not distinguish between the exact value and both bounds (up to numerical rounding errors), but it should be noted that the bound of Theorem 3 is the easiest one to obtain.

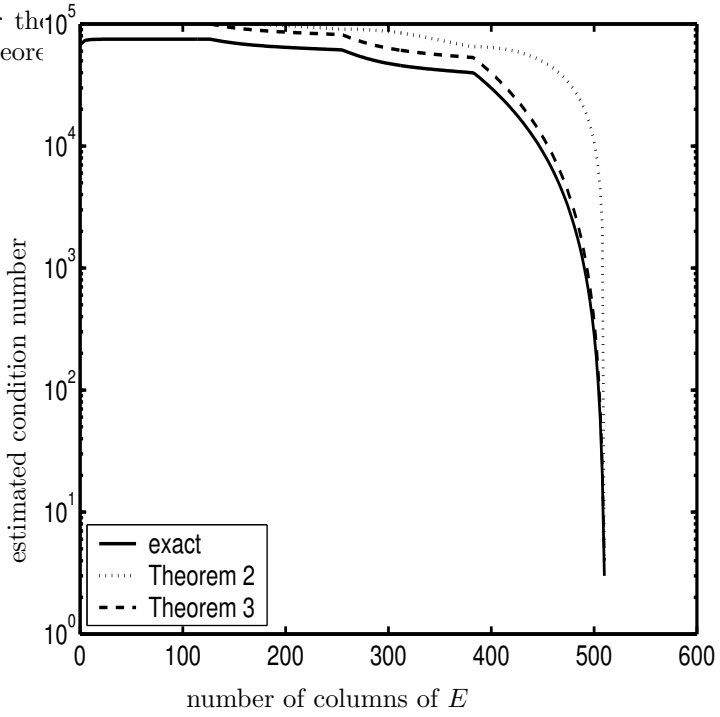
Example 2 illustrates the situation for a simple one–dimensional model problem. Our next example demonstrates that similar results also holds in the two–dimensional case.

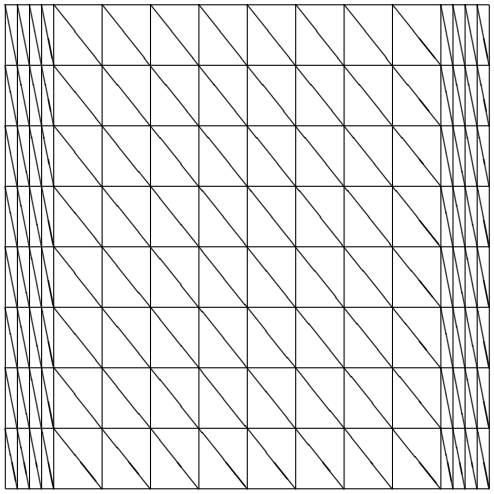
Example 3 *Consider the problem*

$$\begin{aligned} -\varepsilon^2 u_{xx} - u_{yy} &= f \text{ in } [0,1]^2 \\ u &= g \text{ on } \partial[0,1]^2 \end{aligned}$$

where $\varepsilon = 10^{-2}$. Again we use piecewise linear finite elements. The discretization is done using a uniform triangulation with two additional boundary layers of size $\frac{\varepsilon}{4} \times 1$ near the left and also near the right boundary (see the grid below). Within these boundary layers the triangles are condensed by an additional factor $\varepsilon/4$ in x–direction. In y–direction the mesh size is $h = 1/32$.

