CONTROL–THEORY AND ADVANCED TECHNOLOGY
Vol 10, No. 4, Part 5, pp. 2005–2016, December, 1995

Contributed Paper

ON COMPUTING HIGH ACCURACY SOLUTIONS OF A CLASS OF RICCATI EQUATIONS*

A. VARGA[1]

Abstract. The paper is concerned with the computation with high accuracy of the non-negative definite stabilizing solution of the matrix Riccati equation $A^TX + XA + XQX + R = 0$, where A is stable, $Q = Q^T \geq 0$ and $R = R^T \geq 0$. The limiting accuracy solution can be compnted by using an iterative refinement technique based on Newton's method. A proof of global convergence of the iterative process, with final quadratic rate, is given. Several ways to compute with high accuracy the Cholesky factor of the solution are also discussed.

Key Words—Algebraic matrix Riccati equation, Newton's method, spectral factorization, iterative refinement, square-root methods.

1. Introduction

The *algebraic matrix Riccati equation* (AMRE) of the form

$$A^TX + XA + XQX + R = 0, \tag{1}$$

where A is a stable $n \times n$ matrix ($\mathrm{Re}(\lambda_i(A) < 0,\ i = 1, 2, \cdots, n$), $Q = Q^T \geq 0$ and $R = R^T \geq 0$, arises in solving spectral factorization problems. Consider, for instance, a stable continuous-time state-space realization (A, B, C, D) of a system with m inputs and p outputs ($p \leq m$), having the transfer-function matrix $G(s)$ given by

$$G(s) = C(sI - A)^{-1}B + D.$$

Let $\Phi(s) = G(s)G^T(-s)$ be the power spectrum matrix, and let $W(s)$ be a square minimum phase righr spectral factor of Φ, satisfying

$$\Phi(s) = G(s)G^T(-s) = W^T(-s)W(s).$$

By assuming that $\mathrm{rank}(D) = p$, a state-space realization (A_w, B_w, C_w, D_w) of $W(s)$ is given by (Anderson, 1967),

$$(A_w, B_w, C_w, D_w) = (A, BD^T + PC^T, E^{-\frac{1}{2}}(C - B_w^TX), E^{\frac{1}{2}}),$$

* Received by the editors January 19, 1994 and in revised form February 10, 1995.
This work was partly carried out with the assistance of the Alexander von Humboldt Foundation, Bonn, F.R.G.

[1] Institute for Robotics and System Dynamics, German Aerospace Research Establishment (DLR), Oberpfaffenhofen, P.O. Box 1116, D-82230, Wessling, Germany.

where

$$E = DD^T,$$

and P and X satisfy the equations,

$$AP + PA^T + BB^T = 0, \tag{2}$$

$$A^TX + XA + (C - B_w^TX)^T(DD^T)^{-1}(C - B_w^TX) = 0. \tag{3}$$

P and X are, respectively, the controllability gramian of the system (A, B, C, D) and the observability gramian of the system (A, B_w, C_w, D_w). The last equation can be rewritten as

$$(A - B_wE^{-1}C)^TX + X(A - B_wE^{-1}C) + XB_wE^{-1}B_w^TX + C^TE^{-1}C = 0, \tag{4}$$

which is in the form (1) with $A - B_wE^{-1}C$ stable. This equation belongs to a class of Riccati equations specific to spectral factorization problems, with the last two terms in (4) having the same signs.

The need to solve (4) with high accuracy arises in conjunction with the square-root methods for model reduction (Safonov and Chiang, 1988; Varga and Fasol, 1993), where we have to compute the Cholesky factors of the gramians P and X with comparable accuracy in order to ensure the effectiveness of the square-root accuracy enhancing approach. The main difficulties are encountered when the original system is not observable or is nearly unobservable because in this case the gramian X is singular, or nearly singular, respectively. The difficulties are caused by the fact that the rank information on X usually is recovered later, after multiplying this matrix with P or after multiplying the corresponding Cholesky factors. Therefore, in these applications, it is very important that such rank decisions be based on results computed with the highest achievable accuracy.

