SOME NEW METHODS OF SOLVING THE EQUATION AX + XB = C WITH THE AID OF SIMILARITY TRANSFORMATIONS

By: Antony Jameson

Technical Memorandum ENG FS MGR 70-4
May, 1970

The matrix equation

$$AX + XB = C \tag{1}$$

in which the unknown matrix X and the given matrix C are $M \times N$ matrices, A is an $M \times M$ matrix, and B is an $N \times N$ matrix, is encoutered in a wide range of applications. It occurs, for example, in the design of observers in control theory [1], and it stability theory [2] in symmetric form, with $B = A^T$, and C symmetric. It also arises in the analysis of beam gridworks [3].

if u and v are eigenvectors of A and B^T with the corresponding eigenvalues λ and μ, then

$$Auv^T + uv^T B = (\lambda + \mu)\, uv^T$$

Thus the eigenvalues of the system (1) are $\lambda_i + \mu_j$, where λ_i and μ_j are the eigenvalues of A and B, and (1) can be solved uniquely if and only if

$$\lambda_i + \mu_j \neq 0 \tag{2}$$

for all i, j. If the system (1) is written in full as a set of equations of MN unknowns, its solution by Gaussian elimination requires a number of multiplications proportional to $(MN)^3$. It is obviously desirable to find some alternative method to avoid this extremely rapid growth with the size of the unknown matrix. One approach is to eliminate A and B by using their characteristic equations [4]. This may suffer from a loss of precision due to the introduction of high powers of the matrices. Another approach is to simplify (1) by preliminary similarity transformations of either A or B, or both A and B. Four such methods will be discussed. The first, the well known method in which both A and B are reduced to diagonal form [4], [5], is included for the purpose of comparison. The other methods require less complicated preliminary transformations, at the expense of some increase in the subsequent computations.

Let P and Q be matrices similar to A and B, so that

$$A = UPU^{-1}\,, \quad B = VQV^{-1} \tag{3}$$

Then (1) can be written as

$$PY + YQ = D \tag{4}$$

where

$$Y = U^{-1}XV\,, \quad D = U^{-1}CV \tag{5}$$

If (4) can be solved for Y, then X is recovered by a single inversion from (5).

The most complete simplification of (1) is used in method 1.

Method 1: diagonalization of A and B

If U and V can be found such that

$$P = \mathrm{diag}\,(\lambda_i)\,, \quad Q = \mathrm{diag}\,(\mu_i)$$

where λ_i and μ_j are the eigenvalues of A and B, then (4) becomes

$$(\lambda_i + \mu_j)\,y_{ij} = d_{ij} \tag{6}$$

which is solved trivially. This method is effective as long as the required U and V can be found. It is well known that this is possible if the λ_i are distinct, and the μ_j are alsi distinct [6, p 4].

If either A or B has multiple eigenvalues, the reduction to diagonal form may be impossible. If the system is not symmetric, and only one of A and B can be diagonalized, one may resort to method 2.

Method 2: diagonalization of A or B

Let V be such that

$$Q = \text{diag}\,(\mu_j)$$

and set U equal to I. Then if y_j and d_j are the j^{th} columns of Y and D, equation (4) reduces to

$$(A + \mu_j I)\, y_j = d_{ij} \tag{7}$$

Thus Y is obtained by solving N sets of M equations, requiring a number of multiplications proportional to NM^3. It is evident that equations (7) can be solved if condition (2) is satisfied. Diagonalization of A leads to similar set of equations for the rows of Y, requiring for their solution a number of multiplications proportional to MN^3.

The reduction to diagonal form, if at all possible, may be expensive. Method 3 calls only for reduction to triangular form. It is well known [6, p 46] that the required similarity transformation always exists. The QR algorithm, for example, is a procedure for generating the necessary transformations which can be proved to converge for arbitrary matrices whether or not they have multiple eigenvalues [6, p 540].

Method 3: Triangularization of A and B

Let U and V be such that P is in lower triangular form, with the eigenvalue λ_i of A on its diagonal, and Q is in upper trigular form, with the eigenvalues μ_j of B on its diagonal. The method of solving (4) for Y is most conveniently demonstrated by an example. Suppose that Y is a 3×3 matrix. Then equation (4) can be expanded as

$$\begin{bmatrix} \lambda_1 & 0 & 0 \\ p_{21} & \lambda_2 & 0 \\ p_{31} & p_{32} & \lambda_3 \end{bmatrix} \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \end{bmatrix} + \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \end{bmatrix} \begin{bmatrix} \mu_1 & q_{12} & q_{13} \\ 0 & \mu_2 & q_{23} \\ 0 & 0 & \mu_3 \end{bmatrix} = \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix}$$

Thus

$$\begin{aligned} \lambda_1 + y_{11} + y_{11}\mu_1 &= d_{11} \\ p_{21}y_{11} + \lambda_2 y_{21} + y_{21}\mu_1 &= d_{21} \\ p_{31}y_{11} + p_{32}y_{21} + \lambda_3 y_{31} + y_{31}\mu_1 &= d_{31} \\ \lambda_1 y_{12} + y_{11}q_{12} + y_{12}\mu_2 &= d_{12} \end{aligned} \tag{8}$$

$$p_{21}y_{12} + \lambda_2 y_{22} + y_{21}q_{12} + y_{22}\mu_2 \quad = \quad d_{22}$$

whence the elements y_{ij} can be evaluated successively as

$$\begin{aligned}
y_{11} &= \frac{d_{11}}{\lambda_1 + \mu_1} \\
y_{21} &= \frac{d_{21} - p_{21}y_{11}}{\lambda_2 + \mu_1} \\
y_{31} &= \frac{d_{31} - p_{31}y_{11} - p_{32}y_{21}}{\lambda_3 + \mu_1} \\
y_{12} &= \frac{d_{12} - y_{11}q_{12}}{\lambda_1 + \mu_2} \\
y_{22} &= \frac{d_{31} - p_{21}y_{12} - y_{21}q_{12}}{\lambda_2 + \mu_2} \\
\ldots
\end{aligned}$$

