SIAM J. MATRIX ANAL. APPL.
Vol. 20, No. 1, pp. 228–243

NONSYMMETRIC ALGEBRAIC RICCATI EQUATIONS AND HAMILTONIAN-LIKE MATRICES[*]

JONQ JUANG[†] AND WEN-WEI LIN[‡]

Abstract. We consider a nonsymmetric algebraic matrix Riccati equation arising from transport theory. The nonnegative solutions of the equation can be explicitly constructed via the inversion formula of a Cauchy matrix. An error analysis and numerical results are given. We also show a comparison theorem of the nonnegative solutions.

Key words. Hamiltonian, algebraic Riccati equation, M-matrices, nonnegative matrices, transport theory

AMS subject classifications. 15A24, 82C70

PII. S0895479897318253

1. Introduction.

In transport theory, a variation of the usual one-group neutron transport equation [2, 8, 10] is formulated as

$$\left\{(\mu+\alpha)\frac{\partial}{\partial x}+1\right\}\varphi(x,\mu)=\frac{c}{2}\int_{-1}^{1}\varphi(x,\omega)d\omega, \tag{1a}$$

$$\varphi(0,\mu)=f(\mu),\ \mu>-\alpha,|\mu|\le 1, \tag{1b}$$

$$\lim_{x\to\infty}\varphi(x,\mu)=0. \tag{1c}$$

Here φ is the neutron flux, α $(0\le\alpha<1)$ is an angular shift, and c is the average of the total number of particles emerging from a collision, which is assumed to be conservation; i.e., $0\le c\le 1$.

The scattering function (see, e.g., [15]) for particle transport (or radiative transfer) in the half-space can be derived from (1) via invariant embedding [2]. Such a scattering function satisfies the following integrodifferential equation (see the appendix in [15]):

$$\left(\frac{1}{\mu+\alpha}+\frac{1}{\nu-\alpha}\right)X(\mu,\nu)=c\left[1+\frac{1}{2}\int_{-\alpha}^{1}\frac{X(\omega,\nu)}{\omega+\alpha}d\omega\right]\left[1+\frac{1}{2}\int_{\alpha}^{1}\frac{X(\mu,\omega)}{\omega-\alpha}d\omega\right], \tag{2}$$

with $(\mu,\nu)\in[-\alpha,1]\times[\alpha,1]$. Here the function $X:[-\alpha,1]\times[\alpha,1]\to\mathbf{R}$ is called a scattering function. For the case in which $c=0$ or $\alpha=1$, (2) has a trivial solution. When $\alpha=0$, the existence of nonnegative solutions of (2) has been studied by many authors (see, e.g., [15] and the works cited therein). In fact, for this case, the two integrals in (2) are the usual Chandrasekhar H-function [5, 15].

Discretization of the integrodifferential equation of (2) yields an algebraic matrix Riccati equation. To see this, let $\{\omega_k\}_{k=1}^{n}$ and $\{c_k\}_{k=1}^{n}$ denote the sets of the Gauss–Legendre nodes and weights, respectively, on $[0,1]$ with

$$0<\omega_n<\cdots<\omega_2<\omega_1<1 \tag{3a}$$

[*] Received by the editors March 7, 1997; accepted for publication (in revised form) by V. Mehrmann January 28, 1998; published electronically September 23, 1998. This work was supported in part by the National Science Council of the R.O.C. on Taiwan.

http://www.siam.org/journals/simax/20-1/31825.html

[†] Department of Applied Mathematics, National Chiao Tung University, Hsinchu, Taiwan 31015, Republic of China (jjuang@math.nctu.edu.tw).

[‡] Institute of Applied Mathematics, National Tsing Hwa University, Hsinchu, Taiwan 30043, Republic of China (wwlin@am.nthu.edu.tw).

and

$$\sum_{k=1}^{n} c_k = 1,\ c_k > 0,\ k = 1, 2, \ldots, n. \tag{3b}$$

Transforming the Gauss–Legendre nodes and weights on $[0, 1]$ to the intervals $[-\alpha, 1]$ and $[\alpha, 1]$, respectively, we have the following relationships:

$$\omega_k^- = \{(1+\alpha)\omega_k - \alpha\} \in [-\alpha, 1], \quad c_k^- = c_k(1+\alpha), \tag{4a}$$

$$\omega_k^+ = \{(1-\alpha)\omega_k + \alpha\} \in [\alpha, 1], \quad c_k^+ = c_k(1-\alpha) \tag{4b}$$

for $k = 1, \ldots, n$. Let $X_{ij} = X(\omega_i^-, \omega_j^+),\ i, j = 1, \ldots, n$. Replacing μ, ν with ω_i^- and ω_j^+, respectively, in (2), the integrals in (2) can be approximated by

$$\int_{-\alpha}^{1} \frac{X(\omega, \omega_j^+)}{\omega + \alpha} d\omega \sim \sum_{k=1}^{n} \frac{c_k^- X_{kj}}{\omega_k^- + \alpha}$$

and

$$\int_{\alpha}^{1} \frac{X(\omega_i^-, \omega)}{\omega - \alpha} d\omega \sim \sum_{k=1}^{n} \frac{c_k^+ X_{ik}}{\omega_k^+ - \alpha}.$$

Consequently, the descretized version of (2) becomes

$$\begin{aligned} &\frac{1}{c(\omega_i^- + \alpha)} X_{ij} + \frac{1}{c(\omega_j^+ - \alpha)} X_{ij} \\ &= 1 + \frac{1}{2}\sum_{k=1}^{n} \frac{c_k^- X_{kj}}{\omega_k^- + \alpha} + \frac{1}{2}\sum_{k=1}^{n} \frac{c_k^+ X_{ik}}{\omega_k^+ - \alpha} + \frac{1}{4}\sum_{k=1}^{n}\sum_{l=1}^{n} \frac{X_{ik} c_k^+ c_l^- X_{lj}}{(\omega_k^+ - \alpha)(\omega_l^- + \alpha)}. \end{aligned} \tag{5}$$

Substituting (4) into (5) and writing the resulting equation in matrix form, we get an $n \times n$ nonsymmetric algebraic matrix Riccati equation in X:

$$B - AX - XD + XCX = 0, \tag{6}$$

where A, B, C, and D have the following structures:

$$A = \operatorname{diag}[\delta_1, \delta_2, \ldots, \delta_n] - eq^T, \tag{7a}$$

$$B = ee^T, \tag{7b}$$

$$C = qq^T, \tag{7c}$$

and

$$D = \operatorname{diag}[d_1, d_2, \ldots, d_n] - qe^T, \tag{7d}$$

where

$$\delta_i = \frac{1}{cw_i(1+\alpha)}, \qquad d_i = \frac{1}{cw_i(1-\alpha)}, \tag{8a}$$

and

$$e = [1, 1, \ldots, 1]^T, \qquad q = [q_1, q_2, \ldots, q_n]^T \text{ with } q_i = \frac{c_i}{2w_i}. \tag{8b}$$

In studying (6), we may assume that all the data are real and that $0 < c \leq 1$, $0 \leq \alpha < 1$, and (3) are satisfied. Consequently, we may assume that

$$0 < \delta_1 < \delta_2 < \cdots < \delta_n \tag{9a}$$

and

$$0 < d_1 < d_2 < \cdots < d_n. \tag{9b}$$

In addition, we may assume that

$$d_i = \delta_i \text{ for } \alpha = 0, \; d_i > \delta_i \text{ for } \alpha \neq 0, \quad i = 1, 2, \ldots, n. \tag{9c}$$

Recently, the existence of nonnegative solutions (in the componentwise sense) of (6) was demonstrated via the degree theory by Juang [13]. Some iterative procedures [14] have been developed for finding the nonnegative solutions of (6). However, for the case in which $c \approx 1$ and $\alpha \approx 0$, the convergence rates of these procedures are very slow, which is unsatisfactory.