The standard method to compute the stabilizing, non-negative definite solution $\bar{X}$ of (1) is the *Schur vectors method* proposed by Laub (1979). $\bar{X}$ is computed as

$$\bar{X} = Z_2Z_1^{-1}, \tag{5}$$

where $Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$ is a $2n \times n$ matrix with orthonormal columns which span the stable eigenspace of the Hamiltonian matrix,

$$H = \begin{bmatrix} A & Q \\ -R & -A^T \end{bmatrix}. \tag{6}$$

Although each computational step to determine $\bar{X}$ can be performed by using numerically stable procedures, the Schur method cannot be considered unconditionally numerically stable. It was pointed out by Petkov et al. (1987) that even for well conditioned problems, poor accuracy solutions can occasionally result. As pointed out by Kenney et al. (1989), an appropriate scaling of the Hamiltonian matrix improves sometimes the accuracy of the solution, but, unfortunately, there exists no generally applicable straightforward scaling procedure. Therefore, we are faced frequently with computed solutions whose accuracy is lower than the highest attainable accuracy permitted by the conditioning of the problem. In order to attain this limiting accuracy solution, an iterative refinement technique, similar to that suggested by Arnold and Laub (1984) and Kenney et al. (1990), is usually necessary to be performed.

In this paper, we investigate the use of Newton's method for solving equation (1) and for iterative refinement of approximate solutions. The main result is the theorem presented in the next section, which states the global convergence of this method with a final quadratic convergence rate. The result is similar to those obtained by Kleinman (1968) and Sandell (1974) for the standard Riccati equation, where the constant and quadratic terms have different signs. In Sec. 3, we discuss several possibilities to compute the Cholesky factor of the limiting accuracy solution. An example illustrating the effectiveness of the proposed approach is given in Sec. 4. Some conclusions are presented in Sec. 5.

2. Newton's Iterative Method

We assume that the AMRE (1) has a unique positive semi-definite stabilizing solution; that is, there exists $\bar{X} \geq 0$ such that $A + Q\bar{X}$ is stable. The existence condition of such a solution of (1), provided A is stable, is that the Hamiltonian matrix H in (6) has no eigenvalues on the imaginary axis (Kano and Nishimura, 1993). In what follows, we assume this condition fulfilled.

The problem addressed in this section is how to solve iteratively the Eq. (1), provided an approximation X_{j-1} of $\bar{X}$ is available. Consider the Lyapunov equation,

$$(A + QX_{j-1})^TX_j + X_j(A + QX_{j-1}) + R - X_{j-1}QX_{j-1} = 0, \tag{7}$$

which produces the next iterate X_j when applying Newton's method to solve iteratively (1). The following theorem is the main result of the paper.

Theorem 1. Let $X_0 = X_0^T$ be a stabilizing approximation of $\bar{X} = \bar{X}^T \geq 0$, the unique stabilizing solution of the AMRE (1). Let X_j be the symmetric solution of the Lyapunov equation (7), for $j = 1, 2, \cdots$. Then:

(a) $X_1 \leq X_2 \leq \cdots \leq X_j \leq X_{j+1} \leq \cdots \leq \bar{X}$.

(b) $\lim_{j\to\infty} X_j = \bar{X}$.

(c) There exists $\gamma > 0$ such that $\|\bar{X} - X_j\| \leq \gamma\|\bar{X} - X_{j-1}\|^2$.

In (c) and in what follows, $\|\cdot\|$ stays for any consistent matrix norm. In order to prove the theorem, we prove first the following two lemmas.

Lemma 1. If X_{j-1} is a stabilizing approximation of $\bar{X}$, then the solution X_j of (7) satisfies $X_j \leq \bar{X}$, and $A + QX_j$ is stable.

Proof. By subtracting (7) from (1), where we replaced X by $\bar{X}$, we obtain

$$(A + QX_{j-1})^T(\bar{X} - X_j) + (\bar{X} - X_j)(A + QX_{j-1}) + X_{j-1}QX_{j-1} + \bar{X}Q\bar{X} - X_{j-1}Q\bar{X} - \bar{X}QX_{j-1} = 0$$

or

$$(A+QX_{j-1})^T(\bar{X}-X_j)+(\bar{X}-X_j)(A+QX_{j-1})$$
$$+(\bar{X}-X_{j-1})Q(\bar{X}-X_{j-1})=0.$$

By applying Lyapunov's theorem (Barnett, 1971), we have $\bar{X}\geq X_j$.

