A preconditioning technique for indefinite linear systems

A Preconditioning Technique for Indefinite Linear Systems

W.H.A. Schilders

Abstract

In this report we present a preconditioning technique which can be used when solving symmetric indefinite linear systems, in augmented form. The coefficient matrices in the resulting preconditioned systems have eigenvalues with positive real parts, and a number of them are equal to unity. When a special ordering is used, all eigenvalues are real.

1 Introduction

Consider linear systems of the form

$$M\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} = \begin{pmatrix} A & B \\ B^T & 0\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} = \begin{pmatrix}\mathbf{b}\\ \mathbf{c}\end{pmatrix}, \tag{1}$$

where A is a positive definite $n \times n$ matrix, and B is an $n \times m$ matrix of full rank. Often, $m < n$. By writing

$$\begin{pmatrix} A & B \\ B^T & 0\end{pmatrix} = \begin{pmatrix} A & 0 \\ B^T & I\end{pmatrix}\begin{pmatrix} A^{-1} & 0 \\ 0 & -B^TA^{-1}B\end{pmatrix}\begin{pmatrix} A & B \\ 0 & I\end{pmatrix},$$

we immediately see that the coefficient matrix in (1) has exactly n positive and m negative eigenvalues. Therefore, the system is indefinite. In [3], (theoretical) lower and upper bounds on these eigenvalues are given in terms of the extremal eigenvalues of A and the extremal singular values of B.

Systems of the form (1) occur in the numerical solution of a large variety of problems. In fact, it is often the basic system of equations, containing relations between different kinds of unknowns and some conservation law. Hence, it is not surprising that many researchers have focussed their attention to the solution of such systems. In Iain Duff's paper [1] a review of methods for augmented systems can be found. Roughly speaking, these methods can be divided into two categories:

- range-space methods
- null-space methods

The former of these methods is probably the most wellknown, and also the most obvious. Namely, the unknown vector $\mathbf{x}$ is eliminated from the first set of equations, making use of the non singularity of the matrix A:

$$\mathbf{x} = A^{-1}(\mathbf{b} - B\mathbf{y}).$$

This expression is substituted in the second set of equations, and leads to:

$$B^T A^{-1} B\mathbf{y} = B^T A^{-1}\mathbf{b} - \mathbf{c}.$$

This is an $m \times m$ positive definite system. Unfortunately, the coefficient matrix is usually not sparse, even if M is a sparse matrix. If an iterative solution method is used, this complication can be avoided, since only matrix vector multiplications are needed. Indeed, multiplying a vector $\mathbf{y}$ by the coefficient matrix $B^T A^{-1} B$ involves two matrix vector multiplications (by B and B^T, respectively), and one solution of a linear system. This means that the iterative solution method must contain an inner loop which solves systems with a coefficient matrix A. Unfortunately, iterative solution methods often converge within a reasonable number of iterations only if some kind of preconditioning is used. Most preconditioning techniques require the coefficient matrix in closed form, so that the possible benefit of iterative methods sketched in the foregoing is lost.

From the foregoing it is clear that range-space methods may not be ideal for solving systems of the form (1) in an efficient way. Therefore, we now turn to the second class of methods for solving augmented systems, the so-called null-space methods. These methods are based on the observation that, instead of eliminating $\mathbf{x}$ from the system, one may also eliminate the unknown vector $\mathbf{y}$. This can be done in the following way. First we observe that a special solution, $B\widehat{\mathbf{y}}$, of the system

$$B^T \mathbf{x} = \mathbf{c},$$

can be found by solving the system

$$B^T B\widehat{\mathbf{y}} = \mathbf{c}.$$

Here, the assumption that B has full column rank ($\text{rank}(B) = m$) is needed. Using this special solution, we find that $\mathbf{x}$ must satisfy

$$B^T(\mathbf{x} - B\widehat{\mathbf{y}}) = \mathbf{0},$$

so that $\mathbf{x} - B\widehat{\mathbf{y}}$ must be in the null-space of B^T. Suppose now that a basis for the null-space is known: $\{\mathbf{c}_1, ..., \mathbf{c}_{n-m}\}$. Then there exist coefficients z_1, ...,z_{n-m} such that

$$\mathbf{x} - B\widehat{\mathbf{y}} = \sum_{i=1}^{n-m} z_i \mathbf{c}_i. \tag{2}$$