Figure 2: Bounds for the exact condition number versus Δ (Theore





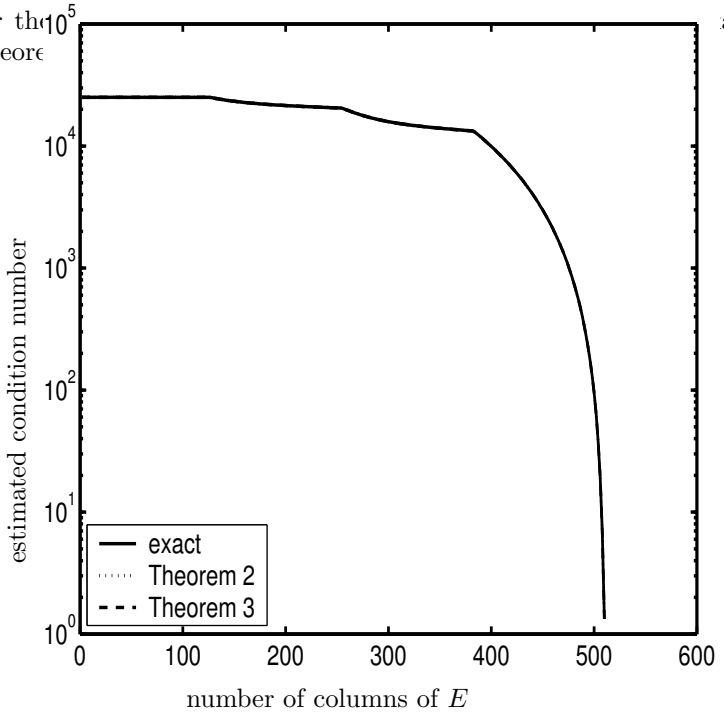
As in Example 2, we compare the exact condition number of the preconditioned system matrix for $M_i^{(1)}$ and $M_i^{(2)}$ with the bounds of Theorems 2 and 3.

Figure 4 shows the comparison for the preconditioner $M^{(1)}$. It shows again that the bound of Theorem 3 is much better than that of Theorem 2.

In Figure 5 we depict the results for $M^{(2)}$ where again the bounds and the exact value are almost indistinguishable but again the one from Theorem 3 is the most easy bound to obtain.

The bounds obtained by Theorem 3 are in general sharper and easier to obtain than those of Theorem 2, since in Theorem 3 both bounds for $\hat{\Delta}$ only require the subspace that is given by the M–orthogonal complement of P. On the other hand the bounds of Theorem 2 can be more easily generalized to the general case $l > 1$, see [6]). It is possible, though more technical, to generalize the sharper bounds of Theorem 3 to the multilevel case ($l > 1$). The difficulties arise from the fact that Γ is not as well isolated in (23) as in (19).

Figure 3: Bounds for the exact condition number versus Δ (Theore



In this section we have derived sharp bounds for the condition numbers of the multilevel preconditioned systems. In practice to determine these bounds still requires the solution of an eigenvalue problem. However, one could estimate $\hat{\Delta}$, e.g. by applying a test vector x to the inequalities (22),(24). For more details see Section 4. The results so far require two spaces (or matrices) P and W that are orthogonal with respect to the inner product defined by $M = L^\top AL$. One can construct a sequence of well–suited matrices P_k in each step k using the results of Theorems 2 and 3. This construction leads to a multilevel hierarchy. This will be the topic of the next section.

4 A Simplified Coarsening Scheme

The approximation bounds that we have derived in Section 3 can be utilized to construct an algebraic multilevel preconditioner by skilfully choosing columns of the residual matrix E. Here we focus only on a sketch of the main idea and the basic ingredients.

The basic components of the construction are:

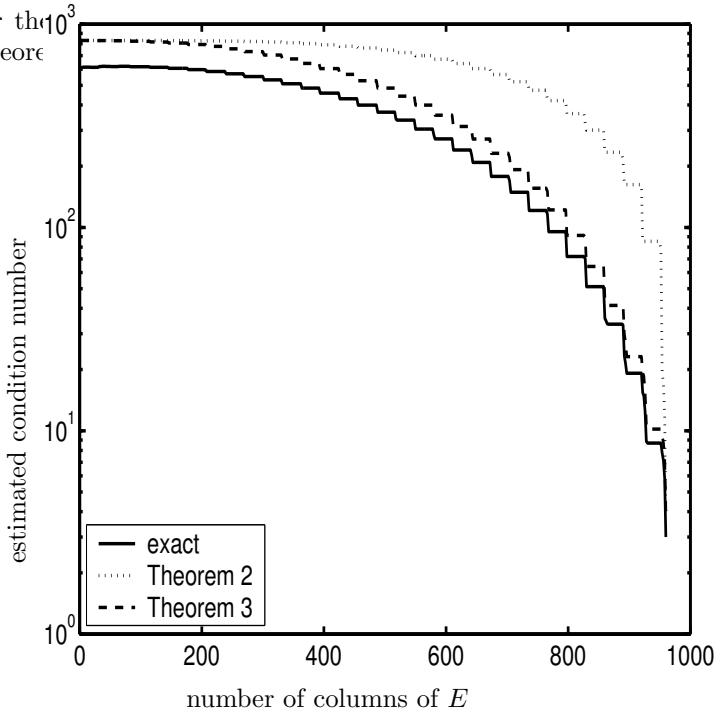
1. the residual matrix $E = I - M = I - L^\top AL$
2. the QR decomposition for E with respect to a special inner product
3. the pivoting strategy to detect suitable columns of E to define P