By rewriting differently the above expression, we obtain

$$(A+QX_j)^T(\bar{X}-X_j)+(\bar{X}-X_j)(A+QX_j)$$
$$+(X_j-X_{j-1})Q(X_j-X_{j-1})+(\bar{X}-X_j)Q(\bar{X}-X_j)=0.$$

By applying again Lyapunov's theorem (Barnett, 1971), the stability of $A+QX_j$ follows because $\bar{X}-X_j\geq 0$.

Lemma 2. If X_j and X_{j+1} are successive approximations computed from a stabilizing approximation X_{j-1} of $\bar{X}$, then X_j and X_{j+1} satisfy $X_j\leq X_{j+1}\leq\bar{X}$.

Proof. From Lemma 1, it follows that $X_j\leq\bar{X}$ and $X_{j+1}\leq\bar{X}$, and, thus, we have to show that $X_j\leq X_{j+1}$. Let $E(Y)$ be the residual for Y of the AMRE (1), defined by

$$E(Y)\triangleq A^TY+YA+YQY+R. \tag{8}$$

The residual for X_j, the solution of Eq. (7), is

$$\begin{aligned}E(X_j)&=A^TX_j+X_jA+X_jQX_j+R\\&=(A+QX_{j-1})^TX_j+X_j(A+QX_{j-1})+R-X_{j-1}QX_{j-1}\\&\quad-X_{j-1}QX_j-X_jQX_{j-1}+X_jQX_j+X_{j-1}QX_{j-1}\\&=(X_j-X_{j-1})Q(X_j-X_{j-1})\geq 0.\end{aligned}$$

From the equation satisfied by X_{j+1}, we have

$$(A+QX_j)^T(X_{j+1}-X_j)+(X_{j+1}-X_j)(A+QX_j)+E(X_j)=0.$$

By applying Lyapunov's theorem, we obtain $X_j\leq X_{j+1}$.

Proof of Theorem 1. The proof of the theorem parallels a similar proof for the standard Riccati equation done by Mehrmann (1991). Lemma 1 shows that for a given stabilizing approximation X_0 of $\bar{X}$, the sequence $X_1, X_2, \cdots$, is well defined; that is,

$$X_j\leq\bar{X},\quad j=1,2,\cdots,$$

and the matrices $A+QX_j$, $j=1,2,\cdots$, are stable. By Lemma 2, the sequence $X_1, X_2, \cdots$, is nondecreasing and bounded above by $\bar{X}\geq 0$, the stabilizing solution of (1). Therefore, by the Bolzano-Weierstraß theorem, this sequence is convergent to the unique stabilizing solution $\bar{X}$. Thus, points (a) and (b) of the theorem are proven.

In order to prove (c), we subtract (7) from (1), in which we replaced X by $\bar{X}$, to obtain

$$(A+Q\bar{X})^T(\bar{X}-X_j)+(\bar{X}-X_j)(A+Q\bar{X})$$
$$+(X_j-X_{j-1})Q(X_j-X_{j-1})-(\bar{X}-X_j)Q(\bar{X}-X_j)=0.$$

This implies

$$\begin{aligned}0&\leq\bar{X}-X_j\\&\leq\int_0^\infty\exp(t(A+Q\bar{X})^T)[(X_j-X_{j-1})Q(X_j-X_{j-1})]\exp(t(A+Q\bar{X}))dt\\&\leq\|X_j-X_{j-1}\|^2\int_0^\infty\exp(t(A+Q\bar{X})^T)Q\exp(t(A+Q\bar{X}))dt\\&\triangleq\|X_j-X_{j-1}\|^2Y,\end{aligned}$$

where $Y\geq 0$ satisfies

$$(A+Q\bar{X})^TY+Y(A+Q\bar{X})+Q=0.$$

Thus, we obtain

$$\begin{aligned}\|\bar{X}-X_j\|&\leq\|Y\|\,\|X_j-X_{j-1}\|^2\\&=\gamma\|X_j-X_{j-1}\|^2\\&\leq\gamma\|\bar{X}-X_{j-1}\|^2,\end{aligned}$$

since $X_j-X_{j-1}\leq\bar{X}-X_{j-1}$.

Remarks:

(i) A straightforward choice for initializing the sequence X_j, $j=1,2,\cdots$ is $X_0=0$, the matrix A being stable. With this initialization, $X_1\geq 0$ and Lemma 1 insure a monotonic convergence to the solution $\bar{X}$. By (c), the convergence has finally a quadratic rate.