Denoting the matrix with columns $\mathbf{c}_1$,...,$\mathbf{c}_{n-m}$ by C and the vector of unknown coefficients z_i by $\mathbf{z}$, (2) can be rewritten as:

$$\mathbf{x} = C\mathbf{z} + B\widehat{\mathbf{y}}.$$

Substituting this in the first set of equations, we then find:

$$AC\mathbf{z} + B\mathbf{y} = \mathbf{b} - AB\widehat{\mathbf{y}}.$$

If we now multiply this equation by C^T, and make use of the fact that $C^T B = 0$, we finally obtain

$$C^T AC\mathbf{z} = C^T(\mathbf{b} - AB\widehat{\mathbf{y}}),$$

which is an $(n-m)\times(n-m)$ system of equations for the unknown vector $\mathbf{z}$. As in the case of the range-space methods, the coefficient matrix C^TAC is positive definite and symmetric. If both C and A are sparse, the coefficient matrix will also be sparse. Again, problems similar to those described for the range-space methods occur: preconditioning of the system usually requires the coefficient matrix in closed form. Fortunately, however, it is sometimes possible here to directly assemble the matrix C^TAC. This is the case, for example, in electrical circuit simulation. The matrix B^T then contains information about the topology of the circuit, and its null space consists of so-called loop currents which can be readily identified. This also holds in certain areas of electromagnetics [4].

From the foregoing it is clear that both range space methods and null space methods suffer from the problem that the coefficient matrix is not known explicitly, but only as a product of matrices. This makes preconditioning rather difficult. In some cases, successful approaches have been described. For the bi-harmonic equation, the product B^TB is a matrix which closely resembles A, so that $B^TA^{-1}B$ is a good approximation of the identity matrix. In this case, the range space method works reasonably well. For most other cases, however, we must resort to different strategies for preconditioning. In the next section, we will describe a new preconditioning technique for augmented systems, and prove a number of attractive properties of this method.

2 A new preconditioning technique

The basic idea of the preconditioning method is to construct a preconditioning matrix which has the same number of positive and negative eigenvalues as the original matrix. If this is the case, it may be expected that the coefficient matrix of the preconditioned system (which is the inverse of the preconditioning matrix applied to the original matrix) only has positive eigenvalues, so that conjugate gradient type methods can be applied. The problem is whether the aforementioned goal can be achieved. Fortunately, it can be shown that this is the case for the method developed. In the following, the method will be described, and a summary of theoretical results obtained will be given.

Consider the matrix

$$M = \begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix},$$

where $A \in R^{n\times n}$ is symmetric and positive definite, and $B \in R^{n\times m}$ such that

$$B_{i.j} \in \{-1, 0, 1\} \quad \forall_{1\le i\le n, 1\le j\le m},$$

where $m \le n$. In addition, we assume that each row of B contains at most two non-zero elements of opposite sign:

$$\sum_{j=1}^{m} |B_{i,j}| \le 2,$$

$$-1 \le \sum_{j=1}^{m} B_{i,j} \le 1.$$

Finally, we also assume that $\mathrm{rank}(B) = m$.

Now let $P : \{1, ..., n\} \to \{1, ..., n\}$ be a permutation with the property that

$$B_{P(i),i} \neq 0, \qquad i \leq m.$$

Then we define the permutation matrix Q by

$$Q = \left(\mathbf{e}_{P(1)}, \mathbf{e}_{n+1}, ..., \mathbf{e}_{P(m)}, \mathbf{e}_{n+m}, \mathbf{e}_{P(m+1)}, ..., \mathbf{e}_{P(n)}\right),$$

where $\mathbf{e}_i \in R^{n+m}$ is the i-th unit vector.

After permutation of rows and columns, we obtain the matrix

$$\tilde{M} = Q^T M Q,$$

the "diagonal" of which being of the form

$$\text{"diag"}(\tilde{M}) = \begin{pmatrix} A_{P(1),P(1)} & B_{P(1),1} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ B^T_{1,P(1)} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & A_{P(m),P(m)} & B_{P(m),m} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & B^T_{m,P(m)} & 0 & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & A_{P(m+1),P(m+1)} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & A_{P(n),P(n)} \end{pmatrix}.$$

