PERTURBATION ANALYSIS FOR SOLUTIONS OF ALGEBRAIC RICCATI EQUATIONS*

CHEN CHUN-HUI (陈春晖)
(*Peking University, Beijing, China*)

Abstract

This paper discusses the conditioning of algebraic Riccati equations, i. e. the influence of perturbations in data on the positive semi-definite solution. A perturbation bound for the solution is given.

Notation. The symble $\mathbb{C}^{m\times n}$ denotes the set of complex $m\times n$ matrices, and $\mathbb{C}^n=\mathbb{C}^{n\times 1}$. $\|\cdot\|_2$ denotes the spectral norm and the Euclidean vector norm. The superscript H is for conjugate transpose. $A\geqslant 0$ means that matrix A is positive semi-definite. $\lambda(A)$ denotes the spectrum of a matrix A. I_n denotes the n-th order identity matrix. Re λ denotes the real part of a complex number λ.

§ 1. Introduction

Algebraic Riccati equations arise in optimal control applications. The algebraic Riccati equation for continuous-time systems takes the form

$$A^H X + XA - XNX + K = 0, \tag{1.1}$$

where A, N, $K\in\mathbb{C}^{n\times n}$, $N^H=N\geqslant 0$, $K^H=K\geqslant 0$. The positive semi-definite solution $X=X^H\geqslant 0$ of (1.1) is required.

Let $N=BB^H$ and $K=C^HC$ be full-rank factorizations of N and K, respectively. Under the assumption that (A, B) is stabilizable and (C, A) is detectable, (1.1) is known to have a unique positive semi-definite solution X, and $A-NX$ is stable.

Definition 1.1. *$M\in\mathbb{C}^{2n\times 2n}$ is said to be Hamiltonian if $J^{-1}MJ=-M^H$, where* $J=\begin{pmatrix}0 & I_n\\ -I_n & 0\end{pmatrix}$.

Now consider the Hamiltonian matrix

$$M=\begin{pmatrix}A & N\\ K & -A^H\end{pmatrix}. \tag{1.2}$$

Under the assumption above, the eigenvalues of M have nonzero real part. If $\begin{pmatrix}U_1\\ U_2\end{pmatrix}$ is a $2n\times n$ matrix such that $M\begin{pmatrix}U_1\\ U_2\end{pmatrix}=\begin{pmatrix}U_1\\ U_2\end{pmatrix}S$, where S is stable, U_1 is invertible and $X=-U_2U_1^{-1}$ is the positive semidefinite solution of (1.1).

* Received January 12, 1987.

The conditioning of algebraic Riccati equations, i. e. the influence of perturbations in the data on the solution, was studied to some extent in [3], [4] and [7]. [5] pointed out that it is still an open problem.

By using the perturbation theorem of invariant subspaces of a matrix, [7] obtained some useful results. This paper will continue the discussion on this problem.

§ 2. The Separation of a Stable Matrix

In [6] the separation of two matrices is defined and denoted by $\mathrm{sep}(A, B)$. Now we introduce the following definition.

Definition 2.1. *Let $A\in\mathbb{C}^{n\times n}$. The separation of A is the number* $\mathrm{sep}(A)$ *defined by*

$$\mathrm{sep}(A)=\inf_{\substack{P^H=P\\ \|P\|=1}}\|PA+A^HP\|, \tag{2.1}$$

where $\|\cdot\|$ denotes any consistent norm on $\mathbb{C}^{n\times n}$.

In particular, when the norm in (2.1) is taken to be the spectral norm and Frobenious norm, it is denoted by $\mathrm{sep}_2(A)$ and $\mathrm{sep}_F(A)$, respectively.

By [6], it is easy to prove that $\mathrm{sep}(A)$ has the following properties:

Property 1. Let $A, X\in\mathbb{C}^{n\times n}$ with X nonsingular. Then

$$\mathrm{sep}(X^{-1}AX)\geqslant\frac{\mathrm{sep}(A)}{\varkappa(X)\varkappa(X^H)},$$

where $\varkappa(X)=\|X\|\|X^{-1}\|$. If X is unitary, then

$$\mathrm{sep}_P(X^HAX)=\mathrm{sep}_P(A),\ P=2,\ F.$$

Property 2. Let $A, E\in\mathbb{C}^{n\times n}$. Then

$$\mathrm{sep}(A)-(\|E\|+\|E^H\|)\leqslant\mathrm{sep}(A+E)\leqslant\mathrm{sep}(A)+(\|E\|+\|E^H\|).$$

Property 3. Let $A\in\mathbb{C}^{n\times n}$ with $\lambda(A)=\{\lambda_i: i=1, 2, \cdots, n\}$. Then

$$\mathrm{sep}_P(A)\leqslant 2\min_{1\leqslant i\leqslant n}|\mathrm{Re}\,\lambda_i|,\ P=2, F.$$

On that basis, we will give a further discussion on the property of the separation of a stable matrix. Let $A\in\mathbb{C}^{n\times n}$ be a stable marix with $\lambda(A)=\{\lambda_i(A): i=1, 2, \cdots, n, |\mathrm{Re}\ \lambda_1(A)|\geqslant\cdots\geqslant|\mathrm{Re}\ \lambda_n(A)|\}$. If P is Hermitian, write $\lambda(P)=\{\lambda_i(P): i=1, 2, \cdots, n, |\lambda_1(P)|\geqslant\cdots\geqslant|\lambda_n(P)|\}$.

It is easy to prove the following lemma.