(ii) If X_0 is an arbitrary stabilizing approximation of $\bar{X}$ (computed, for instance, with the Schur vectors method), we have a guaranteed monotonic convergence only after the first iteration. This aspect was already observed by Kenney et al. (1990) for the case of a standard Riccati equation. Thus, generally, we cannot assume that $X_1\geq 0$, and, in fact, X_1 can be even a negative definite matrix. A similar example to that given in Kenney et al. (1990) can be easily constructed for the considered Riccati equation (1). The stabilizing solution for the AMRE ($A=-I$, $Q=I$, $R=0.75I$)

$$-2X+X^2+0.75I=0$$

is $\bar{X}=0.5I$. With $\varepsilon>0$ and $X_0=(1-\varepsilon)I$, the matrix $A+QX_0=-\varepsilon I$ and thus is stable. The first approximation X_1 is

$$X_1=-\frac{(1-\varepsilon)^2-0.75}{2\varepsilon}I,$$

which for small ε is clearly a large negative definite matrix. Therefore, if a poor

approximation X_0 is to be expected (for example, the Riccati equation is ill-conditioned), then it is advisable, in order to speed-up the convergence, to reinitialize the iteration after the first step with $\bar{X}_0 = \max(0, X_1)$.

(iii) Usually, no more than 9–11 iterations of Newton's method are necessary to compute the limiting accuracy solution of (1) when we choose $X_0 = 0$. Thus, the solution of the AMRE (1) requires usually no more than $120n^3$ *flops* (floating point operations), where we assumed 10 iterations and $12n^3$ *flops* were considered to solve the Lyapunov equation. When X_0 is an approximate solution computed by the Schur vectors method, then, at most 3–4 iterations are additionally necessary to achieve the highest attainable accuracy. The whole procedure to compute $\bar{X}$ requires about $120n^3$ *flops* instead of about $75n^3$ *flops* required by using only the Schur method, This increase with about 50% of the computational effort is sometimes necessary in order to attain the necessary accuracy of the solution. Notice that the scaling of the Hamiltonian matrix (6), as suggested in Kenney et al. (1989), for enhancing the numerical accuracy of the Schur method, requires the solution of the Riccati equation twice and is always more expensive and less accurate than the iterative refinement technique.

3. Computation of the Cholesky Factor

In several applications, as, for instance, in the square-root balanced stochastic truncation method (Varga and Fasol, 1993), we need instead of $\bar{X}$, its Cholesky factor S. Even if a sufficiently accurate $\bar{X}$ is available, the direct computation of the Cholesky factorization $\bar{X} = S^T S$ is generally not recommendable from the numerical point of view because the computed factor has usually lower accuracy than the given $\bar{X}$. In the worst case, we can expect a loss of accuracy equivalent to half of the significant digits of the solution. When $\bar{X}$ is nearly or exactly singular, the rank information can even be lost due to roundoff errors when computing the Cholesky factor. For instance, if $\bar{X}$ has a singular value of order ε, then S will have a singular value of order $\varepsilon^{1/2}$, which could be a non-negligible quantity even when ε is negligible. What makes the problem even more difficult in the mentioned model reduction application is that this rank information must be recovered later, after performing some operations with the computed factor. The purpose of this section is to discuss several other possible methods to compute S.

Method 1. We discuss first the computation of the Cholesky factor S by using the Schur method (Laub, 1979). Accuracy losses arise here due to the instability of the inversion step when computing $\bar{X} = Z_2 Z_1^{-1}$. Provided the Riccati equation is well scaled, that is $\|\bar{X}\| \cong 1$ (Kenney et al., 1989), we can avoid the explicit inversion of Z_1 by using the following approach suggested by Singer and Hammarling (1983). We compute first the singular value decomposition,

$$Z_1 = U\Sigma V^T,$$

where U and V are orthogonal matrices and $\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_n)$. Because Z_1 is a part of a matrix with orthogonal columns and is invertible, we have

