ON THE SOLUTION OF A CLASS OF TOEPLITZ SYSTEMS*

Chen Ming-kui

(*Department of Mathematics Xi'an Jiaotong University, Xi'an, China*)

Abstract

The solution of certain Toeplitz linear systems is considered in this paper. This kind of systems are encountered when we solve certain partial differential equations by finite difference techniques and approximate functions using higher order splines. The methods presented here are more efficient than the Cholesky decomposition method and are based on the circulant factorisation of the symmetric "banded circulant" matrix, the Woodbury formula and the algebraic perturbation method.

1. Introduction

We consider a linear system of the form

$$Ax = f, \tag{1.1}$$

where the coefficient matrix is an nth order symmetric banded matrix of Toeplitz form

$$A_t = \begin{bmatrix} \alpha_0 & \alpha_1 & & \cdots & \alpha_p & & & & \\ \alpha_1 & \ddots & \ddots & & \ddots & & & & \\ \vdots & \ddots & \ddots & \ddots & & \ddots & & & \\ \alpha_p & & \ddots & \ddots & \ddots & & \ddots & & \\ & \ddots & & \ddots & \ddots & \ddots & & \ddots & \\ & & \ddots & & \ddots & \ddots & \ddots & & \alpha_p \\ & & & \ddots & & \ddots & \ddots & \ddots & \vdots \\ & & & & \ddots & & \ddots & \ddots & \alpha_1 \\ & & & & & \alpha_p & \cdots & \alpha_1 & \alpha_0 \end{bmatrix}, \tag{1.2}$$

* Received December 5, 1986.

or the symmetric "banded circulant" form

$$A_c = \begin{bmatrix} \alpha_0 & \alpha_1 & & \cdots & \alpha_p & & \alpha_p & \cdots & \alpha_1 \\ \alpha_1 & \ddots & \ddots & & \ddots & & & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \ddots & & & \alpha_p \\ \alpha_p & & \ddots & \ddots & \ddots & & \ddots & & \\ & \ddots & & \ddots & \ddots & \ddots & & \ddots & \\ & & \ddots & & \ddots & \ddots & \ddots & & \alpha_p \\ \alpha_p & & & \ddots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & & & \ddots & & \ddots & \ddots & \alpha_1 \\ \alpha_1 & \cdots & \alpha_p & & & \alpha_p & \cdots & \alpha_1 & \alpha_0 \end{bmatrix}, \tag{1.3}$$

$x = (x_1, x_2, \cdots, x_n)^T$ is the unknown n-vector, and f is the given right-hand side.

This class of linear systems occurs in solving a certain kind of boundary value problems by finite difference techniques, in solving biharmonic equations by the Fourier method, and in higher order spline approximation [2, 3, 4, 5, 6, 11].

System (1.1) with coefficient matrix of form (1.2) can be solved by band Cholesky decomposition [7] or by Toeplitz factorization [6]. Although the operation counts of the two methods are about the same, the latter requires less storage. If the system has a coefficient matrix of form (1.3), then the Cholesky decomposition is expensive, and the circulant factorization presented here is more favorable in terms of not only arithmetic operations but also storage requirements. The methods presented in this paper are based on the fact that under certain conditions the matrix in (1.3) can be factored into two simpler circulant matrices, and the corresponding circulant system may then be solved by using the Woodbury formula [8]. Furthermore, the banded Toeplitz matrix may be treated as a perturbation of a circulant matrix, and Toeplitz systems can be solved by the combination of the circulant factorization and algebraic perturbation method [9].

In §2, we will describe the method for factoring a symmetric banded circulant matrix into two circulant matrices. This factorization was used to solve the band circulant system in [3]. The methods for solving band Toeplitz systems will be studied in §3, and finally, some numerical results will be given in §4.

2. Factorization of Banded Circulant Matrices

To factor the banded circulant matrix given by (1.3) we consider the real function with the elements of the matrix as its coefficients

$$\Phi(z) = \alpha_p z^p + \cdots + \alpha_1 z + \alpha_0 + \alpha_1 z^{-1} + \cdots + \alpha_p z^{-p}, \tag{2.1}$$

the characteristic function of matrix A_c. Assume, without loss of generality, that $\alpha_p = 1$. We have that following theorem.