Systems of the form

$$M \begin{pmatrix} x_1 \\ . \\ . \\ . \\ x_n \\ y_1 \\ . \\ . \\ y_m \end{pmatrix} = \begin{pmatrix} b_1 \\ . \\ . \\ . \\ b_n \\ c_1 \\ . \\ . \\ c_m \end{pmatrix}, \tag{3}$$

transform to

$$\tilde{M} \begin{pmatrix} x_{P(1)} \\ y_1 \\ . \\ . \\ x_{P(m)} \\ y_m \\ x_{P(m+1)} \\ . \\ x_{P(n)} \end{pmatrix} = \begin{pmatrix} b_{P(1)} \\ c_1 \\ . \\ . \\ b_{P(m)} \\ c_m \\ b_{P(m+1)} \\ . \\ b_{P(n)} \end{pmatrix}. \tag{4}$$

In order to find a suitable preconditioning technique for the indefinite system (3), we first transform it into the form (4) and propose a generalised incomplete

Crout preconditioning for this system. After having found this, the preconditioning matrix is transformed back.

The preconditioning matrix for the system (4) is cast into the form

$$P = (\tilde{L} + \tilde{D})\tilde{D}^{-1}(\tilde{L} + \tilde{D})^T, \tag{5}$$

where

$$\tilde{L} = \text{"lower"}(\tilde{M}),$$

and

$$\tilde{D} = \begin{pmatrix} \tilde{D}_1 & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \tilde{D}_m & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \tilde{d}_{m+1} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \tilde{d}_n \end{pmatrix}.$$

Here, "lower" is to be understood in a sense which matches "diag". This means that $\tilde{L}$ consists of blocks $\tilde{L}_{i,j}$ which are in $R^{2\times 2}$ if $1 < j \le i \le m$, in $R^{2\times 1}$ if $m+1 \le i \le n, 1 \le j \le m$ and in $R^{1\times 1}$ if $m+1 < j \le i \le n$. In the first of these cases, it is easy to check that

$$\tilde{L}_{i,j} = \begin{pmatrix} A_{P(i),P(j)} & B_{P(i),j} \\ B^T_{i,P(j)} & 0 \end{pmatrix}.$$

The matrices $\tilde{D}_1, ..., \tilde{D}_m$ and the scalars $\tilde{d}_{m+1}, ..., \tilde{d}_n$ are required to be such that

$$\text{"diag"}\left((\tilde{L} + \tilde{D})\tilde{D}^{-1}(\tilde{L} + \tilde{D})^T\right) = \text{"diag"}(\tilde{M}).$$

<u>Lemma 1</u>:
There exist $\tilde{d}_1, ..., \tilde{d}_m$ such that, for $1 \le i \le m$,

$$\tilde{D}_i = \begin{pmatrix} \tilde{d}_i & B_{P(i),i} \\ B^T_{i,P(i)} & 0 \end{pmatrix}.$$

<u>Proof</u>:
The proof proceeds by induction. It is easily verified that

$$\tilde{D}_1 = \begin{pmatrix} A_{P(1),P(1)} & B_{P(1),1} \\ B^T_{1,P(1)} & 0 \end{pmatrix},$$

so that $\tilde{d}_1 = A_{P(1),P(1)}$. Now assume that $\tilde{D}_1, ..., \tilde{D}_{i-1}$ are of the desired form (where $2 \le i \le m$). Then $\tilde{D}_i$ is determined by the condition

$$\begin{pmatrix} A_{P(i),P(i)} & B_{P(i),i} \\ B^T_{i,P(i)} & 0 \end{pmatrix} = \tilde{D}_i + \sum_{j=1}^{i-1} \tilde{L}_{i,j}\tilde{D}_j^{-1}\tilde{L}^T_{i,j}.$$

By the induction hypothesis and the fact that $B^2_{P(j),j} = 1$ for all $1 \leq j \leq m$, we find that

$$\tilde{D}_j^{-1} = \begin{pmatrix} 0 & B_{P(j),j} \\ B^T_{j,P(j)} & -\tilde{d}_j \end{pmatrix}.$$

Hence,

$$\tilde{L}_{i,j}\tilde{D}_j^{-1}\tilde{L}_{i,j}^T = \begin{pmatrix} 2A_{P(i),P(j)}B_{P(j),j}B_{P(i),j} - \tilde{d}_j B^2_{P(i),j} & B_{P(i),j}B_{P(j),j}B_{P(j),i} \\ B_{P(i),j}B_{P(j),j}B_{P(j),i} & 0 \end{pmatrix}.$$