Now, let H denote a 2×2 block matrix of the form

$$H := \begin{bmatrix} D & -C \\ B & -A \end{bmatrix}, \tag{10}$$

where A, B, C, and D are as defined in (7). We call this matrix H a Hamiltonian-like matrix of (6). In this paper, we develop a different approach to constructing the nonnegative solutions of (6) based on computing the invariant subspaces of H corresponding to some specified eigenvalues. The inversion formula of a Cauchy matrix is also used to explicitly construct such solutions. Our approach gives a complete representation and bifurcation diagram, with respect to parameters c and α, for nonnegative solutions of (6). Furthermore, it provides a numerical algorithm for computing the nonnegative solutions of (6) and avoids the deficiencies inherent in the iterative procedures mentioned above.

Symmetric algebraic Riccati equations arising from linear-quadratic control problems are often solved by computing the "stable" invariant subspace of the corresponding Hamiltonian matrix $\tilde{H}$ (see, e.g., [17, p. 55]). Such equations have been treated at length in the literature (see, e.g., [17] and the works cited therein). Here $\tilde{H}$ is of the form $\tilde{H} = \left[\begin{smallmatrix} \tilde{A}^T & -\tilde{C} \\ \tilde{B} & -\tilde{A} \end{smallmatrix}\right]$, where $\tilde{B}, \tilde{C}$ are symmetric and $\tilde{A}$ is arbitrary. On the other hand, nonsymmetric Riccati equations (see, e.g., [7, 18]) are less well understood than their symmetric counterparts. Note that H, given in (10), is a Hamiltonian matrix only when $c = 1$ and $\alpha = 0$, which is why we call it a Hamiltonian-like matrix. Moreover, we are seeking a nonnegative solution of (6), as opposed to the positive semidefinite solutions found in linear-quadratic control problems or the nonsingular solutions found in polynomial factorizations.

This paper is organized as follows. In section 2, we analyze the eigenvalue distribution of H and characterize the components of the associated eigenvectors. In section 3, a complete representation and bifurcation diagram of the nonnegative solutions of (6) are established. In particular, we show that (6) has a unique nonnegative solution when $c = 1$ and $\alpha = 0$ and two nonnegative solutions otherwise. An error analysis and some numerical experiments for the computation of the nonnegative solutions are given in section 4. In section 5, some comparison results are derived. Specifically, we are able to show that the minimal solution of (6) is increasing in c and decreasing in α. Our concluding section primarily contains some thoughts regarding possible future research related to the results presented here. For completeness and ease of reference, we conclude this introductory section by recording some well-known results.

In what follows, we shall give the definition of the M-matrix and its properties (see, e.g., [19, p. 54]).

DEFINITION 1.1. *A real $n \times n$ matrix A is an M-matrix if there exists a nonnegative matrix B with a maximal eigenvalue r such that $A = cI_n - B$, where $c \geq r$.*

THEOREM 1.2. *Let matrix A be an $n\times n$ nonsingular real matrix with nonpositive off-diagonal elements. Then the following are equivalent:*
(i) *A is an M-matrix.* (ii) *Every real eigenvalue in A is nonnegative.* (iii) *A^{-1} is nonnegative.*

THEOREM 1.3. *Let two $n \times n$ matrices A_i, $i = 1, 2$, be decomposed into $A_i = D_i - B_i$, respectively, where D_i, $i = 1, 2$, are diagonal parts of A_i, $i = 1, 2$. Suppose A_1 is an invertible M-matrix, $D_1 \leq D_2$, and $B_1 \geq B_2 \geq 0$. A_2 is then an M-matrix and $A_2^{-1} \leq A_1^{-1}$.*

2. Properties of the Hamiltonian-like matrix H.

In this section, we analyze the eigenvalue distribution of H given in (10) and characterize the components and properties of the associated eigenvectors.

LEMMA 2.1. *The matrix H, as defined in* (10)*, has only real eigenvalues $\{-\mu_n, \ldots, -\mu_1, \lambda_1, \ldots, \lambda_n\}$, which are arranged in an ascending order and satisfy the following inequalities:*

$$-\delta_n < -\mu_n < -\delta_{n-1} < \ldots < -\delta_2 < -\mu_2 < -\delta_1 < -\mu_1 \leq 0, \tag{11a}$$

$$0 \leq \lambda_1 < d_1 < \lambda_2 < d_2 < \ldots < \lambda_n < d_n. \tag{11b}$$

Moreover, the following hold: (i) *$\mu_1 = 0$ only if $c = 1$.* (ii) *$\lambda_1 = 0$ only if $c = 1$ and $\alpha = 0$.* (iii) *$\mu_i = \lambda_i, i = 1, 2, \ldots, n$, for $\alpha = 0$.*

Proof. Let

$$D_1 = \mathrm{diag}[d_1, d_2, \ldots, d_n] \quad \text{and} \quad \Delta_1 = \mathrm{diag}[\delta_1, \delta_2, \ldots, \delta_n]. \tag{12}$$

We rewrite $H - \lambda I$ as

$$H - \lambda I = \begin{bmatrix} D_1 & 0 \\ 0 & -\Delta_1 \end{bmatrix} - \lambda I - \begin{bmatrix} q \\ -e \end{bmatrix} \begin{bmatrix} e^T, & q^T \end{bmatrix}. \tag{13}$$

The secular equation (see, e.g., [3, 6, 11]) of $H - \lambda I$ is given by

$$\begin{aligned} f(\lambda) &= 1 - (e^T, q^T) \begin{bmatrix} (D_1 - \lambda I)^{-1} & 0 \\ 0 & -(\Delta_1 + \lambda I)^{-1} \end{bmatrix} \begin{bmatrix} q \\ -e \end{bmatrix} \\ &= 1 - \sum_{i=1}^{n} \frac{q_i}{d_i - \lambda} - \sum_{i=1}^{n} \frac{q_i}{\delta_i + \lambda}. \end{aligned} \tag{14}$$

Since $\lambda = d_i$ and $-\delta_i$, $i = 1, \ldots, n$, are not eigenvalues of H, finding eigenvalues of H is equivalent to locating the roots of $f(\lambda)$. Using (14), we immediately have the following asymptotic properties:

$$\lim_{\lambda \to \pm\infty} f(\lambda) = 1, \quad \lim_{\lambda \to d_i^{\pm}} f(\lambda) = \pm\infty, \quad \lim_{\lambda \to -\delta_i^{\pm}} f(\lambda) = \mp\infty, \quad i = 1, 2, \ldots, n.$$

The intermediate value theorem indicates that $f(\lambda)$ must have at least one root in each of the intervals (d_i, d_{i+1}) and $(-\delta_{i+1}, -\delta_i)$, where $i = 1, 2, \ldots, n-1$. Thus, there are at least $2n-2$ roots in those intervals. We next examine the number of possible roots of $f(\lambda)$ in the interval $(-\delta_1, d_1)$. To this end, we evaluate $f(0)$ and its rate of change, $f^{'}(0)$.