Lemma 2.1. *Let $H\in\mathbb{C}^{n\times n}$ be Hermitian. Then*

$$\|H\|_2=\max_{\substack{x\in\mathbb{C}^n\\ \|x\|_2=1}}|x^HHx|.$$

In addition, if the signs of eigenvalues of H are the same, then

$$|\lambda_n(H)|=\min_{\substack{x\in\mathbb{C}^n\\ \|x\|_2=1}}|x^HHx|.$$

By Lemma 2.1, we can estimate a lower bound of the separation of a stable matrix.

Theorem 2.1. *Let $A\in\mathbb{C}^{n\times n}$ be stable.*

(1) *If A is normal, then*

$$\operatorname{sep}_2(A)=2\min_{1\le i\le n}\operatorname{Re}|\lambda_i(A)|. \tag{2.2}$$

(2) *If A is diagonalizable and $X\in\mathbb{C}^{n\times n}$ is nonsingular such that $X^{-1}AX=\Lambda$, where Λ is diagonal, then*

$$\operatorname{sep}_2(A)\geqslant\frac{2}{\varkappa_2^2(X)}\min_{1\le i\le n}|\operatorname{Re}\lambda_i(A)|. \tag{2.3}$$

(3) *If A is undiagonalizable and $X\in\mathbb{C}^{n\times n}$ is nonsingular such that $X^{-1}AX=J_A$, where J_A is the Jordan canonical form, and if ν denotes the highest order of the Jordan blocks of J_A, then*

$$\operatorname{sep}_2(A)\geqslant\begin{cases}\dfrac{1}{\varkappa_2^2(x)}\cdot\dfrac{1}{\nu-1}\left(\dfrac{2\nu-2}{2\nu-1}\min\limits_{1\le i\le n}|\operatorname{Re}\lambda_i(A)|\right)^{2\nu-1}, & \min\limits_{1\le i\le n}|\operatorname{Re}\lambda_i(A)|\leqslant 1,\\ \dfrac{1}{\varkappa_2(x)}\max\left\{2(\min\limits_{1\le i\le n}|\operatorname{Re}\lambda_i(A)|-1),\ \dfrac{1}{\nu-1}\left(\dfrac{2\nu-2}{2\nu-1}\min\limits_{1\le i\le n}|\operatorname{Re}\lambda_i(A)|\right)^{2\nu-1}\right\}, & \\ & 1<\min\limits_{1\le i\le n}|\operatorname{Re}\lambda_i(A)|\leqslant\dfrac{2\nu-1}{2\nu-2},\\ \dfrac{1}{\varkappa_2(x)}(\min\limits_{1\le i\le n}|\operatorname{Re}\lambda_i(A)|-1), & \min\limits_{1\le i\le n}|\operatorname{Re}\lambda_i(A)|>\dfrac{2\nu-1}{2\nu-2}.\end{cases} \tag{2.4}$$

Proof. (1) If A is normal, there exists a unitary matrix X such that $X^{-1}AX=\Lambda$, where Λ is diagonal. By Property 2 and Lemma 2.1, we have

$$\operatorname{sep}_2(A)=\operatorname{sep}_2(\Lambda)=\inf_{\substack{P^H=P\\ \|P\|_2=1}}\|P\Lambda+\Lambda^HP\|_2=\inf_{\substack{P^H=P\\ \|P\|_2=1}}\max_{\substack{x\in\mathbb{C}^n\\ \|x\|_2=1}}|x^H(P\Lambda+\Lambda^HP)x|.$$

For the given Hermitian matrix P, there exists a unit vector x_P such that $Px_P=\lambda_1(P)x_P$. Thus.

$$\begin{aligned}\operatorname{sep}_2(A)&\geqslant\inf_{\substack{P^H=P\\ \|P\|_2=1}}(\lambda_1(P)|\cdot|x_P^H(\Lambda+\Lambda^H)x_P|)=\inf_{\substack{P^H=P\\ \|P\|_2=1}}|x_P^H(\Lambda+\Lambda^H)x_P|\\ &\geqslant\min_{\substack{x\in\mathbb{C}^n\\ \|x\|_2=1}}|x^H(\Lambda+\Lambda^H)x|=|\lambda_n(\Lambda+\Lambda^H)|\geqslant 2\min_{1\le i\le n}|\operatorname{Re}\lambda_i(A)|.\end{aligned}$$

Combining with Property 3, we get (2.2) at once.

(2) By Property 2 and (1), (2.3) is immediate.

(3) From $X^{-1}AX=J_A$, it follows that

$$\operatorname{sep}_2(A)\geqslant\frac{1}{\varkappa_2^2(X)}\operatorname{sep}_2(J_A). \tag{2.5}$$

Let

$$J_A=\operatorname{diag}(J_1,J_2,\cdots,J_k),\ J_i=\begin{pmatrix}\lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1\\ & & & \lambda_i\end{pmatrix}\in\mathbb{C}^{\nu_i\times\nu_i},\quad i=1,2,\cdots,k.$$

Then $\nu=\max\limits_{1\le i\le n}\{\nu_i\}$. Let $D=\operatorname{diag}(D_1,D_2,\cdots,D_k)$, $D_i=\operatorname{diag}(1,\varepsilon,\cdots,\varepsilon^{\nu_i-1})$, where $\varepsilon\in(0,1]$ is to be determined. Then

$$D^{-1}J_AD=J_A^{(\varepsilon)},\ J_A^{(\varepsilon)}=\operatorname{diag}(J_1^{(\varepsilon)},J_2^{(\varepsilon)},\cdots,J_k^{(\varepsilon)}), \tag{2.6}$$

$$J_i^{(\varepsilon)}=\begin{pmatrix}\lambda_i & \varepsilon & & \\ & \ddots & \ddots & \\ & & & \varepsilon \\ & & & \lambda_i\end{pmatrix}, \quad i=1, 2, \cdots, k. \tag{2.7}$$