Theorem 2.1. *If matrix A_C is strictly diagonally dominant, i.e.* $|\alpha_0| > 2(|\alpha_1| + \cdots + |\alpha_p|)$, *then there exists a real polynomial* $l(z) = \beta_0 + \beta_1 z + \cdots + \beta_p z^p$, $|\beta_0| > \beta_p|$, *with all roots outside the unit circle such that the characteristic function* $\Phi(z)$ *can be factored as*

$$\Phi(z) = l(z) \cdot l(z^{-1}). \tag{2.2}$$

Proof. We show first that the function $\Phi(z)$ has no root on the unit circle. If there exists a number z_0 on the unit circle which is a root of the equation

$$\Phi(z) = 0, \tag{2.3}$$

then $z_0 = e^{i\theta}$ for some real $\theta, 0 < \theta < 2\pi$. Substituting z_0 into (2.3) we have

$$\begin{aligned} \alpha_0 &= -\big[\alpha_1(e^{i\theta} + e^{-i\theta}) + \cdots + \alpha_p(e^{ip\theta} + e^{-ip\theta})\big] \\ &= -2\big[\alpha_1 \cos\theta + \cdots + \alpha_p \cos p\theta\big]. \end{aligned}$$

It follows that

$$|\alpha_0| \le 2(|\alpha_1| + \cdots + |\alpha_p|),$$

which is a contradiction to the assumption of the theorem.

We now note that

$$\Phi(z) = \Phi(z^{-1}),$$

and (2.3) is a reciprocal equation [1]. Thus if z_0 is a root of (2.3), then so is z_0^{-1}. It follows that $\Phi(z)$ has p pairs of roots $z_1^{(k)}, z_2^{(k)}$, such that

$$z_1^{(k)} = 1/z_2^{(k)}, \qquad k = 1, 2, \cdots, p,$$

and $z_1^{(k)}$ are outside the unit circle.

Let

$$l(z) = \prod_{k=1}^{p} (z - z_1^{(k)}). \tag{2.4}$$

We now prove that $l(z)$ is a real polynomial. If all the roots $z_1^{(k)}$ are real, then $l(z)$ is real; if some of $z_1^{(k)'}s$ are complex, then their conjugate complex numbers, which are outside the unit circle too, are the roots of (2.3) since the coefficients of the equation are real. So it is obvious that $l(z)$ is a real polynomial and satisfies (2.2), and the proof is completed.

It is easy to verify that the corresponding circulant matrix A_c can be factored as

$$A_C = \tilde{L}\tilde{L}^T, \tag{2.5}$$

where

$$\tilde{L} = \begin{bmatrix} \beta_0 & & \beta_p & \cdots & \cdots & \beta_1 \\ \vdots & \ddots & & & \ddots & \vdots \\ \beta_p & & \ddots & & & \beta_p \\ & \ddots & & \ddots & & \\ & & \ddots & & \ddots & \\ & & & \beta_p & \cdots & \beta_0 \end{bmatrix}.$$

To compute the factor $l(z)$, we solve equation (2.3). When $p=2$, it is well known [1, 5] that the roots of equation (2.3) are given by

$$\begin{cases} \rho_1 = \frac{1}{2}\Big(\eta_1 + (\eta_1^2 - 4)^{1/2}\Big), \\ \rho_2 = \frac{1}{2}\Big(\eta_1 - (\eta_1^2 - 4)^{1/2}\Big), \\ \rho_3 = \frac{1}{2}\Big(\eta_2 + (\eta_2^2 - 4)^{1/2}\Big), \\ \rho_4 = \frac{1}{2}\Big(\eta_2 - (\eta_2^2 - 4)^{1/2}\Big), \end{cases} \tag{2.6}$$

where

$$\begin{cases} \eta_1 = \frac{1}{2}\Big(- \alpha_1 + (\alpha_1^2 - 4\alpha_0 + 8)^{1/2}\Big), \\ \eta_2 = \frac{1}{2}\Big(- \alpha_1 - (\alpha_1^2 - 4\alpha_0 + 8)^{1/2}\Big). \end{cases} \tag{2.7}$$

Having computed the roots we choose two roots the absolute values of which are greater than 1, say $z_1^{(1)}$ and $z_1^{(2)}$, and form the coefficients of the factor $l(z)$ via

$$\begin{cases} \beta_0 = z_1^{(1)} z_1^{(2)}, \\ \beta_1 = -(z_1^{(1)} + z_1^{(2)}), \\ \beta_2 = 1/\beta_0. \end{cases} \tag{2.8}$$