From (14), (8), and (3b), it can be determined that

$$\begin{aligned} f(0) &= 1 - \sum_{i=1}^{n} \frac{c_i}{2\omega_i} c\omega_i(1-\alpha) - \sum_{i=1}^{n} \frac{c_i}{2\omega_i} c\omega_i(1+\alpha) \\ &= 1 - c \begin{cases} > 0 & \text{for } 0 \le c < 1, \\ = 0 & \text{for } c = 1 \end{cases} \end{aligned} \tag{15a}$$

and

$$\begin{aligned} f'(0) &= -\frac{1}{2}\sum_{i=1}^{n} c_i c^2 \omega_i (1-\alpha)^2 + \frac{1}{2}\sum_{i=1}^{n} c_i c^2 \omega_i (1+\alpha)^2 \\ &= 2\alpha c^2 \sum_{i=1}^{n} c_i \omega_i \begin{cases} > 0 & \text{for } \alpha > 0, \\ = 0 & \text{for } \alpha = 0. \end{cases} \end{aligned} \tag{15b}$$

Since $\lim_{\lambda \to -\delta_1^+} f(\lambda) = \lim_{\lambda \to d_1^-} f(\lambda) = -\infty$, we conclude, via (15), that, for $0 < c < 1$, $f(\lambda)$ has two other roots. Specifically, one is in $(-\delta_1, 0)$ and the other is in $(0, d_1)$. It follows from (15) that, for $c = 1$ and $0 < \alpha < 1$, one root of $f(\lambda)$ is zero and the other root is in $(0, d_1)$, and for $c = 1$ and $\alpha = 0$, $f(\lambda)$ has a zero root of multiplicity 2. We thus complete the proof of the lemma. □

The following results can be easily obtained by studying the secular equations of A and D. The proof of the lemma is thus omitted.

LEMMA 2.2. *The eigenvalues of A and D are real and positive.*

We now turn our attention to the eigenvectors corresponding to the eigenvalues λ_k and μ_k of H for $k = 1, \ldots, n$.

LEMMA 2.3. *Let $[x_{1,k}, \ldots, x_{2n,k}]^T$ and $[z_{1,k}, \ldots, z_{2n,k}]^T$ be the eigenvectors of H corresponding to λ_k and $-\mu_k$, respectively, for $k = 1, \ldots, n$. Then it holds that*

$$x_{i,k} = \frac{q_i(\delta_n + \lambda_k)}{d_i - \lambda_k} \quad \textit{and} \quad x_{n+i,k} = \frac{\delta_n + \lambda_k}{\delta_i + \lambda_k}, i = 1, \ldots, n, \tag{16}$$

$$z_{i,k} = \frac{q_i(\delta_n - \mu_k)}{d_i + \mu_k} \quad \textit{and} \quad z_{n+i,k} = \frac{\delta_n - \mu_k}{\delta_i - \mu_k}, i = 1, \ldots, n, \tag{17}$$

for $k = 1, \ldots, n$.

Proof. Let $x_k = [x_{1,k}, x_{2,k}, \ldots, x_{2n,k}]^T$ be the eigenvector corresponding to λ_k; i.e., $(H - \lambda_k I)x_k = 0$. Writing $H - \lambda_k I$ in the form of (13) and using the last component -1 in $\left[\begin{smallmatrix} q \\ -e \end{smallmatrix}\right]$ as a pivotal element to eliminate the other elements of $\left[\begin{smallmatrix} q \\ -e \end{smallmatrix}\right]$, we get

(18)

$$\begin{bmatrix} \tilde{d}_{1,k} & 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 & -q_1\tilde{\delta}_{n,k} \\ 0 & \ddots & \ddots & & & & & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & & & & \vdots & \vdots \\ \vdots & & \ddots & \tilde{d}_{n,k} & 0 & & & \vdots & -q_n\tilde{\delta}_{n,k} \\ \vdots & & & 0 & -\tilde{\delta}_{1,k} & \ddots & & \vdots & \tilde{\delta}_{n,k} \\ \vdots & & & & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & & & & & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 & -\tilde{\delta}_{n-1,k} & \tilde{\delta}_{n,k} \\ 1 & \cdots & \cdots & 1 & q_1 & \cdots & \cdots & q_{n-1} & -\tilde{\delta}_{n,k}+q_n \end{bmatrix} \begin{bmatrix} x_{1,k} \\ \vdots \\ \vdots \\ x_{n,k} \\ x_{n+1,k} \\ \vdots \\ \vdots \\ x_{2n-1,k} \\ x_{2n,k} \end{bmatrix} = 0.$$

Here, $\tilde{d}_{i,k} = d_i - \lambda_k$ and $\tilde{\delta}_{i,k} = \delta_i + \lambda_k$. Letting $x_{2n,k} = 1$, we see, via (18), that

$$x_{i,k} = \frac{q_i\tilde{\delta}_{n,k}}{\tilde{d}_{i,k}} = \frac{q_i(\delta_n + \lambda_k)}{d_i - \lambda_k} \text{ and } x_{n+i,k} = \frac{\tilde{\delta}_n}{\tilde{\delta}_i} = \frac{\delta_n + \lambda_k}{\delta_i + \lambda_k} \tag{19}$$

for $i = 1, \ldots, n$. The eigenvectors corresponding to $-\mu_k$ can be obtained in a similar way. □

Proposition 2.4. *For $c = 1$ and $\alpha = 0, \lambda_1 = \mu_1 = 0$ is a zero eigenvalue of H that has an algebraic multiplicity of two and a geometric multiplicity of one.*

Proof. From Lemma 2.1 we see that the eigenvalues $\lambda_1 = \mu_1 = 0$ of H has algebraic multiplicity of two. To see the geometric multiplicity of $\lambda_1 = \mu_1 = 0$, by noting the leading principal $(2n-1) \times (2n-1)$ minor of the coefficient matrix in (18) with $\tilde{d}_{i,k} = d_i$ and $\tilde{\delta}_{i,k} = \delta_i$ is nonzero, we conclude that the geometric multiplicity of $\lambda_1 = \mu_1 = 0$ is one. □

Let

$$W = \left[\frac{1}{d_i - \lambda_j}\right]_{i,j=1}^{n} := [w_{i,j}], \tag{20}$$

where $\{d_i\}_{i=1}^n$ and $\{\lambda_j\}_{j=1}^n$ are given in Lemma 2.3. Then W is a Cauchy matrix. We next state a result of [9] which is very useful in proving our main results.

Lemma 2.5. (i) (See *Theorem* 3.1 *of* [9].) *The matrix W defined in* (20) *is nonsingular, and its inverse is given by*

$$W^{-1} = D_1 W^T D_2, \tag{21}$$

where $D_1 = \operatorname{diag}(\alpha_1, \ldots, \alpha_n)$ and $D_2 = \operatorname{diag}(\beta_1, \ldots, \beta_n)$ with

$$\alpha_i = \frac{-\prod\limits_{j=1}^{n}(\lambda_i - d_j)}{\prod\limits_{j=1, j\neq i}^{n}(\lambda_i - \lambda_j)} \quad \text{and} \quad \beta_i = \frac{\prod\limits_{j=1}^{n}(d_i - \lambda_j)}{\prod\limits_{j=1, j\neq i}^{n}(d_i - d_j)}, \tag{22}$$

for i = 1,..., n. *Moreover, α_i, $\beta_i > 0$ for $1 \le i \le n$. For $n = 1$ the denominators in* (24) *is interpreted as 1.*

(ii) *Let $\alpha = [\alpha_1, \ldots, \alpha_n]^T$, $\beta = [\beta_1, \ldots, \beta_n]^T$. Then $W\alpha = e$ and $W^T\beta = e$. Here e is given in* (8b).