$$1 \geq \sigma_1 \geq \cdots \geq \sigma_n > 0.$$

The solution $\bar{X}$ can be expressed as

$$\bar{X} = U(\Sigma^{-2} - I)^{\frac{1}{2}} U^T$$

and, thus, the Cholesky factor can be computed as

$$S = (\Sigma^{-2} - I)^{\frac{1}{2}} U^T.$$

Notice that the computation of V is not necessary in order to determine S. It is easy to see that $\|\bar{X}\|_2 = \sqrt{\sigma_n^{-2} - 1}$ and, thus, we have an immediate indication of the expected accuracy of the computed solution. Even the information on the rank of $\bar{X}$ can be computed from the computed singular values. The rank deficiency of $\bar{X}$ (and thus of S, too) is given by the number of singular values which are equal to 1. In order not to lose this information because of roundoff errors, we can set the respective singular values *exactly* to 1 before computing S. However, when using this approach, always we have to take into account the possibility of losing half of the accuracy due to the square root operations in evaluating S.

Method 2. Let $Q = V^T V$ and $R = U^T U$ be the Cholesky factorizations of the positive semidefinite matrices Q and R. Although generally not unique, such factorizations always exist with V and U upper triangular and with non-negative diagonal elements. A square-root version of Newton's method can be derived by rewriting (7) as

$$(A + QX_{j-1})^T(X_j - X_{j-1}) + (X_j - X_{j-1})(A + QX_{j-1}) + E(X_{j-1}) = 0 \qquad (9)$$

and taking into account that for $j \geq 2$, the term $E(X_{j-1})$ can be expressed as

$$E(X_{j-1}) = (X_{j-1} - X_{j-2})Q(X_{j-1} - X_{j-2}).$$

By Theorem 1, $X_j - X_{j-1} \geq 0$ for $j \geq 2$, and, thus, there exists the Cholesky factorization,

$$X_j - X_{j-1} = H_j^T H_j.$$

The Cholesky factor S_j of X_j can be recursively computed from the equality,

$$S_j^T S_j = S_{j-1}^T S_{j-1} + H_j^T H_j,$$

by using a stable algorithm for updating the Cholesky factorization (Gill et al., 1974). The following procedure represents a square-root version of Newton's method to solve the AMRE (1):

1) Set $S_0 = 0$ and compute H_0 by solving

$$A^T(H_0^T H) + (H_0^T H)A + U^T U = 0.$$

2) For $j = 1, 2, \cdots$ (until convergence),
   2.1) Compute H_j by solving

$$(A + QS_{j-1}^T S_{j-1})^T (H_j^T H_j) + (H_j^T H_j)(A + QS_{j-1}^T S_{j-1}) + (VH_{j-1}^T H_{j-1})^T (VH_{j-1}^T H_{j-1}) = 0 .$$

2.2) Compute S_j from

$$S_j^T S_j = S_{j-1}^T S_{j-1} + H_j^T H_j .$$

This algorithm is primarily intended as an alternative more accurate approach for the iterative solution of the AMRE (1). It is slightly more expensive than the original version of Newton's method discussed in the previous section. Both methods are particularly well suited when, due to the ill-conditioning of the AMRE (1), the initial approximation (computed, for instance, with the Schur method) is not stabilizing, or the Schur method fails (for instance, the Hamiltonian (6) loses the dichotomy property due to roundoff). An additional advantage of the square-root approach is that it provides directly the Cholesky factor of the stabilizing solution. Note that the proposed square-root method to solve the AMRE (1) is of the *incremental* type and, thus, of a different natnre than the square-root iteration proposed by Hammarling (1982 a) for solving the standard Riccati equation.

The use of this algorithm for iterative refinement raises some problems. Generally, we cannot assume that an arbitrary positive semidefinite stabilizing approximation $\bar{X}_0$ satisfies $\bar{X}_0 \leq \bar{X}$, and, thus, it is suitable to initialize the square-root iteration. Therefore, we have to compute first $X_{-1} \leq \bar{X}$ and $X_0 \leq \bar{X}$ by performing two standard Newton iterations (7) initialized with $\bar{X}_0$. By Lemma 1, $X_{-1} \leq X_0$, and we can compute the Cholesky factors H_0 and S_0 of $X_0 - X_{-1}$ and of X_0, respectively. These computations can lead, however, to accuracy losses, especially if $\bar{X}$ is (nearly) singular, an aspect already mentioned before.