By Property 2, we have

$$\mathrm{sep}_2(J_A)\geqslant\frac{1}{\kappa_2^2(D)}\,\mathrm{sep}_2(J_A^{(\varepsilon)})=\frac{1}{\varepsilon^{2(\nu-1)}}\,\mathrm{sep}_2(J_A^{(\varepsilon)}). \tag{2.8}$$

In a manner similar to the proof in (1), we get

$$\mathrm{sep}_2(J_A^{(\varepsilon)})\geqslant\min_{\substack{x\in C^n\\ \|x\|_2=1}}|x^H(J_A^{(\varepsilon)}+(J_A^{(\varepsilon)})^H)x|, \tag{2.9}$$

where

$$J_A^{(\varepsilon)}+(J_A^{(\varepsilon)})^H=\mathrm{diag}(J_1^{(\varepsilon)}+(J_1^{(\varepsilon)})^H, \cdots, J_k^{(\varepsilon)}+(J_k^{(\varepsilon)})^H),$$

$$J_i^{(\varepsilon)}+(J_i^{(\varepsilon)})^H=\begin{pmatrix}2\,\mathrm{Re}\,\lambda_i & \varepsilon & & \\ \varepsilon & \ddots & \ddots & \\ & \ddots & \ddots & \varepsilon \\ & & \varepsilon & 2\,\mathrm{Re}\,\lambda_i\end{pmatrix}=2\,\mathrm{Re}\,\lambda_i\begin{pmatrix}1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1\end{pmatrix}+\varepsilon\begin{pmatrix}0 & 1 & & \\ 1 & 0 & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & 0\end{pmatrix}.$$

The eigenvalues of $J_i^{(\varepsilon)}+(J_i^{(\varepsilon)})^H$ are

$$\lambda_j^{(i)}=2\,\mathrm{Re}\,\lambda_i+2\varepsilon\cos\frac{j\pi}{\nu_i+1}, \quad j=1, 2, \cdots, \nu_i.$$

If we set $0<\varepsilon<\min\limits_{1\leqslant i\leqslant n}|\mathrm{Re}\,\lambda_i(A)|$, then $\lambda_j^{(i)}<0$, $j=1, \cdots, \nu_i$, $i=1, \cdots, k$. By Lemma 2.1, we have

$$\mathrm{sep}_2(J_A^{(\varepsilon)})\geqslant|\lambda_n(J_A^{(\varepsilon)}+(J_A^{(\varepsilon)})^H)|\geqslant2(\min_{1\leqslant i\leqslant n}|\mathrm{Re}\,\lambda_i(A)|-\varepsilon).$$

If

$$\min_{1\leqslant i\leqslant n}|\mathrm{Re}\,\lambda_i(A)|\leqslant\frac{2\nu-1}{2\nu-2},$$

set

$$\varepsilon_0=\frac{2\nu-2}{2\nu-1}\min_{1\leqslant i\leqslant n}|\mathrm{Re}\,\lambda_i(A)|.$$

Then $0<\varepsilon_0\leqslant1$, and $\dfrac{\min\limits_{1\leqslant i\leqslant n}|\mathrm{Re}\,\lambda_i(A)|-\varepsilon}{1/\varepsilon^{2(\nu-1)}}$ attains its extremum at ε_0. Substituting $\varepsilon=\varepsilon_0$ into (2.8), we have

$$\mathrm{sep}_2(J_A)\geqslant\frac{1}{\nu-1}\left(\frac{2\nu-2}{2\nu-1}\min_{1\leqslant i\leqslant n}|\mathrm{Re}\,\lambda_i(A)|\right)^{2\nu-1}. \tag{2.10}$$

If $\min\limits_{1\leqslant i\leqslant n}|\mathrm{Re}\,\lambda_i(A)|>1$, set $\varepsilon_0=1$, We have

$$\mathrm{sep}_2(J_A)\geqslant2\min_{1\leqslant i\leqslant n}(|\mathrm{Re}\,\lambda_i(A)|-1). \tag{2.11}$$

Combining (2.5) with (2.10) and (2.11), we get (2.4) at once. ∎

By Theorem 2.1 we know that the requirement that the separation of a stable matrix A be relatively large is equivalent to the requirement that $\min\limits_{1\leqslant i\leqslant n}\mathrm{Re}|\lambda_i(A)|$ be relatively large, i.e., the distance between the spectrum of A and the imaginary axis be relatively large.

§ 3. Perturbation Theorem

Before describing the perturbation theorem, we introduce the concept of unitary symplectic matrix and the singular value decomposition theorem for a unitary symplectic matrix.

Definition 3.1. *$S\in\mathbb{C}^{2n\times 2n}$ is said to be symplectic if $J^{-1}SJ=S^{-H}$, where* $J=\begin{pmatrix}0 & I_n\\ -I_n & 0\end{pmatrix}$.