Proof. We need only to prove the second part of the lemma. To this end, let $W^T\beta = f = [f_1, f_2, \dots, f_n]^T$. By (20) and (22),

$$f_i = \sum_{k=1}^{n} \frac{\prod\limits_{j=1, j\neq i}^{n}(d_k - \lambda_j)}{\prod\limits_{j=1, j\neq k}^{n}(d_k - d_j)} =: \sum_{k=1}^{n} r_k.$$

Let $\phi_i(\lambda) = \prod\limits_{j=1, j\neq i}^{n}(\lambda - \lambda_j)$, which is an $n-1$ order monic polynomial. Set

$$\psi(\lambda) := \sum_{k=1}^{n} r_k \prod_{j=1, j\neq k}^{n}(\lambda - d_j).$$

Then $\psi(\lambda)$ is the Lagrangian interplating polynomial of $\phi_i(\lambda)$ at the points $d_1, \dots, d_n$. That is, $\phi_i(d_j) = \psi(d_j),\ j = 1, \dots, n$. Because the order of $\psi(\lambda)$ is also $n-1$, we have $\phi_i(\lambda) \equiv \psi(\lambda)$. By comparing the first coefficient, we get $f_i = 1$. So we have $W^T\beta = e$. Finally, $W^{-1}e = D_1 W^T D_2 e = D_1 W^T \beta = D_1 e = \alpha$. We thus complete the proof of the lemma. □

Remark 2.1. Let α_i and β_i be as given in (22) except with λ_1 on the respective products replaced by $-\mu_k$. Denote such new α_i and β_i by $\alpha_{i,k}$ and $\beta_{i,k}$, respectively. Then the assertions in the second part of Lemma 2.5 still hold for the corresponding W since the interlace property remains true.

We next show the matrices X_1 and $Z_1^{(k)}$, given in (23), are invertible. Such assertions will be used in constructing the solution of the algebraic Riccati equation (6) in section 3.

THEOREM 2.6. *Let*

$$X_1 = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,n} \end{bmatrix}, \quad Z_1^{(k)} = \begin{bmatrix} z_{1,k} & x_{1,2} & \cdots & x_{1,n} \\ z_{2,k} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & & \vdots \\ z_{n,k} & x_{n,2} & \cdots & x_{n,n} \end{bmatrix}, \tag{23}$$

where $x_{i,k}$ and $z_{i,k}$, $1 \le i, k \le n$, are defined in (16) *and* (17), *respectively. Then X_1 and $Z_1^{(k)}$, $\ k = 1, 2, \dots, n$, are invertible.*

Proof. From (16), decompose X_1 into

$$X_1 = D_q W D_\delta, \tag{24}$$

where

$$D_q = \mathrm{diag}[q_1, q_2, \dots, q_n], \tag{25a}$$

$$D_\delta = \mathrm{diag}[\delta_n + \lambda_1, \delta_n + \lambda_2, \dots, \delta_n + \lambda_n], \tag{25b}$$

and W is defined as in (20). Thus, the nonsingularity of W, and therefore of X_1, follows immediately from Lemma 2.5. The assertion for $Z_1^{(k)}, k = 1, \dots, n$, can be similarly obtained. □

COROLLARY 2.7. *Let $X_1^{-1} = [\widetilde{x}_{i,j}]_{i,j=1}^{n}$. Then it holds that*

$$\widetilde{x}_{i,j} > 0 \quad \mathit{for} \quad i \le j \qquad \mathit{and} \quad \widetilde{x}_{i,j} < 0 \quad \mathit{for} \quad i > j. \tag{26}$$

Similarly, let $(Z_1^{(k)})^{-1} = [\widetilde{z}_{i,j}^{(k)}]_{i,j}^{n}$. The corresponding elements $\widetilde{z}_{i,j}^{(k)}$, $i, j = 1, 2, \dots, n$, then satisfy the relationship shown in (26).

Proof. The statement (26) follows immediately from (21), (22), and (24). □

3. Existence and multiplicity of nonnegative solutions. Our object in this section is to study the existence and multiplicity of the nonnegative solutions of (6). To derive the main results, we first write (6) in the form

$$\begin{bmatrix} D & -C \\ B & -A \end{bmatrix}\begin{bmatrix} I \\ X \end{bmatrix} = \begin{bmatrix} I \\ X \end{bmatrix}(D - CX). \tag{27}$$

It is easily seen that the Span $\{\left[\begin{smallmatrix} I \\ X \end{smallmatrix}\right]\}$ forms an invariant subspace of H corresponding to the matrix $D - CX$. We first recall the following well-known theorem (see, e.g., [17]).

THEOREM 3.1. *If the Span* $\{\left[\begin{smallmatrix} X_1 \\ X_2 \end{smallmatrix}\right]\}$ *forms an invariant subspace of* H *associated with the matrix* $\Lambda \in R^{n\times n}$ *and if* X_1 *is invertible, then* $X = X_2X_1^{-1}$ *is a solution of* (6).

PROPOSITION 3.2. *If* X *is a nonnegative solution of* (6)*, then* $\{\lambda_2, \lambda_3, \ldots, \lambda_n\}$ *must be the eigenvalues of* $D - CX$*. Consequently,* (6) *has at most* $n+1$ *nonnegative solutions and at most* n *nonnegative solutions when* $c = 1$ *and* $\alpha = 0$.

Proof. Let X be a nonnegative solution of (6). Then,

$$D - CX = D_1 - q(e^T + q^TX) := D_1 - q\tilde{q}^T$$

with $\tilde{q}_i > 0$ for all i. The secular equation of $D_1 - q\tilde{q}^T$ is

$$s(\lambda) = 1 - \sum_{i=1}^{n} \frac{q_i\tilde{q}_i}{d_i - \lambda}.$$

Since $s(-\infty) > 0, s(d_1^-) < 0$, and $s(d_i^+)s(d_{i+1}^-) < 0$ for $i = 1, 2, \ldots, n-1$, we may conclude that $D - CX$ has n distinct real eigenvalues $\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_n$. Moreover, there are at least $n-1$ positive eigenvalues, say, $\tilde{\lambda}_i > 0, i = 2, \ldots, n$. Since $\sigma(D - CX)$, the spectrum of $D - CX$, is contained in $\sigma\left(\left[\begin{smallmatrix} D & -C \\ B & -A \end{smallmatrix}\right]\right)$, it then follows from Lemma 2.1 that $\lambda_i = \tilde{\lambda}_i$ for $i = 2, 3, \ldots, n$. The assertions in the proposition then follow from Theorem 3.1 and Proposition 2.4. □

Remark 3.1. From Lemma 2.1, Theorem 3.1, and Proposition 2.4, we conclude that (6) has at most $\binom{2n}{n} - \binom{2n-2}{n-1}$ solutions for $c = 1$ and $\alpha = 0$ and at most $\binom{2n}{n}$ solutions otherwise.

We next prove the following main result.