Now suppose that $B_{P(i),j}B_{P(j),j}B_{P(j),i} \neq 0$. Then $B_{P(j),i} \neq 0$ and $B_{P(i),j} \neq 0$. By assumption on the permutation, we also have that $B_{P(i),i} \neq 0$ and $B_{P(j),j} \neq 0$. Thus, $B_{P(j),j}$ and $B_{P(j),i}$ are the only non zero elements in the row $P(j)$, whereas $B_{P(i),j}$ and $B_{P(i),i}$ are the only non zero elements in row $P(i)$. Hence, either subtracting or adding these rows leads to a row consisting entirely of zeroes. This contradicts the assumption that $\text{rank}(B) = m$. Thus, $B_{P(i),j}B_{P(j),j}B_{P(j),i} = 0$, and we conclude that

$$\tilde{L}_{i,j}\tilde{D}_j^{-1}\tilde{L}_{i,j}^T = \begin{pmatrix} 2A_{P(i),P(j)}B_{P(j),j}B_{P(i),j} - \tilde{d}_j B^2_{P(i),j} & 0 \\ 0 & 0 \end{pmatrix}.$$

So,

$$\tilde{D}_i = \begin{pmatrix} \tilde{d}_i & B_{P(i),i} \\ B^T_{i,P(i)} & 0 \end{pmatrix},$$

with

$$\tilde{d}_i = A_{P(i),P(i)} + \sum_{j=1}^{i-1} B^2_{P(i),j}\tilde{d}_j - 2A_{P(i),P(j)}B_{P(j),j}B_{P(i),j}.$$

This completes the proof.

Remark 1:
The above lemma demonstrates that it is possible to find the first n 2×2 blocks in the preconditioning matrix. The remaining part of the preconditioning may be constructed in several ways, depending on properties of the original system.

Remark 2:
The constraints on the matrix B can be somewhat relaxed; if, after permutation, the top part (of dimension $m \times m$) is a non-singular lower triangular matrix, the lemma also holds.

Remark 3:
If B has the properties listed in the above, it can be shown that a permutation matrix Q exists which renders the top part of the matrix B lower triangular.

The linear system to be solved is now multiplied by the preconditioning matrix as given in (5), so that a system with coefficient matrix $P^{-1}M$ has to be solved. Note that, due to the fact that this matrix will not be symmetric in general, the (symmetric) conjugate gradient method cannot be used. Hence, CGS, biCGSTAB or related methods must be used. Notwithstanding the fact that these methods are designed to solve nonsymmetric systems of equations, a desirable property is that the eigenvalues of the coefficient matrix are all located in one half plane through the origin. Fortunately, it turns out that the matrix $P^{-1}M$ has this property. In fact, a detailed analysis reveals the following facts, which have been proven theoretically under the assumption (this is not a limitation in view of Remark 3) that the top part of the matrix B is lower triangular (after permutation):

- the matrix $P^{-1}M$ has at least $2m$ eigenvalues $\lambda = 1$
- if the preconditioning of the remaining $(n - m) \times (n - m)$ part of the matrix exists, all eigenvalues of $P^{-1}M$ are real and positive

If the matrix B does not have the form indicated, the first of these facts remains. However, pairs of complex conjugate eigenvalues may occur. From experiments it follows that the real parts of these eigenvalues are again positive; however, this has not yet been verified theoretically. The proof of the above facts is rather involved, and will be published in a forthcoming report.

3 Conclusion

From the discussion above, it follows that the proposed preconditioning technique has desirable properties, and is very suitable for linear systems in which the coefficient matrix is indefinite. Because matrices arising in circuit problems are of a similar structure, the preconditioning technique can easily be generalised to this case.

The method can not only be used for preconditioning purposes; it can also be used to avoid pivoting in direct solution methods, since the 2×2 diagonal blocks are always non-singular and can serve as 2×2 pivots. In this sense, the method resembles that of Bunch-Kaufman-Parlett [2], but in this more general method the number of 2×2 pivots is not known and must be determined during the solution process. The advantage of our method is that it is known a priori how many 2×2 pivots are necessary.

References

[1] I.S. Duff, The solution of augmented systems, *Report RAL-93-084*, Rutherford Appleton Laboratory, Chilton, Didcot (1993)

[2] G.H. Golub and C.F. van Loan, *Matrix Computations*, John Hopkins Univ. Press, London (1989)

[3] T. Rusten and R. Winther, A dynamic circuit partitioning and iterative methods for saddlepoint problems, *ACM Trans. Math. Software*, vol. 17, pp. 189-193 (1991)

[4] M.E. Verbeek, Iterative solvers and preconditioning for electromagnetic boundary integral equations, *Ph.D. Thesis*, Utrecht University (2001)