Definition 3.2. *$Q\in\mathbb{C}^{2n\times 2n}$ is called a unitary symplectic matrix if Q is unitary and symplectic.*

The following facts are well known (see [4]). Any unitary symplectic matrix $Q\in\mathbb{C}^{2n\times 2n}$ can be written in the form $Q=\begin{pmatrix}Q_{11} & Q_{12}\\ -Q_{12} & Q_{11}\end{pmatrix}$, where Q_{11}, $Q_{12}\in\mathbb{C}^{n\times n}$ satisfy $Q_{11}Q_{11}^H+Q_{12}Q_{12}^H=I_n$ and $Q_{11}Q_{12}^H=Q_{12}Q_{11}^H$. Let $M\in\mathbb{C}^{2n\times 2n}$ be Hamiltonian whose eigenvalues have nonzero real part. Then there exists a unitary symplectic matrix Q such that

$$Q^HMQ=\begin{pmatrix}T & R\\ 0 & -T^H\end{pmatrix}, \tag{3.1}$$

where T, $R\in\mathbb{C}^{n\times n}$ and T is stable.

The singular value decomposition theorem for a unitary symplectic matrix can be described as

Theorem 3.1[4]. *Let $Q\in\mathbb{C}^{2n\times 2n}$ be a unitary symplectic matrix. Then there exist two unitary matrices U and V in $\mathbb{C}^{n\times n}$ such that*

$$\mathrm{diag}(U^H,\ U^H)Q\,\mathrm{diag}(V,\ V)=\begin{pmatrix}\Sigma & \Delta\\ -\Delta & \Sigma\end{pmatrix}, \tag{3.2}$$

where

$$\begin{aligned}&\Sigma=\mathrm{diag}(\sigma_1,\ \sigma_2,\ \cdots,\ \sigma_n),\ 0\leqslant\sigma_1\leqslant\sigma_2\leqslant\cdots\leqslant\sigma_n\leqslant 1,\\ &\Delta=\mathrm{diag}(\delta_1,\ \delta_2,\ \cdots,\ \delta_n),\ \delta_i=\pm(1-\sigma_i^2)^{1/2},\quad i=1,\ 2,\cdots,\ n.\end{aligned} \tag{3.3}$$

Applying Theorem 4.11 in [6] to a Hamiltonian matrix and considering the properties of Hamiltonian matrices, we can get the following perturbation theorem of a Hamiltonian matrix.

Theorem 3.2. *Let M, $\delta M\in\mathbb{C}^{2n\times 2n}$ be Hamiltonian with the eigenvalues of M having nonzero real part. The unitary symplectic matrix Q, $n\times n$ matrices T and R are defined by* (3.1). *Let $Q^H\delta MQ$ be partitioned conformally with* (3.1) *in the form*

$$Q^H\delta MQ=\begin{pmatrix}E_{11} & E_{21}\\ E_{21} & -E_{11}^H\end{pmatrix}. \tag{3.4}$$

Let

$$\delta=\mathrm{sep}(T)-(\|E_{11}\|+\|E_{11}^H\|). \tag{3.5}$$

If

$$\frac{\|E_{21}\|(\|R\|+\|E_{12}\|)}{\delta^2}\leqslant\frac{1}{4}, \tag{3.6}$$

there is a Hermitian matrix $P\in\mathbb{C}^{n\times n}$ satisfying

$$\|P\| < \frac{2\|E_{21}\|}{\delta} \tag{3.7}$$

such that

$$\tilde{Q}^H(M+\delta M)\tilde{Q} = \begin{pmatrix} \tilde{T} & \tilde{R} \\ 0 & -\tilde{T}^H \end{pmatrix}, \tag{3.8}$$

where $\tilde{Q} = Q\begin{pmatrix} I_n & -P \\ P & I_n \end{pmatrix}\begin{pmatrix} (I_n+P^2)^{-\frac{1}{2}} & 0 \\ 0 & (I_n+P^2)^{-\frac{1}{2}} \end{pmatrix}$ *is a unitary symplectic matrix.*

In Theorem 3.2, $\tilde{T} = (I_n+P^2)^{\frac{1}{2}}(T+E_{11}+(R+E_{12})P)(I_n+P^2)^{-\frac{1}{2}}$. We have that $\tilde{T}$ is similar to $T+E_{11}+(R+E_{12})P$. Now we discuss that under what conditions $\tilde{T}$ is stable.

Let $X_T \in \mathbb{C}^{n\times n}$ be nonsingular such that

$$X_T^{-1}TX_T = J_T, \tag{3.9}$$

where J_T is the Jordan canonical form. The orders of Jordan blocks in J_T are m_1, m_2, $\cdots$, m_k, respectively. Let

$$m = \max_{1\le i\le k}\{m_i\}. \tag{3.10}$$

By Theorem 8 in [2] any $\lambda \in \lambda(\tilde{T})$ there corresponds to a $\mu \in \lambda(T)$ such that

$$\frac{|\lambda-\mu|^m}{(1+|\lambda-\mu|)^{m-1}} \le \varkappa_2(X_T)\|E_{11}+(R+E_{12})P\|_2.$$

If

$$\|E_{11}+(R+E_{12})P\|_2 < \frac{1}{\varkappa_2(X_T)}\frac{\min_{1\le i\le n}|\mathrm{Re}\,\lambda_i(T)|^m}{(1+\min_{1\le i\le n}|\mathrm{Re}\,\lambda_i(T)|)^{m-1}}, \tag{3.11}$$

then

$$\frac{|\lambda-\mu|^m}{(1+|\lambda-\mu|)^{m-1}} < \frac{\min_{1\le i\le n}|\mathrm{Re}\,\lambda_i(T)|^m}{(1+\min_{1\le i\le n}|\mathrm{Re}\,\lambda_i(T)|)^{m-1}}.$$

Because $\frac{x^m}{(1+x)^{m-1}}$ is strictly increasing in $x \ge 0$, we have $|\lambda-\mu| < \min_{1\le i\le n}|\mathrm{Re}\,\lambda_i(T)|$. It is easy to see that $\mathrm{Re}\,\lambda < 0$, $\lambda \in \lambda(\tilde{T})$, i.e., $\tilde{T}$ is stable.