When p is greater then 2, it is natural to solve (2.3) by using some numerical method and then use the relation between the roots and coefficients to calculate the factor $l(z)$, but it is preferable to compute $l(z)$ directly. Since (2.2) is equivalent to

$$\sum_{j=0}^{p-i} \beta_j \beta_{j+i} = \alpha_i, \qquad i = 0, 1, \cdots, p, \tag{2.9}$$

we can solve (2.9), which is a system of nonlinear equations, for $\beta_j' s$. If we denote

$$f_i(b) = \sum_{j=0}^{p-i} \beta_j \beta_{j+i} - \alpha_i, \qquad i = 0, 1, \cdots, p,$$

then (2.9) can be written as

$$f(b) = 0, \tag{2.10}$$

where $f(b) = (f_0(b), f_1(b), \cdots, f_p(b))^T, b = (\beta_0, \beta_1, \cdots, \beta_p)^T$, and the application of the Newton-Raphson method to system (2.10) gives

$$b^{(m+1)} = b^{(m)} - T(b^{(m)})^{-1} f(b^{(m)}), \tag{2.11}$$

or

$$T(b^{(m)}) b^{(m+1)} = T(b^{(m)}) b^{(m)} - f(b^{(m)}) = T_1(b^{(m)}) b^{(m)} - \alpha,$$

where

$$T = \left[\frac{\partial f_i}{\partial \beta_j}\right] = T_1 + T_2,$$

$$T_1 = \begin{bmatrix} \beta_0 & \beta_1 & \cdots & \beta_p \\ \beta_1 & & \cdot & \\ \vdots & \cdot & & \\ \beta_p & & & \end{bmatrix}, \qquad T_2 = \begin{bmatrix} \beta_0 & \beta_1 & \cdots & \beta_p \\ & \ddots & & \vdots \\ & & \ddots & \beta_1 \\ & & & \beta_0 \end{bmatrix},$$

and $\alpha = (\alpha_0, \alpha_1, \cdots, \alpha_p)^T$.

It has been shown [10] that with the starting values

$$\beta_0^{(0)} = (\alpha_0 + 2\sum_{i=1}^{p} \alpha_i)^{1/2}, \quad \beta_i^{(0)} = 0, \quad i = 1, 2, \cdots, p,$$

the iteration (2.12) is always convergent.

3. Band Toeplitz Systems

The band Cholesky decomposition is an efficient method for solving general band symmetric systems [7], and it can of course be used to solve the symmetric band Toeplitz system

$$A_t x = f. \tag{3.1}$$

But the application of this method to Toeplitz systems not only costs a lot of arithmetic operations but also requires a great amount of storage since it does not take the advantage of the structure of the Toeplitz matrix. Fischer et al. [6] proposed the Toeplitz factorization method for the solution of band Toeplitz systems, which has some advantages in terms of arithmetic operations and storage requirements. In this section we will use the circulant method given in [3] to develop an alternative to the Toeplitz factorization for solving band Toeplitz system (3.1), and use the name BCS to refer to the algorithm Banded Circulant Solver (see [3] for details).

Algorithm BCS can be modified to compute the inverse of a banded circulant matrix. Since A_c is a symmetric circulant matrix, its inverse A^{-1} is also a symmetric circulant, which is uniquely defined by its first column, that is the solution of the equation

$$A_c u = (1, 0, \cdots, 0)^T. \tag{3.2}$$

The algorithm BCS may directly be employed to solve equation (3.2). But in this case the first two steps of the algorithm are essentially the same, so the algorithm for inverting a banded circulant matrix requires $O(4pn)$ operations.