THEOREM 3.3. *Let* $\left[\begin{smallmatrix} X_1 \\ X_2 \end{smallmatrix}\right]$ *and* $\left[\begin{smallmatrix} Z_1^{(1)} \\ Z_2^{(1)} \end{smallmatrix}\right]$ *be the eigenvector matrices of* H *corresponding to* $\Lambda = \mathrm{diag}[\lambda_1, \lambda_2, \ldots, \lambda_n]$ *and* $\Gamma_1 = \mathrm{diag}[-\mu_1, \lambda_2, \ldots, \lambda_n]$*, respectively. Then* $X = X_2X_1^{-1}$ *and* $Z = Z_2^{(1)}(Z_1^{(1)})^{-1}$ *are positive solutions of Riccati equation* (6)*. Moreover,* $Z \geq X > 0$.

Proof. We first prove that $X_2X_1^{-1}$ is positive. Let $W_2 = \left[\frac{1}{\delta_i + \lambda_j}\right]$, D_δ, and D_q be given in (25). Using (16), we see that $X_2 = W_2D_\delta$ and $X_1 = D_qWD_\delta$. Hence,

$$X = W_2W^{-1}D_q^{-1} = W_2D_1W^TD_2D_q^{-1},$$

where D_1 and D_2 are given in (21). Let $X = [\chi_{i,j}]$. We see that

$$\chi_{i,j} = \left\{\sum_{\ell=1}^{n}\left(\frac{1}{\delta_i + \lambda_\ell}\right)\left(\frac{1}{d_j - \lambda_\ell}\right)\alpha_\ell\right\}\left(\frac{\beta_j}{q_j}\right).$$

Using the identity

$$\frac{1}{(\delta_i+\lambda_\ell)(d_j-\lambda_\ell)} = \frac{1}{\delta_i+d_j}\left(\frac{1}{\delta_i+\lambda_\ell}+\frac{1}{d_j-\lambda_\ell}\right)$$

and recognizing

$$\sum_{\ell=1}^{n}\frac{\alpha_\ell}{d_j-\lambda_\ell} = (W\alpha)_j = 1, \tag{28}$$

we see that

$$\chi_{i,j} = \frac{\beta_j}{q_j(\delta_i+d_j)}\left(1+\sum_{\ell=1}^{n}\frac{\alpha_\ell}{\delta_i+\lambda_\ell}\right) > 0. \tag{29}$$

The last equality in (28) is justified by Lemma 2.5 (ii). To complete the proof of the theorem, let $Z = Z_2^{(1)}(Z_1^{(1)})^{-1} = [\zeta_{i,j}]$. Here $Z_1^{(1)}$ is given in (23) and

$$Z_2^{(1)} = \begin{bmatrix} z_{n+1,1} & x_{n+1,2} & x_{n+1,3} & \cdots & x_{n+1,n} \\ z_{n+2,1} & x_{n+2,2} & x_{n+2,3} & \cdots & x_{n+2,n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ z_{n+n,1} & x_{n+n,2} & x_{n+n,3} & \cdots & x_{n+n,n} \end{bmatrix},$$

where $z_{n+i,1}$ and $x_{n+i,k}$ are defined in (16) and (17), respectively.

To complete the proof of the theorem, it remains to show that $Z - X \geq 0$. To this end, we see that

$$\begin{aligned} Z - X &= (Z_2^{(1)} - XZ_1^{(1)})(Z_1^{(1)})^{-1} \\ &= \{(Z_2^{(1)} - X_2) + X(X_1 - Z_1^{(1)})\}(Z_1^{(1)})^{-1} \\ &=: F(Z_1^{(1)})^{-1}. \end{aligned}$$

Using (16), (17) and doing some direct calculations, we see that F must be of the form $F = ge_1^T$, where g is a nonnegative vector and $e_1^T = (1,0,\ldots,0)^T$. Note, via Corollary 2.7, that $e_1^T(Z_1^{(1)})^{-1} \geq 0$. Thus $Z - X = F(Z_1^{(1)})^{-1} = ge_1^T(Z_1^{(1)})^{-1} \geq 0$. The proof of the theorem is thus complete. □

Remark 3.2. Using Remark 2.1 and a procedure similar to that above, we have that $\zeta_{i,j}$ are defined as in (29) except with λ_1 in the respective products and summation taken as $-\mu_1$; i.e.,

$$\zeta_{i,j} = \frac{\beta_{j,1}}{q_j(\delta_i+d_j)}\left(1+\frac{\alpha_{1,1}}{\delta_i-\mu_1}+\sum_{\ell=2}^{n}\frac{\alpha_{\ell,1}}{\delta_i+\lambda_\ell}\right). \tag{30}$$

THEOREM 3.4. *Equation* (6) *has a unique nonnegative solution when* $c = 1$ *and* $\alpha = 0$*; otherwise, when* $0 < c < 1$ *and* $0 < \alpha < 1$*, it has two nonnegative solutions.*

Proof. From Proposition 3.2, it suffices to show, at this stage, that for $k = 2,\ldots,n$, letting $\left[\begin{smallmatrix} Z_1^{(k)} \\ Z_2^{(k)} \end{smallmatrix}\right]$ be the eigenvector matrices of H corresponding to $\{-\mu_k, \lambda_2, \ldots, \lambda_n\}$, respectively, results in $Z^{(k)} = Z_2^{(k)}(Z_1^{(k)})^{-1}$ being other than nonnegative. However, these assertions follow directly from Corollary 2.7 and (17). □

4. Error analysis and numerical experiments.

In this section, we first provide a perturbation analysis of Riccati equation (6). For the case that $0 \le c < 1$ and $0 < \alpha < 1$, one can apply the standard theory as discussed in Byers [4] and Kenney and Laub [16]. Let $X = X_2 X_1^{-1}$ be the positive solution of (6). Let $\|\cdot\|$ denote the Frobenius norm and $P_X(\delta)$ be the set of perturbations with respect to X:

$$P_X(\delta) = \left\{ \frac{\|\Delta X\|}{\delta \|X\|} : \frac{\|\Delta A\|}{\|A\|} \le \delta,\ \frac{\|\Delta D\|}{\|D\|} \le \delta,\ \frac{\|\Delta C\|}{\|C\|} \le \delta \right\}.$$

Here $\widetilde{A} = A + \Delta A$, $\widetilde{D} = D + \Delta D$, and $\widetilde{C} = C + \Delta C$ are some perturbed matrices in a δ-neighborhood of A, D, and C, respectively, and $\widetilde{X} = X + \Delta X$ is the corresponding perturbed solution. Note that $B = ee^T$ has no perturbation error. Following Rice [20], we define the (asymptotic) condition number of (6) as follows:

$$\nu_X(A, C, D) = \lim_{\delta \to 0} \sup\{P_X(\delta)\}. \tag{31}$$

The magnitude of $\nu_X(A,C,D)$ is used to measure the sensitivity of the solution of (6) to perturbations in the data. If $\nu_X(A,C,D)$ is "large," then small changes in the data make large changes in the solution. Consequently, the Riccati equation (6) is ill conditioned. If $\nu_X(A,C,D)$ is of "modest magnitude," then the small changes in the data make small changes in the solution. Hence, the corresponding Riccati equation (6) is well conditioned. Define

$$K_X(A, C, D) = \frac{\|\Theta_X\| \max\{\|A\|, \|D\|\} + \|\Pi_X\| \|C\|}{\|X\|}, \tag{32}$$

where Θ_X and Π_X are linear operators on $\mathbf{R}^{n \times n}$, respectively, given by