A detailed (and more technical) description of the implementation can be found in [6]. We refer to this paper for numerical comparisons with other algebraic multilevel methods.

To describe the coarsening process, suppose that we have constructed a QR decomposition with column pivoting of E,

$$E \underbrace{[\Pi_1, \Pi_2]}_{\Pi} = \underbrace{[P, W]}_{Q} \underbrace{\begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix}}_{R},$$

Figure 4: Bounds for the act condition number versus Δ (Theore



where Q is orthogonal in the inner product given by M, i.e., $Q^\top MQ = I$, then we already have $P^\top MW = 0$. P and W fulfill the orthogonality condition of Theorems 2 and 3 and thus the quality of P as coarse grid projection can be measured via the bounds in these two theorems. If we find a suitable permutation of E such that sensible estimates for the bounds of Theorems 2 and 3 are small, then P can serve as coarse grid projection. There are several approaches to compute an approximate QR–decomposition. One possibility is to adapt a QR–like decomposition as in [21] but other constructions are possible as well. We refer to [5, 6] for a detailed description of this quite technical construction. There is one drawback of choosing P itself as coarse grid projection: in general P is not sparse! But Theorems 2 and 3 are independent of the specific choice of the basis in the space that is spanned by the columns of P. Since

$$E\Pi_1 = PR_{11},$$

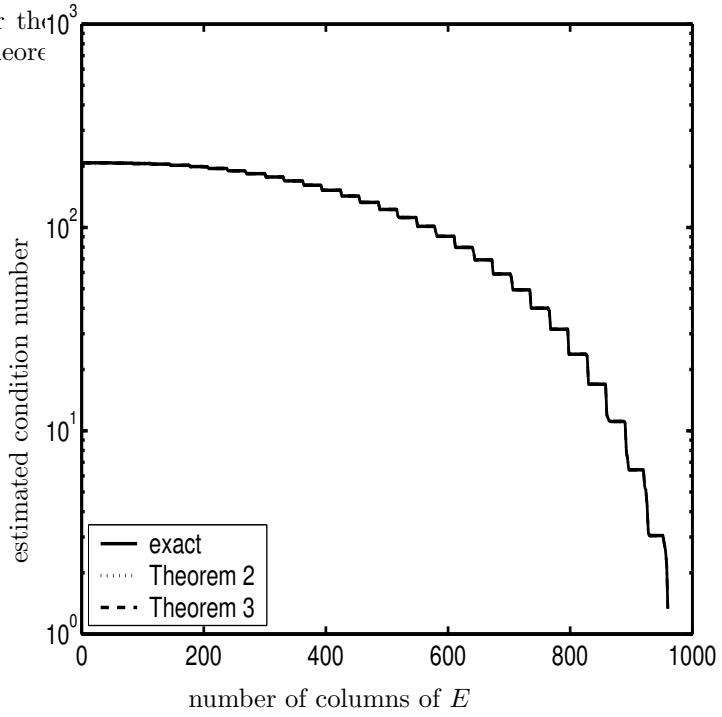
the columns of $E\Pi_1$ and P span the same space. In addition $E\Pi_1$ refers to specific columns of E and therefore we can expect $E\Pi_1$ to be sparse! We use this relation for the coarse grid projection and finally substitute

$$P \to E\Pi_1$$

as coarse grid projection.