The banded Toeplitz matrix A_t may be considered to be a $(2p)$-rank perturbation of the banded circulant matrix A_c, i.e.

$$A_t = A_c - \begin{pmatrix} I_p \\ 0 \end{pmatrix} U (0^T I_p) - \begin{pmatrix} 0 \\ I_p \end{pmatrix} U^T (I_p 0^T), \tag{3.3}$$

where

$$U = \begin{bmatrix} \alpha_p & \cdots & \alpha_1 \\ & \ddots & \vdots \\ & & \alpha_p \end{bmatrix}.$$

Substituting (3.3) into (3.1) we have

$$A_c x - \begin{pmatrix} I_p \\ 0 \end{pmatrix} U(0^T I_p)x - \begin{pmatrix} 0 \\ I_p \end{pmatrix} (I_P 0^T)x = f. \tag{3.4}$$

If matrix A_t is strictly diagonally dominant, then so is the corresponding circulant matrix A_c, and therefore A_c is nonsingular, and from (3.4) we háve

$$x - A_c^{-1} \begin{pmatrix} I_p \\ 0 \end{pmatrix} U(0^T I_p)x - A_c^{-1} \begin{pmatrix} 0 \\ I_p \end{pmatrix} U^{(T)} (I_p 0^T)x = A_c^{-1} f. \tag{3.5}$$

Let $x^{(1)} = (x_1, \cdots, x_p)^T, x^{(2)} = (x_{p+1}, \cdots, x_{n-p})^T, x^{(3)} = (x_{n-p+1}, \cdots, x_n)^T$, and

$$B_1 = A_c^{-1} \begin{pmatrix} I_p \\ 0 \end{pmatrix},$$

$$B_3 = A_c^{-1} \begin{pmatrix} 0 \\ I_p \end{pmatrix},$$

which are the n-by-p submatrices consisting of the first and the last p columns of matrix A_c^{-1}, respectively. Then equation (3.5) becomes

$$x = y + B_1 U_x^{(3)} + B_3 U^T x^{(1)}, \tag{3.6}$$

which shows that the solution to equation (3.1) is the linear combination of the solution of the corresponding circulant system

$$A_c y = f \tag{3.7}$$

and the first p and the last p colums of the inverse of the corresponding circulant matrix.

The solution to (3.7) can be obtained by algorithm BCS in $O(5pn)$ operations, and the inverse of A_c can be calculated in $O(4pn)$ operations. The inverse A_c^{-1} is, as we pointed out above, symmetric circulant and defined by its first column, the elements of which are denoted by $u_1, u_2, \cdots, u_n$ satisfying

$$u_{n-i} = u_{i+2}, \qquad i = 0, 1, \cdots, \lfloor (n-1)/2 \rfloor,$$

where $\lfloor z \rfloor$ is the integer floor function of z. We then have

$$A_c^{-1} = \begin{bmatrix} u_1 & u_2 & \cdots & \cdots & u_n \\ u_2 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & u_2 \\ u_n & \cdots & \cdots & u_2 & u_1 \end{bmatrix},$$

and therefore

$$B_1 = \begin{bmatrix} u_1 & \cdots & \cdots & u_p \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \vdots & & & u_1 \\ u_{n-p+1} & & & \vdots \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ u_n & \cdots & \cdots & u_{n-p+1} \end{bmatrix}, \tag{3.8}$$

and

$$B_3 = \begin{bmatrix} u_{n-p+1} & \cdots & \cdots & u_n \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \vdots & & & u_{n-p+1} \\ u_1 & & & \vdots \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ u_p & \cdots & \cdots & u_1 \end{bmatrix}. \tag{3.9}$$

To compute the first p and the last p components of the unknown vector x, we premultiply equation (3.6) by $(I_p 0^T)$ and $(0^T I_P)$, respectively, resulting in the following linear system

$$\begin{cases} (I_p - M_{1p}U^T)x^{(1)} - M_{11}Ux^{(3)} = y^{(1)}, \\ -M_{11}U^T x^{(1)} + (I_p - M_{1p}^T U)x^{(3)} = y^{(3)}, \end{cases} \tag{3.10}$$

where M_{11} and M_{1p} are pth order submatrices of A_c^{-1} at the left upper and right upper corner, respectively, i.e.

$$M_{11} = \begin{bmatrix} u_1 & \cdots & \cdots & u_p \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ u_p & \cdots & \cdots & u_1 \end{bmatrix}, \quad M_{1p} = \begin{bmatrix} u_{n-p+1} & \cdots & \cdots & u_n \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ u_{n-2p+2} & \cdots & \cdots & u_{n-p+1} \end{bmatrix},$$

and $y^{(1)}, y^{(3)}$ are p-vectors with the first and the last p components of vector y as their elements, respectively.