$$\Theta_X(V) = \Omega_X^{-1}(V^T X + XV) \ \text{ and } \ \Pi_X(V) = \Omega_X^{-1}(XVX)$$

with

$$\Omega_X(V) = (-XC + A)V + V(D - CX).$$

By a similar argument in [4], one can also show that

$$\frac{1}{9} K_X(A, C, D) \le \nu_X(A, C, D) \le 4 K_X(A, C, D). \tag{33}$$

From (27), we have that the eigenvalues of Ω_X are of the form $\lambda_k + \mu_\ell$, where $\{\lambda_k\}_{k=1}^n$ and $\{\mu_\ell\}_{\ell=1}^n$, defined in Lemma 2.1, are eigenvalues of $(D - CX)$ and $(A - XC)$, respectively. Expressions (32) and (33) show that Riccati equation (6) is ill conditioned when $\|\Omega_X^{-1}\|$ is large. The quantity $\|\Omega_X^{-1}\|^{-1}$ is usually measured by $sep((D - CX), -(A - XC))$ [21]. From the definition of "*sep*" it follows that

$$\frac{1}{\displaystyle\min_{1 \le k, \ell \le n} |\lambda_k + \mu_\ell|} \le \|\Omega_X^{-1}\| = [sep((D - CX), -(A - XC))]^{-1}. \tag{34}$$

We next apply the above perturbation analysis to the case that $c \approx 1$ and $\alpha \approx 0$. From Lemma 2.1, we see that $\lambda_1 \to 0^+$ and $-\mu_1 \to 0^-$ as $c \to 1^-$ and $\alpha \to 0^+$. In this case, $\frac{1}{|\lambda_1 + \mu_1|}$; hence, $\nu_X(A,C,D)$ become very large. Therefore, the Riccati equation (6) for $c \approx 1$ and $\alpha \approx 0$ is very ill conditioned. This shows that the convergence rates of some iterative methods of [14] for solving Riccati equation (6) are very slow and unsatisfactory.

We now turn our attention to the formulae derived in (29) and (30). The nonnegative solutions $X = X_2 X_1^{-1}$ and $Z = Z_2^{(1)} Z_1^{(1)-1}$ are as in Theorem 3.4 and thus can be computed directly by (29), (30), and (22). In the following we give an error analysis on the method of (29) and (30). For simplicity, we suppose the given data $\{d_i, \delta_i\}_{i=1}^n$ have no error propagation in computation and let ϵ_{λ_i} be the relative error of λ_i caused by computation. By a standard technique of error analysis (see, e.g., [22, Chap. 1]), one can derive the relative errors for the computation of $\{\alpha_i\}_{i=1}^n$ and $\{\beta_i\}_{i=1}^n$ of (22) as follows:

$$\text{rel}(\alpha_i) \equiv \epsilon_{\alpha_i} = \sum_{j=1}^{n} \frac{\lambda_i}{\lambda_i - d_j} \epsilon_{\lambda_i} - \sum_{j=1, j\neq i}^{n} \left(\frac{\lambda_i}{\lambda_i - \lambda_j} \epsilon_{\lambda_i} - \frac{\lambda_j}{\lambda_i - \lambda_j} \epsilon_{\lambda_j} \right) \tag{35a}$$

and

$$\text{rel}(\beta_i) \equiv \epsilon_{\beta_i} = \sum_{j=1}^{n} \frac{\lambda_j}{\lambda_j - d_i} \epsilon_{\lambda_j}. \tag{35b}$$

Here $\text{rel}(x)$ denotes the relative error for the computation of x. Let

$$\epsilon_{\max} := \max\{|\epsilon_{\lambda_i}|,\ i = 1, \ldots, n\}, \tag{36a}$$

$$\theta_{\max} := \max\left\{ \frac{|\lambda_i|}{|\lambda_i - d_i|},\ \frac{|\lambda_i|}{|\lambda_i - d_{i-1}|},\ i = 1, \ldots, n,\ d_0 = -\infty \right\}, \tag{36b}$$

$$\tilde{\theta}_{\max} := \max\left\{ \left| \frac{\lambda_i + \lambda_{i+1}}{\lambda_{i+1} - \lambda_i} \right|,\ i = 1, \ldots, n \right\}, \tag{36c}$$

$$d_{\min} := \min\left\{ |d_i - d_{i+1}|,\ i = 1, \ldots, n \right\}. \tag{36d}$$

From (35), (36), and (11) we can estimate $|\epsilon_{\alpha_i}|$ and $|\epsilon_{\beta_i}|$ as follows:

$$\begin{aligned}
|\epsilon_{\alpha_i}| &\leq \left[\left(\frac{|\lambda_i|}{|\lambda_i - d_i|} + \frac{|\lambda_i|}{|\lambda_i - d_{i-1}|} + \left(\sum_{j=1}^{i-2} + \sum_{j=i+1}^{n} \right) \frac{|\lambda_i|}{|\lambda_i - d_j|} \right) \right. \\
&\qquad \left. + \left(\frac{|\lambda_{i+1} + \lambda_i|}{|\lambda_{i+1} - \lambda_i|} + \frac{|\lambda_i + \lambda_{i-1}|}{|\lambda_i - \lambda_{i-1}|} + \left(\sum_{j=1}^{i-2} + \sum_{j=i+1}^{n} \right) \frac{|\lambda_i + \lambda_j|}{|\lambda_i - \lambda_j|} \right) \right] \epsilon_{\max} \\
&\leq \left[2\left(\theta_{\max} + \tilde{\theta}_{\max} \right) + \frac{2d_i + d_n}{d_{\min}} \left(\ell n(4(i-2)(n-i)) \right) \right] \epsilon_{\max} \\
&\leq \left[2\left(\theta_{\max} + \tilde{\theta}_{\max} \right) + \frac{6d_n}{d_{\min}} \ell n(2n) \right] \epsilon_{\max}
\end{aligned} \tag{37}$$

and

$$\begin{aligned}
|\epsilon_{\beta_i}| &\leq \left[\frac{|\lambda_i|}{|\lambda_i - d_i|} + \frac{|\lambda_{i+1}|}{|\lambda_{i+1} - d_i|} + \left(\sum_{j=1}^{i-2} + \sum_{j=i+1}^{n} \right) \frac{|\lambda_j|}{|\lambda_j - d_i|} \right] \epsilon_{\max} \\
&\leq \left[2\theta_{\max} + \frac{d_n}{d_{\min}} \left(\ell n(4(i-1)(n-i+1)) \right) \right] \epsilon_{\max} \\
&\leq \left[2\theta_{\max} + \frac{2d_n}{d_{\min}} \ell n(2n) \right] \epsilon_{\max}.
\end{aligned} \tag{38}$$

Table 4.1
($c = 0.999999,\ \alpha = 10^{-8}$).

	$n = 32$	$n = 64$	$n = 128$	$n = 256$
r_X	2.1807e-13	4.3211e-12	3.1650e-11	1.0723e-10
r_Z	2.1926e-13	4.4682e-12	3.1528e-11	1.2455e-10
λ_1	1.73206684765e-3	1.73206684059e-3	1.73206683757e-3	1.73206678707e-3
$-\mu_1$	–1.73203684762e-3	–1.73203684057e-3	–1.73203683731e-3	–1.73203678614e-3

Table 4.2
($c = 1.0,\ \alpha = 10^{-14}$).