Here we concentrate on the pivoting strategy. Beside many technical details in the implementation of a sparse approximate QR factorization in [6], there some important components that make this approach effective and that are not as issue of the implementation. They are related to the method of determining the pivots and even show up when using a full QR decomposition. At first glance the best we can do for pivoting would be to locally maximize Δ in inequalities (18), (20), to obtain a feasible coarse grid matrix $\hat{P} = LP$ (or equivalently $\hat{P} \to LE\Pi_1$) for $M^{(1)}$ in (6) and for $M^{(2)}$ in (7). However, for fixed r there exist $\binom{n}{r}$ permutations which have to be checked and for any of these choices one has to compute a QR decomposition of an $n \times r$ matrix $E\Pi_1$ to get the corresponding Δ. Already for small r the costs are prohibitively expensive, e.g. for $r = 2$, $n(n-1)/2$ possibilities have to be checked. So in practice not more than $r = 1$ can be achieved

Figure 5: Bounds for the exact condition number versus Δ (Theore



in one step. Using $P^\top MP = I$ we set

$$T = I - PP^\top M \tag{26}$$

and it is easy to see that the M–orthogonal complement W of P is given by

$$W = TE\Pi_2. \tag{27}$$

Using T from (26), the bounds (18) and (20) in Theorem 2 can be expressed as

$$\frac{1}{\Delta} = \min_{Tx\neq 0} \frac{x^\top T^\top MTx}{x^\top T^\top Tx}, \tag{28}$$

and

$$\frac{1}{\Delta} = \min_{Tx\neq 0} \frac{x^\top (M - EME)x}{x^\top T^\top MTx}, \tag{29}$$

respectively.

Analogously, for the bounds in Theorem 3 we have

$$\frac{1}{\hat{\Delta}} = \min_{Tx\neq 0} \frac{x^\top T^\top M^2 Tx}{x^\top T^\top MTx}, \tag{30}$$

and

$$\frac{1}{\hat{\Delta}} = \min_{Tx\neq 0} \frac{x^\top T^\top (M - EME)Tx}{x^\top T^\top MTx} \tag{31}$$

respectively.

To compute one of these Rayleigh quotients is still not feasible, since this would have to be done in every step of the QR decomposition. Even more so, if we have already computed r columns of the QR–decomposition, then every choice of column $r+1$ has to be checked, i.e., if

$$P_r = [p_1, \ldots, p_r]$$

are the first r columns of Q in the QR–decomposition, then any choice of

$$T = I - [P_r, p_{r+1}][P_r, p_{r+1}]^\top M$$

has to be checked. This is far too expensive and even if this were possible, there is no guarantee that choosing the locally optimal p_{r+1} will globally give a good contribution to the prolongation matrix P.

The easiest simplification would be to replace the minimum over all x by a specific choice of x. An obvious candidate would be the eigenvector of M associated with its smallest eigenvalue. However, as we will demonstrate in the following examples, local optimization does not necessarily lead to a globally good choice.

Example 4 *Consider the matrix*

$$A = \begin{bmatrix} T & -e_n & 0 \\ -e_n^\top & 1+\alpha & -\alpha e_1^\top \\ 0 & -\alpha e_1 & \alpha T \end{bmatrix}, \text{ where } T = \begin{bmatrix} 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{bmatrix} \in \mathbb{R}^{n,n}.$$

for $\alpha = 100$. This example can be viewed as discretization of the problem $-(au')' = f$ in $[0,1]$ with $a(x) = 1$ for $x \in [0,.5]$ and $a(x) = \alpha$, $x \in [.5,1]$. We will use diagonal preconditioning and the exact eigenvector of the preconditioned system M associated with the smallest eigenvalue. We repeat the coarsening process on every level and switch to Cholesky decomposition, once the QR–decomposition needs more than 75% of the columns of E for the prolongation matrix. Since the coarsening process stops at this stage, we denote this situation in Table 1 with ***"stopped"****.*

Table 1 (as well as the following tables) shows the size of the original system (resp. the sequence of coarse grid matrices) and the related fill–in of these matrices for all levels. The fill–in is measured by the density of the matrix compared with a full system (100% means that the matrix is full). In Table 1 one can observe how much the system is reduced in each level of the coarsening process. We denote by Est_1 the estimate obtained using (28) with the smallest eigenvector and by Est_2 the estimate obtained from (29).