Forming the coefficients of equation (3.10) will cost $O(2p^2)$ operations and (3.10) can be solved by Gaussian elimination with $O(8p^3)$ operations. Having calculated $y, u, x^{(1)}$ and $x^{(3)}$, the subvector $x^{(2)}$ can be obtained via (3.6) with $O(2pn)$ operations. performing the iteration (2.12) once would cost $O(p^3)$ operations, and there are usually a few iterations needed for convergence. So the amount of work to compute the factor $l(z)$ would be $O(p^3)$, the same order of solving equation (3.10), and hence the total amount of work to solve symmetric band Toeplitz systems by using the algorithm is $O(11pn) + O(p^3)$. Comparing with the band Cholesky method, which requires $O(np^2 + 3np) - O(p^3)$ operations, when $8 < p << n$, the algorithm is more favorable than Cholesky factorization. In this case, the asymptotic operation count of the algorithm would be $O(11pn)$. The algorithm thus proceeds as folows.

Algorithm BTS (Band Toeplitz Solver) solves the symmetric band Toeplitz system (3.1). Assume that the parameters $\beta_0, \beta_1, \cdots, \beta_p$ are precomputed.

1. Solve for y equation (3.7) by using algorithm BCS.
2. Compute the first column vector u of A_c^{-1} using algorithm BCS.
3. Form and solve equation (3.10) for $x^{(1)}$ and $x^{(3)}$.
4. Compute vector $x^{(2)}$ via (3.6), which along with $x^{(1)}$ and $x^{(3)}$ is the solution.

endalgorithm.

4. Numerical Experiments

The algorithms described in this paper were tried on the APVAX of the Department of Computer Science, Yale University, and compared with Toeplitz factorization and Cholesky decomposition. The program was written in FORTRAN.

To obtain some insight into the accuracy of the algorithms, we generated a number of vectors randomly, which were considered to be the "exact" solutions, and multiplied them by the coefficient matrices to generate the corresponding right hand sides. The equations were solved by using the algorithms BTS and BCS as well as the Toeplitz factorization and the Cholesky method. In all the experiments the results differ from the "exact" solutions only in the last digit, indicating that the algorithms presented in this paper and [3] are stable.

In our all tests we let $p = 2$ and choose several matrices satisfying the assumption in Theorem 2.1. The execution time of algorithm BTS and of the Toeplitz factorization are almost the same. In solving circulant systems the algorithm BCS is about twenty times faster than the Cholesky method in our tests, and saves a lot of storages.

Acknowledgment. This work was completed while the author was visiting the Department of Computer Science, Yale University. The author would like to thank Professor Tony F. Chan and Professor Martin H. Schultz for making such a pleasant stay at the department.

References

[1] W. S. Burnside, The Theory of Equations, Dover Publications, Inc., New York, 1960.

[2] Chen Ming-kui, Modified doublesweep method for solving special tridiagonal systems of linear equations, *Journal of Xi'an Jiaotong University*, 16 : 5 (1982), 85–94.

[3] Chen Ming-kui, On the solution of circulant linear systems, *SIAM J. Numer. Anal.*, **24** (1987), 668–683.

[4] D. J. Evans, On the solution of certain Toeplitz tridiagonal linear systems, *SIAM J. Numer. Anal.*, **17** (1980), 675–680.

[5] H. B. Fine, College Algebra, Dover Publications, Inc., New York 1961.

[6] D. Fischer, G. Golub, O. Hald, C. Leiva , O. Widlund, On Fourier-Toeplitz methods for separable elliptic problems, *Math. Comp.*, **28** (1974), 349–368.

[7] G. H. Golub, C. F. Van Loan, Matrix Computation, The Johns Hopkins University Press, Baltimore. Maryland, 1983.

[8] A. S. Householder, The Theory of Matrices in Numerical Analysis, Blaisdell, New York, 1964.

[9] L. B. Rall, Perturbation methods for the solution of linear problems, M. Z. Nashed ed., Functional Analysis Methods in Numerical Analysis, Springer, 1979.

[10] G. Wilson, Factorisation of the covariance generating function of a pure moving average process, *SIAM J, Numer. Anal.*, **6** (1969), 1–7.

[11] W. L. Wood, Periodicity effects on the iterative solution of elliptic difference equations, *SIAM J, Numer. Anal.*, **8** (1971), 439–464.