Evidently condition (2) is sufficient to ensure a solution. The elements of the first column require successively $0, 1, 2, \ \ldots \ M-1$ multiplications, those of the second column successively $1, 2, 3, \ \ldots \ M$ multiplications, and so on. Thus the total number of multiplications is

$$\frac{1}{2}M\,(M{-}1 + M{+}1 \ \ldots \ + M{+}2N{-}3) = \frac{1}{2}MN\,(M{+}N{-}2)$$

It is immaterial whether successive rows or columns of Y are evaluated, the dependence on previous elements being in the direction of the diagonal. If A and B are both reduced to upper triangular form, the elements may be evaluated sequentially in a similar manner, starting with y_{M1}. In fact, if A and B are reduced to any combination of triangular forms, the same method can be used, starting from one of the four corners of Y.

In the symmetric case, when $B = A^T$, it is of course desirable to preserve the symmetry by setting $V = U^T$, so that $Q = P^T$, and Y is symmetric. It is then only necessary to evaluate the off-diagonal elements of Y on one side of the diagonal. Also, since $p_{ij} = q_{ji}$ and $y_{ij} = y_{ji}$, only half as many multiplications are required in the evaluation of the diagonal elements. The total number of multiplications is thus exactly half the number required for a non-symmetric case, that is $\frac{1}{2}N^2\,(N-1)$.

If the system is not symmetric, a substantial simplication results if either A or B alone is reduced to triangular for. This leads to method 4, which saves the expense of finding two separate triangularization matrices.

Method 4: Triangularization of A or B

Let V be such that Q is in upper triangular form, with the eigenvalues μ_j of B on its diagonal, and set U equal to I. Again using a 3×3 example to illutrate the procedure, equation (4) can now be expanded as

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \end{bmatrix} + \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \end{bmatrix} \begin{bmatrix} \mu_1 & q_{12} & q_{13} \\ 0 & \mu_2 & q_{23} \\ 0 & 0 & \mu_3 \end{bmatrix} = \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix}$$

If y_j and d_j are the columns of Y and D, it can be seen that it is possible to solve for y_j successively from the equation

$$\begin{aligned} (A + \mu_1 I)\, y_1 &= d_1 \\ (A + \mu_2 I)\, y_2 &= d_2 - q_{12} y_1 \\ (A + \mu_3 I)\, y_3 &= d_3 - q_{13} y_1 - q_{23} y_2 \end{aligned} \tag{9}$$

In general the number of multiplications required to set up the right hand sides is

$$M\,(0 + 1\ \ldots\ + N - 1) = \frac{1}{2} MN\,(N - 1)$$

and as with method 2 the solution of the equations requires a number of multiplications proportional to NM^3. Again it is evident that condition (2) ensures that each of the equations can be solved. If A is reduced to triangular form, a corresponding set of equations is obtained for the rows of Y.

One of the most efficient methods of determining the transformation to diagonal form is to obtain the eigenvalues by a preliminary triangularization, and to construct the required transformation matrix with the corresponding eigenvectors as its columns. After accomplishing the preliminary transformation, the additional labour required to complete the solution by method 3 in comparison with method 1, or method 4 in comparison with method 2, appears to be less than the additional labour needed to compute the eigenvectors in methods 1 and 2. For this reason, and because they also avoid the possibility of failure due to a deficiency of eigenvectors, the triangularization methods are attractive in comparison with the diagonalization methods. Given a matrix of order N, the number of multiplications required to carry our a preliminary reduction to Hessenberg form is proportional to N^3. Then the number of multiplications required to carry out a double step of the QR algorithm and accumulate the transformation matrix is intially proportional to N^2, decresing for the later iterations as subdiagonal elements are eliminated. The additional labour in carrying out an extra trigularization in method 3 is thus likely to be less expensive than the additional labour required to complete the solution by method 4. Assuming that as the order of the matrix is increased, the growth in the number of iterations required by the QR algorithm is roughly linear, the number of multiplications required by method 3 to complete a solution for a square matric can be expected to be proportional to N^3.

References:

1. D. G. Luenberger, Observers for multivariable control systems, I.E.E.E. Trans. on Automatic Control, AC 11 (1966), pp. 190-197.

2. S. Lefschetz and J. P. LaSalle, Stability by Lyapunov's Direct Method, Academic Press, New York, 1961.

3. Er-Chieh Ma, A finite series solution of the matrix equation $AX - XB = C$, S.I.A.M. J. on Appl. Math., 14 (1966) pp. 490-495.

4. A. Jameson, Solution of the equation $AX + XB + C$ by inversion of $M \times M$ or $N \times N$ matrix, S.I.A.M. J. on Appl. Math., 16 (1968), pp. 1020-1023.

5. F. R. Gantmacher, The Theory of Matrices, Vol. 1, Chelsea, New York, 1959.

6. J. H. Wilkinson, The Algebraic Eigenvalue Problem, Clarendon Press, Oxford, 1965.