	$n = 32$	$n = 64$	$n = 128$	$n = 256$
r_X	6.9022e-13	4.3476e-12	4.8312e-11	2.0916e-10
λ_1	1.72951930554e-15	4.14078493715e-15	2.15344021678e-15	2.61125671244e-15

Here ℓn denotes the natural logarithm.

Consequently, from (37) and (38) the relative error for the computation of $x_{i,j}$ is bounded by

$$|\mathrm{rel}(x_{i,j})| \le \left[1 + 4\theta_{\max} + 2\tilde{\theta}_{\max} + \frac{8d_n}{d_{\min}}\ell n(2n)\right]\epsilon_{\max} \tag{39}$$

for $i,\ j = 1, \dots, n$. From (4) and (8) we see that the distances between d_i and d_{i+1} are well separated. Moreover, using the fact that $\lambda_i \in (d_i, d_{i+1})$ and the secular equation $f(\lambda)$ in (14), we see that λ_i is well separated from the end points d_i and d_{i+1}. Therefore, the quantities $\theta_{\max}$ and $\tilde{\theta}_{\max}$ and $\frac{1}{d_{\min}}$ defined in (36) cannot become too large. Thus, the relative error of $x_{i,j}$ depends on the quantity of $\epsilon_{\max}$. Indeed, a bisection method combined with Newton's acceleration scheme can be applied to $f(\lambda)$ in (14) for computing the desired eigenvalues $\{\lambda_i\}_{i=1}^n$ accurately. Numerical stability for the computation $\{x_{i,j}\}_{i,j=1}^n$ and $\{z_{i,j}^{(1)}\}_{i,j=1}^n$ is guaranteed, even when the problem for solving Riccati equation (6) is ill posed for $c \approx 1$ and $\alpha \approx 0$.

In the following, we give the numerical results of our test examples. We compute the nonnegative solutions X and Z by using the formulae (29) and (30) with different matrix sizes, $n = 32$, 64, 128, and 256. In Table 4.1 we compute nonnegative solutions X and Z for $c = 0.999999$ and $\alpha = 10^{-8}$. In Table 4.2 we compute the unique nonnegative solution X for $c = 1.0$ and $\alpha = 10^{-14}$. Here r_X and r_Z denote, respectively, the 2-norm residuals of Riccati equation (6) for the computed solutions X and Z.

As mentioned in section 1 for the case in which $c \approx 1$ and $\alpha \approx 0$, iterative procedures [14] can cause numerical problems during the convergence process. Our numerical result shows that the residual r_X of the computed nonnegative solution is very satisfactory, even when the condition number $\nu_X(A, C, D)$ estimated by (32), (33), and (34) is very "large" for $c = 1$ and $\alpha = 10^{-14}$.

5. Comparison theorems for nonnegative solutions. Noting that only minimal solution is physically meaningful (see, e.g., [2]), we show in this section that the minimal nonnegative solution X of (6) is increasing in c and decreasing in α. The dependency of X on the parameter c is well known. However, the effect of the parameter α on X is less understood. Our assertions here provide a better picture as to how the roles of α are played.

To begin, we consider the following iteration:

$$AX^{(p+1)} + X^{(p+1)}(D - CX^{(p)})) = B = ee^T, \tag{40}$$

with $X^{(0)} = 0$. Let $X = X_2X_1^{-1}$ be as given in Theorem 3.3, and set

$$D - CX^{(p)} := \Lambda_p. \tag{41}$$

Let $X^{(p+1)} = [x_{i,j}^{(p+1)}]$. Equation (40) can be equivalently written as a linear system of the form

$$(A \otimes I + I \otimes \Lambda_p)[x_{11}^{(p+1)}, \ldots, x_{1n}^{(p+1)}, x_{21}^{(p+1)}, \ldots, x_{nn}^{(p+1)}] = [1, 1, \ldots, 1]^T. \tag{42}$$

Here $\otimes$ denotes the Kronecker product (see, e.g., [1]). To save notation, we shall write (42) as

$$(A \otimes I + I \otimes \Lambda_p)X^{(p+1)} = B. \tag{43}$$

We then prove the following lemmas.

Lemma 5.1. *(i) $\Lambda_0 = D$ is an M-matrix. (ii) $\overline{\Lambda} = X_1\Lambda X_1^{-1} = D - CX$ is an M-matrix.*

Proof. To see the first assertion of the lemma, we note, via Lemma 2.2, that the eigenvalues of D are real and nonnegative. Using the fact that off-diagonal elements of D are nonpositive, we conclude that Λ_0 is an M-matrix. The second assertion also follows from the fact that off-diagonal elements of $\overline{\Lambda}$ are nonpositive and its eigenvalues are $\{\lambda_i\}$, which are nonnegative. □

Lemma 5.2. *(i) $A \otimes I + I \otimes \overline{\Lambda}$ is an M-matrix. (ii) Let $p \in N \cup \{0\}$. For matrix $\Lambda_p = D - CX^{(p)}$, with $0 \leq X^{(p)} \leq X$, $A \otimes I + I \otimes \Lambda_p$ is an M-matrix, and $(A \otimes I + I \otimes \overline{\Lambda})^{-1} \geq (A \otimes I + I \otimes \Lambda_p)^{-1}$.*

Proof. We first note, via Lemma 2.2, that the eigenvalues of A are positive. It is then clear that the off-diagonal elements of $A \otimes I + I \otimes \overline{\Lambda}$ are nonpositive and its eigenvalues are nonnegative. Hence, $A \otimes I + I \otimes \overline{\Lambda}$ is an M-matrix. The second part of the lemma follows by applying Theorem 1.3 on $A \otimes I + I \otimes \overline{\Lambda} (= A_1)$ and $A \otimes I + I \otimes \Lambda_p (= A_2)$. □

We are now ready to prove the following theorem.

Lemma 5.3. *Let $X = X_2X_1^{-1}$ be as given in Theorem 3.4. Then $X^{(p)}$, as defined in (40), converge upward to X.*

Proof. We first prove that

$$0 \leq X^{(p-1)} \leq X^{(p)} \leq X \qquad \text{for all } p \in \mathbf{N}.$$

To see this, we note that

$$0 = X^{(0)} \leq X^{(1)}.$$

Moreover, using Lemma 5.2 (ii), we get

$$X = (A \otimes I + I \otimes \overline{\Lambda})^{-1}B \geq (A \otimes I + I \otimes \Lambda_0)^{-1}B = X^{(1)},$$

and so

$$0 = X^{(0)} \leq X^{(1)} \leq X.$$

Suppose (42) holds for $p = k$. Then we see, as in Lemma 5.2 (ii), that

$$(A \otimes I + I \otimes \Lambda_{k-1})^{-1} \leq (A \otimes I + I \otimes \Lambda_k)^{-1} \leq (A \otimes I + I \otimes \overline{\Lambda})^{-1}.$$

Using (41), we obtain that

$$0 \leq X^{(k)} \leq X^{(k+1)} \leq X.$$

Therefore, we conclude, via an induction, that (42) holds as claimed. Let the limit of the sequence $\{X^{(p)}\}$ be denoted by $X^{(\infty)}$. Since $X^{(\infty)}$ is a nonnegative solution of (6) and $X^{(\infty)} \leq X$, it must be that $X^{(\infty)} = X$. □

To emphasize the dependence of X on the parameters c and α, we write X as $X(c, \alpha)$. Likewise, all quantities are similarly written if necessary. We are now ready to state our comparison result.