Now we consider the perturbed equation

$$\begin{aligned} &(A+\delta A)^H(X+\delta X)+(X+\delta X)(A+\delta A) \\ &\quad -(X+\delta X)(N+\delta N)(X+\delta X)+(K+\delta K) \\ &=0, \end{aligned} \tag{3.12}$$

where $(N+\delta N)^H = N+\delta N \ge 0$, $(K+\delta K)^H = K+\delta K \ge 0$. By Theorem 3.2, we can get the perturbation theorem of the positive semi-definite solution of the algebraic Riccati equation (1.1).

Theorem 3.3. *Consider the algebraic Riccati equation* (1.1) *and the perturbed equation* (3.12). *Assume that* (A, B) *is stabilizable and* (C, A) *is detectable. The Hamiltonian matrix* M*, unitary symplectic matrix* Q*,* $n\times n$ *matrices* T*,* R *and* X_T *are defined by* (1.2), (3.1) *and* (3.9), *respectively.* σ_1 *and* m *are defined by* (3.2), (3.3) *and* (3.10), *respectively. Let*

$$\Delta_2 = \|\delta A\|_2 + \max\{\|\delta N\|_2,\ \|\delta K\|_2\} \tag{3.13}$$

and

$$\tilde{\delta} = \mathrm{sep}_2(T) - 2\Delta_2. \tag{3.14}$$

If

$$\Delta_2 \leqslant \min\left\{\frac{1}{4}\,\mathrm{sep}_2(T),\ \frac{\mathrm{sep}_2^2(T)}{4(\|R\|_2+\mathrm{sep}_2(T))},\ \frac{\sigma_1\,\mathrm{sep}_2(T)}{2(1+\sigma_1)},\right.$$
$$\left.\frac{\mathrm{sep}_2(T)}{\varkappa_2(X_T)(4\|R\|_2+\mathrm{sep}_2(T))}\ \frac{\min\limits_{1\leqslant i\leqslant n}|\mathrm{Re}\,\lambda_i(T)|^m}{(1+\min\limits_{1\leqslant i\leqslant n}|\mathrm{Re}\,\lambda_i(T)|)^{m-1}}\right\}, \tag{3.15}$$

we have

$$\frac{\|\delta X\|_2}{\|X\|_2} < \frac{1}{\sqrt{1-\sigma_1^2}}\ \frac{\dfrac{2\sqrt{2}\,\Delta_2}{\tilde{\delta}}}{\sigma_1\sqrt{1+\dfrac{4\Delta_2^2}{\tilde{\delta}^2}}+\sqrt{1+\dfrac{4\Delta_2^2}{\tilde{\delta}^2}}-\dfrac{2\sqrt{2}\,\Delta_2}{\tilde{\delta}}}. \tag{3.16}$$

In addition, if $\Delta_2 \leqslant \dfrac{\sigma_1\,\mathrm{sep}_2(T)}{2(2+\sigma_1)}$, *then*

$$\frac{\|\delta X\|_2}{\|X\|_2} < \frac{1}{\sqrt{1-\sigma_1^2}}\ \frac{8\Delta_2}{\sigma_1\,\mathrm{sep}_2(T)}. \tag{3.17}$$

Proof. Let $\delta M = \begin{pmatrix} \delta A & \delta N \\ \delta K & -(\delta A)^H \end{pmatrix}$. Then the E_{ij} defined by (3.4) satisfies

$$\|E_{ij}\|_2 \leqslant \|\delta M\|_2 \leqslant \left\|\begin{pmatrix} \delta A & 0 \\ 0 & -(\delta A)^H \end{pmatrix}\right\|_2 + \left\|\begin{pmatrix} 0 & \delta N \\ \delta K & 0 \end{pmatrix}\right\|_2$$
$$= \|\delta A\|_2 + \max\{\|\delta N\|_2,\ \|\delta K\|_2\} = \Delta_2.$$

By (3.15) we have $\Delta_2 \leqslant \dfrac{\mathrm{sep}_2^2(T)}{4(\|R\|_2+\mathrm{sep}_2(T))}$; then $\dfrac{\Delta_2(\|R\|_2+\Delta_2)}{\tilde{\delta}^2} \leqslant \dfrac{1}{4}$. Thus condition (3.6) in Theorem 3.2 is satisfied. Then there is a Hermitian matrix $P \in \mathbb{C}^{n\times n}$ satisfying $\|P\|_2 < \dfrac{2\Delta_2}{\tilde{\delta}}$ such that (3.8) is true. By (3.15) we also have

$$\|E_{11} + (R+E_{12})P\|_2 \leqslant \|E_{11}\|_2 + (\|R\|_2 + \|E_{12}\|_2)\|P\|_2 < \Delta_2 + (\|R\|_2+\Delta_2)\frac{2\Delta_2}{\tilde{\delta}}$$
$$= \Delta_2\frac{\mathrm{sep}_2(T)+2\|R\|_2}{\mathrm{sep}_2(T)-2\Delta_2} \leqslant \Delta_2\,\frac{2\,\mathrm{sep}_2(T)+4\|R\|_2}{\mathrm{sep}_2(T)}$$
$$\leqslant \frac{1}{\varkappa_2(X_T)}\cdot\frac{\min\limits_{1\leqslant i\leqslant n}|\mathrm{Re}\,\lambda_i(T)|^m}{(1+\min\limits_{1\leqslant i\leqslant n}|\mathrm{Re}\,\lambda_i(T)|)^{m-1}}.$$

Thus (3.11) is satisfied. Then T is stable.