Method 3. Generally, we cannot assume that $R - X_{j-1}QX_{j-1} \geq 0$ in (7), and, thus, this equation is not suitable for directly computing the Cholesky factor S_j. The following approach, exploiting the linearity of (7), can be used to determine S_j if $X_j > 0$: compute first, by using the method of (Hammarling, 1982 b), S_{j1} and S_{j2}, the solutions of the equations,

$$(A + QX_{j-1})^T (S_{j1}^T S_{j1}) + (S_{j1}^T S_{j1})(A + QX_{j-1}) + U^T U = 0 , \quad (10)$$

$$(A + QX_{j-1})^T (S_{j2}^T S_{j2}) + (S_{j2}^T S_{j2})(A + QX_{j-1}) + (VX_{j-1})^T (VX_{j-1}) = 0 , \quad (11)$$

and, then, determine S_j such that

$$X_j \triangleq S_j^T S_j = S_{j1}^T S_{j1} - S_{j2}^T S_{j2} . \quad (12)$$

S_j can be reliably determined from S_{j1} and S_{j2}, without forming the products $S_{j1}^T S_{j1}$ and $S_{j2}^T S_{j2}$, by using an algorithm for downdating the Cholesky factorization. We performed experiments with the downdating algorithm of Bojanczyk and Steinhardt (1991), but because in our case, generally $X_j \geq 0$, we encountered difficulties in approaching the limiting accuracy solution in those cases when $\bar{X}$ was singular. These difficulties can be explained with the ill-conditioning of the downdating problem for the positive semi-definite case. However,

Eqs. (10)–(12), excluding the computation of S_j, represent an alternative way (although more expensive) to perform Newton's iterative refinement. This approach led sometimes to more accurate solutions than those computed with the iteration based on (7).

Method 4. If an accurate solution $\bar{X}$ is already available, a straightforward way to compute reliably its Cholesky factor S is by solving directly for S either the equation

$$A^T (S^T S) + (S^T S) A + \begin{bmatrix} V\bar{X} \\ U \end{bmatrix}^T \begin{bmatrix} V\bar{X} \\ U \end{bmatrix} = 0$$

or

$$\left(A + \frac{1}{2} Q\bar{X}\right)^T (S^T S) + (S^T S)\left(A + \frac{1}{2} Q\bar{X}\right) + U^T U = 0 . \quad (13)$$

These equations are obtained by differently grouping the terms in the AMRE (1). The stability of $A + Q\bar{X}/2$ is guaranteed by Lyapunov's Theorem.

Numerical experiments indicate that Eq. (13) should be generally preferred. It can be viewed as a step of the following iteration, for $i = 1, 2, \cdots$,

$$\left(A + \frac{1}{2} QX_{j-1}\right)^T X_j + X_j \left(A + \frac{1}{2} QX_{j-1}\right) + R = 0 . \quad (14)$$

Experimentally, it was observed that the iterations based on (14) converge to the stabilizing solution $\bar{X}$ provided the initializing approximation X_0 is not too far from $\bar{X}$. In most cases, the convergence rate was, however, very low, and, therefore, Eq. (14) is not appropriate for iterative refinement. However, its contracting property (a conjecture based on experiments) is, in our case, very useful, allowing us to perform safely a "last step" after the convergence of Newton's method in order to compute the Cholesky factor of the solution.

Summarizing the above considerations, the following two main alternatives are possible to compute accurate Cholesky factors of the solution of the Riccati equation (1). If an accurate solution $\bar{X}$ has been already determined by using, for instance, the Schur method with or without iterative refinement, then Method 4 based on Eq. (13) should be generally preferred because of its already mentioned accuracy refining property. Appropriate quantities to assess the accuracy of the computed solution are the corresponding residual, the norm of $\bar{X}$ and/or the condition number of the Riccati equation. The second alternative is to use Method 2 to solve the Riccati equation and is particularly well suited if the Riccati equation is highly ill-conditioned. Notice that in many cases the iterative solution by using Method 2 is computationally less expensive than by using the Schur method with or even without iterative refinement.