THEOREM 5.4. *The solution $X = X_2 X_1^{-1}$ of* (6) *is increasing in c and decreasing in α. In particular, $X(1, 0) \geq X(c, \alpha)$ for all c, α.*

Proof. For fixed α and $c_1 \leq c_2$, suppose $X^{(p)}(c_1, \alpha) \leq X^{(p)}(c_2, \alpha)$, where $X^{(p)}(c_i, \alpha), i = 1, 2$, are as defined in (40). Then by applying Theorem 1.3 on

$$A_1 = A(c_2, \alpha) \otimes I + I \otimes \Lambda_p(c_2, \alpha) \qquad \text{and} \qquad A_2 = A(c_1, \alpha) \otimes I + I \otimes \Lambda_p(c_1, \alpha),$$

we get

$$(A(c_1, \alpha) \otimes I + I \otimes \Lambda_p(c_1, \alpha))^{-1} \leq (A(c_2, \alpha) \otimes I + I \otimes \Lambda_p(c_2, \alpha))^{-1}.$$

Here $\Lambda_p(c_i, \alpha) := D(c_i, \alpha) - CX^{(p)}(c_i, \alpha)$, $i = 1, 2$. It then follows from (41) that

$$X^{(p+1)}(c_1, \alpha) \leq X^{(p+1)}(c_2, \alpha).$$

Clearly,

$$0 = X^{(0)}(c_1, \alpha) \leq X^{(0)}(c_2, \alpha) = 0.$$

Hence, an induction yields that $X^{(p)}(c_1, \alpha) \leq X^{(p)}(c_2, \alpha)$ for all $p \in \mathbf{N}$. We conclude, via Lemma 5.3, that $X(c_1, \alpha) \leq X(c_2, \alpha)$.

To see $X(c, \alpha)$ is decreasing in α, we first note that

$$\delta_i + d_i = \frac{1}{cw_i(1+\alpha)} + \frac{1}{cw_i(1-\alpha)} = \frac{2}{cw_i(1-\alpha^2)}$$

are increasing in α. Therefore, for fixed c and $\alpha_1 \leq \alpha_2$, suppose $X^{(p)}(c, \alpha_1) \geq X^{(p)}(c, \alpha_2)$. Then by applying Theorem 1.3 on $A_1 = A(c, \alpha_1) \otimes I + I \otimes \Lambda_p(c, \alpha_1)$ and $A_2 = A(c, \alpha_2) \otimes I + I \otimes \Lambda_p(c, \alpha_2)$, we get

$$[A(c, \alpha_1) \otimes I + I \otimes \Lambda_p(c, \alpha_1)]^{-1} \geq [A(c, \alpha_2) \otimes I + I \otimes \Lambda_p(c, \alpha_2)]^{-1}.$$

Noting that

$$0 = X^{(0)}(c, \alpha_1) \geq X^{(0)}(c, \alpha_2) = 0,$$

we conclude, via an induction and Lemma 5.3, that

$$X(c, \alpha_1) \geq X(c, \alpha_2).$$

The proof of the theorem is thus complete. □

6. Concluding remarks.

We conclude with a few suggestions for further related work.

First, the method of invariant embedding has been applied to transport problems (see, e.g., [10]) involving neutrons and gamma rays with realistic energy and angle-dependent cross-sections. It is therefore of interest to study a more general form of algebraic matrix Riccati equations encompassing those cases.

Next, we note that the simple transport model [8, 10] in an isotropically scattering plane-parallel layer of finite thickness would induce a differential Riccati equation of the form

$$X' = B - AX - XD + XCX, \tag{44a}$$

$$X(0) = 0. \tag{44b}$$

Here A, B, C, and D are as defined in (7). It would be worthwhile to pursue the asymptotic characteristics and stability of the nonnegative solutions of (6) with respect to this differential Riccati equation (44).

Finally, it would be desirable to generalize our techniques for solving the corresponding algebraic Riccati equation to infinitely dimensional cases [15].

Acknowledgments. We thank Professor Volker Mehrmann and referees for suggesting numerous improvements to the original draft. In particular, the inversion formula of a Cauchy matrix was brought to our attention, which leads to a much shorter proof of Theorem 3.4 and Lemma 2.5 (ii).

REFERENCES

[1] R. Bellman, *Introduction to Matrix Analysis*, 2nd ed., McGraw-Hill, New York, 1970.

[2] R. Bellman and G. M. Wing, *An Introduction to Invariant Embedding*, John Wiley, New York, 1975.

[3] J. R. Bunch, C. R. Nielsen, and D. C. Sorensen, *Rank-one modification of the symmetric eigenproblem*, Numer. Math., 31 (1978), pp. 31–48.

[4] R. Byers, *Numerical stability and instability in matrix sign function based in algorithms*, in Computational and Combined Methods in Systems Theory, C. I. Byrnes and A. Lindquist, eds., North-Holland, New York, 1986, pp. 185–200.

[5] S. Chandrasekhar, *Radiative Transfer*, Dover, New York, 1960.

[6] J. J. M. Cuppen, *A divide and conquer method for the symmetric tridiagonal eigenproblem*, Numer. Math., 36 (1981), pp. 177–195.

[7] D. J. Clements and B. D. O. Anderson, *Polynomial factorization via the Riccati equation*, SIAM J. Appl. Math., 31 (1976), pp. 179–205.

[8] F. Coron, *Computation of the asymptotic states for linear half space kinetic problem*, Transport Theory Statist. Phys., 19 (1990), pp. 89–114.

[9] T. Finck, G. Heinig, and K. Rost, *An inversion formula and fast algorithms for Cauchy-Vandermonde matrices*, Linear Algebra Appl., 183 (1993), pp. 179–197.

[10] B. D. Ganapol, *An investigation of a simple transport model*, Transport Theory Statist. Phys., 21 (1992), pp. 1–37.

[11] G. H. Golub, *Some modified matrix eigenvalue problems*, SIAM Rev., 15 (1973), pp. 318–334.

[12] G. H. Golub and J. H. Wilkinson, *Ill-conditioned eigensystems and the computations of the Jordan canonical form*, SIAM Rev., 18 (1976), pp. 578–619.

[13] J. Juang, *Existence of algebraic matrix Riccati equations arising in transport theory*, Linear Algebra Appl., 230 (1995), pp. 89–100.

[14] J. Juang and I-Der Chen, *Iterative solution for a certain class of algebraic matrix Riccati equations arising in transport theory*, Transport Theory Statist. Phys., 21 (1993), pp. 65–80.

[15] J. Juang and P. Nelson, *Global existence, asymptotic and uniqueness for the reflection kernel of the angularly shifted transport equation*, Math. Models Methods Appl. Sci., 5 (1995), pp. 239–251 .

[16] C. Kenney and A. J. Laub, *Condition estimation for matrix functions*, SIAM J. Matrix Anal. Appl., 10 (1989), pp. 191–209.

[17] V. L. Mehrmann, *The Autonomous Linear Quadratic Control Problem*, Springer-Verlag, Berlin, 1991.

[18] H. Meyer, *The matrix equation* $AZ + B - ZCZ - ZD = 0$, SIAM J. Appl. Math., 30 (1976), pp. 136-142.

[19] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[20] J. R. Rice, *A theory of condition*, SIAM J. Numer. Anal., 3 (1966), pp. 287–310.

[21] G. W. Stewart, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Rev., 15 (1973), pp. 727–764.

[22] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, Springer-Verlag, New York, 1980.