Let $Q = \begin{pmatrix} Q_{11} & Q_{12} \\ -Q_{12} & Q_{11} \end{pmatrix}$ and $\tilde{Q} = \begin{pmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} \\ -\tilde{Q}_{12} & \tilde{Q}_{11} \end{pmatrix}$. The positive semi-definite solutions of (1.1) and (3.12) are given by $X = Q_{12}Q_{11}^{-1}$ and $X+\delta X = \tilde{Q}_{12}\tilde{Q}_{11}^{-1}$, respectively. Let $\begin{pmatrix} Q'_{11} \\ -Q'_{12} \end{pmatrix} = \begin{pmatrix} \tilde{Q}_{11} \\ -\tilde{Q}_{12} \end{pmatrix} - \begin{pmatrix} Q_{11} \\ -Q_{12} \end{pmatrix}$. We have

$$\delta X=\widetilde{Q}_{12}\widetilde{Q}_{11}^{-1}-Q_{12}Q_{11}^{-1}=\widetilde{Q}_{11}^{-H}\widetilde{Q}_{12}^{H}-Q_{12}Q_{11}^{-1}=\widetilde{Q}_{H}^{-H}(\widetilde{Q}_{12}^{H}Q_{11}-\widetilde{Q}_{11}^{H}Q_{12})Q_{11}^{-1}$$
$$=\widetilde{Q}_{11}^{-H}(Q_{12}^{H}(\widetilde{Q}_{11}-Q_{11}')-\widetilde{Q}_{11}^{H}(\widetilde{Q}_{12}-Q_{11}'))Q_{11}^{-1}=\widetilde{Q}_{11}^{-H}(\widetilde{Q}_{11}^{H}Q_{12}'-Q_{12}^{H}Q_{11}')Q_{11}^{-1}$$
$$=\widetilde{Q}_{11}^{-H}(\widetilde{Q}_{11}^{H},\ \widetilde{Q}_{12}^{H})\begin{pmatrix}Q_{12}'\\-Q_{11}'\end{pmatrix}Q_{11}^{-1}=(I_n, X+\delta X)\begin{pmatrix}0&-I_n\\-I_n&0\end{pmatrix}\begin{pmatrix}Q_{11}'\\-Q_{12}'\end{pmatrix}Q_{11}^{-1}$$

and

$$\|\delta X\|_2=\left\|((I,\ X)+(0,\ \delta X))\begin{pmatrix}0&-I_n\\-I_n&0\end{pmatrix}\begin{pmatrix}Q_{11}'\\-Q_{12}'\end{pmatrix}Q_{11}^{-1}\right\|_2$$
$$\leqslant(\sqrt{1+\|X\|_2^2}+\|\delta X\|_2)\left\|\begin{pmatrix}Q_{11}'\\-Q_{12}'\end{pmatrix}\right\|_2\|Q_{11}^{-1}\|_2. \tag{3.18}$$

From (3.2) and (3.3), we have $Q=U\operatorname{diag}(\sigma_1,\ \cdots,\ \sigma_n)V^H$ and

$$X=U\operatorname{diag}\left(\frac{\delta_1}{\sigma_1},\ \cdots,\ \frac{\delta_n}{\sigma_n}\right)U^H.$$

Then

$$\|Q_{11}^{-1}\|_2=\frac{1}{\sigma_1}\quad\text{and}\quad\|X\|_2=\frac{\sqrt{1-\sigma_1^2}}{\sigma_1}.$$

Combining with (3.18), we obtain that if $\left\|\begin{pmatrix}Q_{11}'\\-Q_{12}'\end{pmatrix}\right\|_2<\sigma_1$, then

$$\frac{\|\delta X\|_2}{\|X\|_2}\leqslant\frac{1}{\sqrt{1-\sigma_1^2}}\frac{\left\|\begin{pmatrix}Q_{11}'\\-Q_{12}'\end{pmatrix}\right\|_2}{\sigma_1-\left\|\begin{pmatrix}Q_{11}'\\-Q_{12}'\end{pmatrix}\right\|_2}. \tag{3.19}$$

Now we estimate $\left\|\begin{pmatrix}Q_{11}'\\-Q_{12}'\end{pmatrix}\right\|_2$. Utilizing Theorem 3.2 we get

$$\left\|\begin{pmatrix}Q_{11}'\\-Q_{12}'\end{pmatrix}\right\|_2=\left\|\begin{pmatrix}\widetilde{Q}_{11}\\-\widetilde{Q}_{12}\end{pmatrix}-\begin{pmatrix}Q_{11}\\-Q_{12}\end{pmatrix}\right\|_2=\left\|Q\begin{pmatrix}I_n\\P\end{pmatrix}(I_n+P^2)^{-\frac{1}{2}}-Q\begin{pmatrix}I_n\\0\end{pmatrix}\right\|_2$$
$$=\left\|Q\begin{pmatrix}-I_n&0\\0&I_n\end{pmatrix}\begin{pmatrix}(I_n+P_2)^{\frac{1}{2}}-I_n\\P\end{pmatrix}(I_n+P^2)^{-\frac{1}{2}}\right\|_2$$
$$\leqslant\left\|\begin{pmatrix}(I_n+P^2)^{\frac{1}{2}}-I_n\\P\end{pmatrix}(I_n+P^2)^{-\frac{1}{2}}\right\|_2.$$