Table 1: Laplace 1–D, Coarsening. Comparison (for each level) of the number of unknowns and the percentage of fill-in (strategy Est_1 versus strategy Est_2)

Strategy	Levels: system size r, fill–in (%)				
	1	2	3	4	5
Est_1	1023	529, 1	527, 1	stopped	—
Est_2		513, 1	257, 1	251, 3	stopped

On the coarser levels $l > 1$ the following happens. The natural choice would be to choose P as every second column of E. This can be seen from the fact that $2P$ exactly corresponds to the prolongation from geometric multigrid (except the first and the last row). This has happened in the first steps for both strategies, but later the strategies choose the remaining nodes near to the coefficient jump in $[0.5,1]$ of the underlying analytic problem. In this form the associated multilevel preconditioners performs very badly.

The situation in Example 4 can also be observed in the two–dimensional case, but here the situation is slightly better.

The example has shown that simply to locally optimize the bounds is not enough. This problem even occurs for the exact QR factorization and represents a fundamental point for the pivoting process. In the sequel we will address this problem and suggest a more effective pivoting strategy.

One way out of this dilemma is to choose multiple columns of E at every step. Since P spans the same space as suitable chosen columns of E, we have that two columns i, j of P or E are M–orthogonal, if their distance is greater than 3 in the graph of M. This can be seen from the fact that E, M have the same graph and $E^\top ME$ may have nonzeros elements only for pairs (i, j) that have a distance less than or equal to 3. Since any two possible choices for p_{r+1} commute if their distance in the undirected graph of M is greater than 3, we can choose as many new nodes in step $r+1$ as there are nodes with distance 4 or more between each other. Hence, after r coarse grid nodes have been chosen, we choose the next coarse grid node such that Δ in (28) is minimized for all T of the form

$$T = I - [P_r, p_{r+1}^{(1)}][P_r p_{r+1}^{(1)}]^\top M$$

Then we continue this procedure for every node of distance greater than 3 from node $r+1$.

$$T = I - [P_r, p_{r+1}^{(1)}, p_{r+1}^{(2)}][P_r, p_{r+1}^{(1)}, p_{r+1}^{(2)}]^\top M$$

We repeat this strategy until there exists no new node with distance greater 3 to all selected nodes of step $r+1$.

Consider again Example 4 using this time multiple columns at one step.

Example 5 *Using multiple columns in Example 4 we obtain the results depicted in Figure 2 that shows the size of the system and the related fill–in for all levels. We denote by Est_{1m} the estimate obtained using (28) with the smallest eigenvector and by Est_{2m} the estimate obtained from (29) in combination with multiple chosen columns.*

Table 2: Laplace 1–D, Coarsening. Comparison (for each level) of the number of unknowns and the fill-in (strategy Est_1 versus strategy Est_2, both with multiple column choice)

Strategy	Levels: system size r, fill–in (%)						
	1	2	3	4	5	6	7
Est_{1m}	1023	512, 1	256, 1	128, 2	63, 5	31, 9	15, 19
Est_{2m}		512, 1	256, 1	128, 2	63, 5	31, 9	15, 19

Here the coarsening process constructs exactly the grid that corresponds to the geometric multigrid case. It follows that the preconditioners are optimal, see Table 3.

Table 3: Laplace 1–D. Preconditioned CG: number of iteration steps and flops for different preconditioners (none/diagonal/both multilevel preconditioners)

Strategy	preconditioned system							
	A		MA		$M_l^{(1)}A$		$M_l^{(2)}A$	
Est_{1m}	5037	$8.2{\cdot}10^7$	1023	$2.1{\cdot}10^7$	26	$1.0{\cdot}10^6$	11	$8.7{\cdot}10^5$
Est_{2m}					26	$1.0{\cdot}10^6$	11	$8.7{\cdot}10^5$

A similar improvement can be observed in the two–dimensional case.