4. Numerical Example

The example given in this section is intended to illustrate the effectiveness of the proposed iterative refinement technique in computing the limiting accuracy solution of the AMRE (4) arising in the balanced stochastic truncation model

reduction metbod (Varga and Fasol, 1993). The computations have been performed ou an IBM AT computer, in double precision, using tbe following 10th order non-minimal (uncontrollable and unobservable) system

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$$

$$= \left[\begin{array}{cccccccccc|cccc} -6 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & -8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -10 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -13 & -3 & 9 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -14 & -9 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 10^{-3} & 10^{-3} & 0 & 0 \\ \hline 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 5\cdot 10^{5} & 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -6 & 1 & -2 & 5\cdot 10^{5} & 0 & 0 & 0 & \alpha \end{array}\right].$$

The parameter α appearing in the expression of D is used to modify the conditioning of the Riccati equation (4) in order to illustrate the effects of the iterative refinement. In Table 1, we included, for values of α ranging from 10^{-4} to 10^{2}, the following results: $\|E(0)\|_1$, $\|E(\tilde{X})\|_1$ and $\|E(\bar{X})\|_1$, the norms of residuals of the AMRE (4) for $X = 0$, $X = \tilde{X}$ (the solution computed with the Schur vectors method) and $X = \bar{X}$ (the high accuracy solution computed by using iterative refinement), respectively; $N(0)$ and $N(\tilde{X})$, the numbers of iterations necessary for the convergence of Newton's method initialized with $X_0 = 0$ and $X_0 = \tilde{X}$, respectively; and $\mathrm{cond}(\bar{X})$, an estimation of the condition number of the Riccati equation (Kenney and Hewer, 1990). The values of α are given decreasingly; the smaller is α, the more ill-conditioned is the AMRE. For all values

Table 1.

α	10^{2}	10	1	10^{-1}
$\|E(0)\|_1$	$5.0\cdot 10^{7}$	$5.0\cdot 10^{9}$	$5.0\cdot 10^{11}$	$5.0\cdot 10^{13}$
$\|E(\tilde{X})\|_1$	$3.2\cdot 10^{-8}$	$4.5\cdot 10^{-7}$	$4.1\cdot 10^{-5}$	$3.2\cdot 10^{-1}$
$\|E(\bar{X})\|_1$	$3.2\cdot 10^{-8}$	$4.5\cdot 10^{-7}$	$3.8\cdot 10^{-5}$	$9.7\cdot 10^{-3}$
$N(0)$	6	9	12	16
$N(\tilde{X})$	0	0	1	1
$\mathrm{cond}(\bar{X})$	$2.7\cdot 10^{5}$	$2.0\cdot 10^{7}$	$2.0\cdot 10^{9}$	$2.0\cdot 10^{11}$
α	10^{-2}	10^{-3}	10^{-4}	
$\|E(0)\|_1$	$5.0\cdot 10^{15}$	$5.0\cdot 10^{17}$	$5.0\cdot 10^{19}$	
$\|E(\tilde{X})\|_1$	$7.5\cdot 10^{2}$	$4.0\cdot 10^{4}$	$5.1\cdot 10^{6}$	
$\|E(\bar{X})\|_1$	$5.0\cdot 10^{-1}$	$3.1\cdot 10^{1}$	$7.0\cdot 10^{3}$	
$N(0)$	20	25	28	
$N(\tilde{X})$	2	3	4	
$\mathrm{cond}(\bar{X})$	$3.3\cdot 10^{13}$	$3.7\cdot 10^{15}$	$3.7\cdot 10^{17}$	

of α, $\|\bar{X}\|_1 \cong 10^{6}$, thus the scaling of the AMRE (4) has practically the same effect on the accnracy of the Schur method for different values of α.

It is apparent from this table that the effectiveness of the iterative refinement increases as the problem becomes more ill-conditioned. With the help of iterative refinement, in all cases it was possible to determine correctly the order (six) of the minimal realization of the system (A, B, C, D) from the product WS^T, where W and S are the Cholesky factors of P and X, respectively. Note that this was not possible for the last two values of α by only using the solution compnted with the Schnr method.

5. Conclusions

A techniqne for computing high accuracy solutions of a class of Riccati equations arising in spectral factorization problems has been proposed. The proposed iterative refinement is based on Newton's method. A proof of convergence with final quadratic rate is given. Square-root versions of the basic algorithm, iterating directly on Cholesky factors, have been discussed. The main applications of these techniques are in insuring the effectiveness of the square-root accuracy enhancing approaches in certain model reduction methods.

References

Anderson, B.D.O. (1967). A system theory criterion for positive real matrices. *SIAM J. Control*, **5**, 2, 171–182.