Because P is Hermitian we have the decomposition $P=U_P\Lambda_PU_P^H$, where U_P is unitary and $\Lambda_P=\operatorname{diag}(\lambda_1(P),\ \cdots,\ \lambda_n(P))\in\mathbb{R}^{n\times n}$. Substituting $P=U_P\Lambda_PU_P^H$ into the inequality described above, we get

$$\left\|\begin{pmatrix}Q_{11}'\\-Q_{12}'\end{pmatrix}\right\|_2\leqslant\left\|\begin{pmatrix}(I_n+\Lambda_p^2)^{\frac{1}{2}}-I_r\\\Lambda_P\end{pmatrix}(I_n+\Lambda_p^2)^{\frac{1}{2}}\right\|_2$$

$$=\left\|\begin{pmatrix} \frac{\sqrt{1+\lambda_1^2(P)}-1}{\sqrt{1+\lambda_1^2(P)}} & & 0 \\ 0 & \ddots & \frac{\sqrt{1+\lambda_n^2(P)}-1}{\sqrt{1+\lambda_n^2(P)}} \\ \frac{\lambda_1(P)}{\sqrt{1+\lambda_1^2(P)}} & & 0 \\ 0 & \ddots & \frac{\lambda_n(P)}{\sqrt{1+\lambda_n^2(P)}} \end{pmatrix}\right\|_2$$

$$\leqslant \max_{1<j<n}\left[\left(\frac{\sqrt{1+\lambda_j^2(P)}-1}{\sqrt{1+\lambda_j^2(P)}}\right)^2+\left(\frac{\lambda_j(P)}{\sqrt{1+\lambda_j^2(P)}}\right)^2\right]^{1/2}$$

$$=\max_{1<j<n}\sqrt{2}\sqrt{1-\frac{1}{\sqrt{1+\lambda_j^2(P)}}}=\sqrt{2}\sqrt{1-\frac{1}{\sqrt{1+\|P\|_2}}}$$

$$=\frac{\sqrt{2}\,\|P\|_2}{\sqrt{1+\|P\|_2^2+\sqrt{1+\|P\|_2^2}}}. \tag{3.20}$$

From (3.15), we have $\Delta_2\leqslant\frac{\sigma_1\operatorname{sep}_2(T)}{2(1+\sigma_1)}$. Then

$$\left\|\begin{pmatrix} Q'_{11} \\ -Q'_{12}\end{pmatrix}\right\|_2\leqslant\frac{\sqrt{2}\,\|P\|_2}{\sqrt{1+\|P\|_2^2+\sqrt{1+\|P\|_2^2}}}\leqslant\|P\|_2<\frac{2\Delta_2}{\tilde{\delta}}\leqslant\sigma_1.$$

Substituting (3.20) into (3.19), we get

$$\frac{\|\delta X\|_2}{\|X\|_2}\leqslant\frac{1}{\sqrt{1-\sigma_1^2}}\cdot\frac{\sqrt{2}\,\|P\|_2}{\sigma_1\sqrt{1+\|P\|_2^2+\sqrt{1+\|P\|_2^2}}-\sqrt{2}\,\|P\|_2}.$$

Combining with

$$\|P\|_2<\frac{2\Delta_2}{\tilde{\delta}},$$

we get (3.16) at once.

If

$$\Delta_2\leqslant\frac{\sigma_1\operatorname{sep}_2(T)}{2(2+\sigma_1)},$$

then

$$\frac{2\Delta_2}{\operatorname{sep}_2(T)-2\Delta_2}\leqslant\frac{1}{2}\sigma_1.$$

Combining (3.19) with

$$\left\|\begin{pmatrix} Q'_{11} \\ -Q'_{12}\end{pmatrix}\right\|_2<\frac{2\Delta_2}{\tilde{\delta}},$$

we have

$$\frac{\|\delta X\|_2}{\|X\|_2}\leqslant\frac{1}{\sqrt{1-\sigma_1^2}}\,\frac{\frac{2\Delta_2}{\tilde{\delta}}}{\sigma_1-\frac{2\Delta_2}{\tilde{\delta}}}=\frac{1}{\sqrt{1-\sigma_1^2}}\,\frac{\frac{2\Delta_2}{\operatorname{sep}_2(T)-2\Delta_2}}{\sigma_1-\frac{2\Delta_2}{\operatorname{sep}_2(T)-2\Delta_2}}$$

$$<\frac{1}{\sqrt{1-\sigma_1^2}}\cdot\frac{\dfrac{2\Delta_2}{\frac{1}{2}\operatorname{sep}_2(T)}}{\frac{1}{2}\sigma_1}=\frac{1}{\sqrt{1-\sigma_1^2}}\cdot\frac{8\Delta_2}{\sigma_1\operatorname{sep}_2(T)}.$$

Thus (3.17) is true. ∎

By Theorem 3.3 we know that the perturbation property of the solution of the algebraic Riccati equation (1.1) is closely related to σ_1 and $\operatorname{sep}_2(T)$. Generally speaking, when $\sigma_1 \operatorname{sep}_2(T)$ is relatively large, the solution of (1.1) is insensitive to perturbations in the data. That coinsides with the conclusions in [3], [4] and [7]. When σ_1 is near to one, δX is relatively large with respect to X, and $\|X\|_2$ is near to zero.

§ 4. The Relation Between $\mu(A, B)$ and σ_1

Referring to the distance between a controllable system and an uncontrollable one[7], we introduce the following definition.

Definition 4.1. *Let $A\in\mathbb{C}^{n\times n}$, $B\in\mathbb{C}^{n\times m}$ with (A, B) stabilizable. Define the distance between (A, B) and a nearest unstabilizable pair by*

$$\mu(A, B)=\min_{\delta A,\delta B}\|(\delta A, \delta B)\|_2, \tag{4.1}$$

where $\delta A\in\mathbb{C}^{n\times n}$, $\delta B\in\mathbb{C}^{n\times m}$ such that $(A+\delta A, B+\delta B)$ is unstabilizable.