Example 6 *Finally we use an example from the Harwell–Boeing collection. The matrix we use is*

$$A = \begin{bmatrix} D & B & & 0 \\ B^\top & D & \ddots & \\ & \ddots & \ddots & B \\ 0 & & B^\top & D \end{bmatrix}, \; D = \begin{bmatrix} 786432 & 0 \\ 0 & 256 \end{bmatrix}, \; B = \begin{bmatrix} -393216 & 6144 \\ -6144 & 64 \end{bmatrix},$$

which essentially corresponds to the matrix LANPRO/NOS2 from the Harwell–Boeing collection. We use this matrix for $n = 190$ (95 diagonal blocks). Although this matrix is not very big, its condition number is very large and it has large positive off–diagonal entries. This matrix is a real challenge for the coarsening process. Here the multiple column based strategies Est_{1m} and Est_{2m} strategies construct coarse grids which end up in perfect multilevel scheme while the single column based versions Est_1 and Est_2 fail already during the generation of the coarse grid. Essentially Est_1 and Est_2 start with similar nodes as Est_{1m} and Est_{2m} but then they start taking nodes in the neighbourhood of the previously chosen nodes leading to a coarse grid with insufficient reduction. Table 4 shows the coarsening process and Table 5 gives the behaviour of the preconditioned cg method.

Table 4: NOS2, Coarsening. Comparison (for each level) of the number of unknowns and the percentage of fill-in (single column strategies Est_1 and Est_2 versus multiple column strategies Est_{1m}, Est_{2m})

Strategy	Levels: system size r, fill–in (%)				
	1	2	3	4	5
Est_1	190, 2	129, 8	128, 18	stopped	—
Est_2		125, 8	118, 19	stopped	—
Est_{1m}		94, 6	46, 13	22, 26	10, 52
Est_{2m}		94, 6	46, 13	22, 26	10, 52

We should point out that the fast convergence of the cg method for the single column strategy is exclusively based on selecting too many nodes and using the Cholesky decomposition for the final system, where the final system has size 128 and 118 which is pretty close to the size of the original system. The multiple column based strategy impressively demonstrates that the coarsening process can do significantly better. An almost perfect multilevel hierarchy with small number of cg steps is generated.

Table 5: NOS2. Preconditioned CG: number of iteration steps and flops for different preconditioners (none/diagonal/both multilevel preconditioners)

Strategy	preconditioned system					
	A		MA		$M_l^{(1)}A$	$M_l^{(2)}A$
Est_1	1062	$3.9{\cdot}10^6$	540	$2.5{\cdot}10^6$	20 $3.7{\cdot}10^5$	8 $2.4{\cdot}10^5$
Est_2					28 $4.9{\cdot}10^5$	12 $3.4{\cdot}10^5$
Est_{1m}					22 $2.0{\cdot}10^5$	9 $1.8{\cdot}10^5$
Est_{2m}					22 $2.0{\cdot}10^5$	9 $1.8{\cdot}10^5$

These examples have illustrated that simply using the locally optimal column is not necessarily optimal from a global point of view, while using multiple columns at every step was very successful. Another technique introduced in [6] consists of locally locking the neighbouring nodes of a chosen column node in the graph theoretical sense. This leads to an alternative way for improving the local optimization. Numerical examples with an efficient implementation of an approximate QR factorization in [6] show that this kind of algebraic multilevel strategy is competitive with other algebraic multigrid methods especially for non-standard situations, i.e. matrices with many positive off-diagonal entries.

5 Conclusion

We have derived estimates for the condition number and hence for the convergence rates of algebraic multilevel preconditioners and demonstrated their quality by examples. These estimates can be used in the construction of algebraic multilevel methods. To do this one computes a special QR–decomposition with column pivoting, where the pivoting strategy is driven by an estimate of the condition number. This process, however, is in general not monotone. Numerical examples show that this process needs to be supplemented with other heuristic techniques to safeguard the local optimization. These techniques often improve the optimization and in many cases they lead to a good approximation of the global optimum.