Arnold III, W.F. and A.J. Laub (1984). Generalized eigenproblem algorithms and software for algebraic Riccati equations. *Proc. of IEEE*, **72**, 12, 1746–1754.

Barnett, S. (1971). *Matrices in Control Theory*. Van Nostrand Reinbold, London.

Bojanczyk, A.W. and A.O. Steinhardt (1991). Stability analysis of a Householder-based algorithm for downdating the Cholesky factorization. *SIAM J. Sci. Stat. Comput.*, **12**, 6, 1255–1265.

Gill, P.E., G.H. Golub, W. Murray and M.A. Saunders (1974). Methods for modifying matrix factorizations. *Moth. Comput.*, **28**, 126, 505–535.

Hammarling, S.J. (1982 a). Newton's method for solving the algebraic Riccati equation. NPL Report DITC 12/82, National Physical Laboratory, Teddington, Middlesex, U.K.

Hammarling, S.J. (1982 b). Numerical solution of the stable, non-negative definite Lyapunov equation. *IMA J. Numer. Anal.*, **2**, 3, 303–323.

Kano, H. and T. Nishimura (1993). Nonnegative-definite solutions of algebraic matrix equations with nonnegative-definite quadratic and constant terms. Research Report TDU-IS-39, Dept. Inf. Scie., Tokyo Denki University.

Kenney, C. and G. Hewer (1990). The sensitivity of the algebraic and differential Riccati eqnations. *SIAM J. Control and Optimization*, **28**, 1, 50–69.

Kenney, C., A.J. Laub and M. Wette (1989). A stability enhancing scaling procedure for Schur-Riccati solvers. *Systems and Control Letters*, **12**, 3, 241–250.

Kenney, C., A.J. Laub and M. Wette (1990). Error bounds for Newton refinement of solutions to algebraic Riccati equations. *Math. Control. Signals, Systems*, **3**, 3, 211–224.

Kleinman, D.L. (1968). On an iterative technique for Riccati equation computations. *IEEE Trans. Automatic Control*, **AC-13**, 1, 114–115.

Laub, A. (1979). A Schur method for solving the algebraic Riccati equation. *IEEE Trans. Automatic Control*, **AC-24**, 6, 913–921.

Mehrmann, V.L. (1991). The autonomous linear quadratic control problem. Lecture Notes in Control and Information Sciences, **163**, Springer-Verlag, Berlin.

Petkov, P.H., N.D. Christov and M.M. Konstantinov (1987). On numerical properties of the Schur approach for solving the matrix Riccati equation. *Systems and Control Letters*, **9**, 3, 197–201.

Safonov, M.G. and R.Y. Chiang (1988). Model reduction for robust control: A Schur relative error method. *Int. J. Adapt. Contr. and Sign. Proc.*, **2**, 259–272.

Sandell, N.R. (1974). On Newton's method for Riccati equation solution. *IEEE Trans. Automatic Control*, **AC-19**, 3, 254–255.

Singer, M.A. and S.J. Hammarling (1983). The algebraic Riccati equation: A summary review of some available results. NPL Report DITC 23/83, National Physical Laboratory, Teddington, Middlesex, U.K.

Varga, A. and K.H. Fasol (1993). A new square-root balancing-free stochastic truncation model reduction algorithm. *Preprints of 12th IFAC World Congress*, Sydney, Australia. **7**, 153–156.

Andras Varga received the diploma in control engineering in 1974 and the Ph.D. degree in electrical engineering in 1981, both from the Polytechnic Institute of Bucharest, Romania.

Between 1974 and 1993, he held various research positions at the Institute for Informatics in Bucharest and at the Ruhr-University of Bochnm, Germany. From 1990 to 1992, he worked at the Ruhr-University of Bochum as visiting research fellow in the framework of a fellowship award of the Alexander von Humboldt Foundation. In 1994, he was for a short period a visiting scholar at Kyoto University. Dr. Varga is currently with the German Aerospace Research Establishment (DLR) in Oberpfaffenhofen.

His main interests lie in the areas of numerical methods for systems analysis and design, linear control systems theory, and the developing of robust numerical software for computer-aided design of control systems. Dr. Varga has published more than 70 papers, and he is a coauthor of two books, *Computer Aided Optimization Practice* (1986) and *System Theory: Robust Synthesis and Numerical Methods* (1995) (both in Romanian).