References

[1] Owe Axelsson, Maya Neytcheva, and Ben Polman, *An application of the bordering method to solve nearly singular systems*, Vestnik Moskovskogo Universiteta, Seria 15, Vychisl. Math. Cybern. **1** (1996), 3–25.

[2] Owe Axelsson, Alexander Padiy, and Ben Polman, *Generalized augmented matrix preconditioning approach and its application to iterative solution of ill-conditioned algebraic systems*, Technical report, Katholieke Universiteit Nijmegen, Fakulteit der Wiskunde en Informatica, 1999.

[3] Michele Benzi, Carl D. Meyer, and Miroslav Tůma, *A sparse approximate inverse preconditioner for the conjugate gradient method*, SIAM J. Sci. Comput. **17** (1996), 1135–1149.

[4] Michele Benzi and Miroslav Tůma, *A sparse approximate inverse preconditioner for nonsymmetric linear systems*, SIAM J. Sci. Comput. **19** (1998), no. 3, 968–994.

[5] Matthias Bollhöfer and Volker Mehrmann, *A new approach to algebraic multilevel methods based on sparse approximate inverses*, Preprint SFB393/99–22, TU Chemnitz, Germany, Dep. of Mathematics, August 1999.

[6] _____, *Algebraic multilevel methods and sparse approximate inverses*, SIAM J. Matrix Anal. Appl. **24** (2002), no. 1, 191–218.

[7] E. Chow and Y. Saad, *Approximate inverse preconditioners for general sparse matrices*, Research Report UMSI 94/101, University of Minnesota, Super Computing Institute, Minneapolis, Minnesota, 1994.

[8] Iain S. Duff, Roger G. Grimes, and John G. Lewis, *Sparse matrix test problems*, ACM Trans. Math. Software **15** (1989), 1–14.

[9] Roland W. Freund, Gene H. Golub, and Noel M. Nachtigal, *Iterative solution of linear systems*, Acta Numerica (1992), 1–44.

[10] Gene H. Golub and Charles F. Van Loan, *Matrix computations*, third ed., The Johns Hopkins University Press, 1996.

[11] Anne Greenbaum, *Iterative methods for solving linear systems*, Frontiers in Applied Mathematics, SIAM Publications, 1997.

[12] Marcus J. Grote and Thomas Huckle, *Parallel preconditioning with sparse approximate inverses*, SIAM J. Sci. Comput. **18(3)** (1997), 838–853.

[13] Wolfgang Hackbusch, *Multigrid methods and applications*, Springer-Verlag, 1985.

[14] _____, *Iterative lösung großer schwachbesetzter Gleichungssysteme*, second ed., B.G. Teubner Stuttgart, 1993.

[15] Thomas Huckle, *Matrix multilevel methods and preconditioning*, Technical report SFB–Bericht Nr. 342/11/98 A, Technische Universität München, Fakulktät für Informatik, 1998.

[16] I. E. Kaporin, *New convergence results and preconditioning strategies for the conjugate gradient method*, Numer. Lin. Alg. w. Appl. **1(2)** (1994), 179–210.

[17] L.Yu. Kolotilina and A.Yu. Yeremin, *Factorized sparse approximate inverse preconditionings. I. Theory*, SIAM J. Matrix Anal. Appl. **14** (1993), 45–58.

[18] Y. Notay, *Using approximate inverses in algebraic multigrid methods*, Numer. Math. **80** (1998), 397–417.

[19] Y. Saad, *Iterative methods for sparse linear systems*, PWS Publishing, Boston, 1996.

[20] Y. Saad and M.H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput. **7** (1986), 856–869.

[21] G.W. Stewart, *Four algorithms for the efficient computation of truncated pivoted QR approximations to a sparse matrix*, Technical report UMIACS TR-98-12 CMSC TR-3875, University of Maryland, Department of Computer Science, 1998, to appear in Numerische Mathematik.

[22] R. S. Varga, *Matrix iterative analysis*, Prentice-Hall, Englewood Cliffs, New Jersey, 1962.