We can obtain an equivalent definition from the following theorem.

Theorem 4.1. *Let (A, B) be stabilizable. Then*

$$\mu(A, B)=\min_{\operatorname{Re} s\geq 0}\sigma_n(sI_n-A, B), \tag{4.2}$$

where $\sigma_n(sI_n-A, B)$ is the smallest singular value of (sI_n-A, B).

Proof. Let $P(s)=(sI_n-A, B)$. That (A, B) is stabilizable is equivalent to that rank $P(s)=n$ for any $s\in\lambda(A)$ and Re $s\geq 0$, and to that rank $P(s)=n$ for any $s\in\mathbb{C}$ and Re $s\geq 0$.

If $\delta A\in\mathbb{C}^{n\times n}$ and $\delta B\in\mathbb{C}^{n\times m}$ such that $(A+\delta A, B+\delta B)$ is unstabilizable, there exists an $s\in\mathbb{C}$ which satisfies Re$s\geq 0$ such that rank $(sI_n-(A+\delta A), B+\delta B)<n$, i.e., $\sigma_n(sI_n-(A+\delta A), B+\delta B)=0$. Observing that $(sI_n-(A+\delta A), B+\delta B)=(sI_n+A, B)+(-\delta A, \delta B)$, by the perturbation theorem of singular values we can get

$$|\sigma_n(sI_n-(A+\delta A), B+\delta B)-\sigma_n(sI_n-A, B)|\leq\|(-\delta A, \delta B)\|_2=\|(\delta A, \delta B)\|_2.$$

Thus $\|(\delta A, \delta B)\|_2\geq\sigma_n(sI_n-A, B)\geq\min_{\operatorname{Re} s\geq 0}\sigma_n(sI_n-A, B)$. Immediately we get

$$\mu(A, B)\geq\min_{\operatorname{Re} s\geq 0}\sigma_n(sI_n-A, B). \tag{4.3}$$

It can be assumed that $\min_{\operatorname{Re} s>0}\sigma_n(sI_n-A, B)$ is attained at $s_0\in\mathbb{C}$, Re $s_0\geq 0$. Applying the singular value decomposition theorem to (s_0I_n-A, B), we have

$$(s_0I_n-A, B)=U(\Sigma, 0)V^H,\ \Sigma=\operatorname{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n).$$

Set $(-\delta A, \delta B)=U(\delta\Sigma, 0)V^H$, where $\delta\Sigma=\operatorname{diag}(0,\cdots, 0, -\sigma_n)$. Then rank $(s_0I_n-(A+\delta A), B+\delta B)<n$, i.e., $(A+\delta A, B+\delta B)$ is unstabilizable. Observing that

$$\|(\delta A,\ \delta B)\|_2=\|(-\delta A,\ \delta B)\|_2=\|U(\Sigma,\ 0)V^H\|_2=\sigma_n(s_0I_n-A,\ B)$$
$$=\min_{\mathrm{Re}\,s\geq 0}\sigma\ (sI_n-A,\ B),$$

we have

$$\mu(A,\ B)\leqslant\min_{\mathrm{Re}\,s\geq 0}\sigma_n(sI_n-A,\ B). \tag{4.4}$$

Combining (4.3) with (4.4), we get (4.2). ∎

Now we come back to the discussion on algebraic Riccati equations. We know that σ_1 and $\mathrm{sep}_2(T)$ determine the perturbation property of the solution of (1.1). The quantity of $\min\limits_{1\leqslant i\leqslant n}|\mathrm{Re}\,\lambda_i(T)|$ determines the quantity of $\mathrm{sep}_2(T)$. From the proof of Theorem 4.1 in [4], it is known that if $(A,\ B)$ is stabilizable, then $\sigma_1>0$. Now we demonstrate that when $\mathrm{sep}_2(T)$ is not very small, if $(A,\ B)$ is very near to an unstabilizale system, then σ_1 is very small provided that $\|B\|_2$ is not very large.

Theorem 4.2. *Consider the algebraic Riccati equation.* (1.1). *Assume that* $(A,\ B)$ *is stabilizable and* $(C,\ A)$ *is detectable. The Hamiltonian matrix* M, *unitary symplectic matrix* Q, $n\times n$ *matrices* T, R *and* X_T *are defined by* (1.2), (3.1) *and* (3.9), *respectively.* σ_1 *and* m *are defined by* (3.2), (3.3) *and* (3.10), *respectively. Let*

$$b=\|B\|_2+\frac{1}{2}+\sqrt{\left(\|B\|_2+\frac{1}{2}\right)^2+1}. \tag{4.5}$$

If

$$\mu(A,\ B)\leqslant\frac{1}{b}\min\Bigg\{1,\ \frac{1}{4}\,\mathrm{sep}_2(T),\frac{\mathrm{sep}_2^2(T)}{4(\|R\|_2+\mathrm{sep}_2(T))},\ \frac{\mathrm{sep}_2(T)}{\varkappa_2(X_T)(4\|R\|_2+2\,\mathrm{sep}_2(T))}$$
$$\cdot\frac{\min\limits_{1\leqslant i\leqslant n}|\mathrm{Re}\,\lambda_i(T)|^m}{(1+\min\limits_{1\leqslant i\leqslant n}|\mathrm{Re}\,\lambda_i(T)|)^{m-1}}\Bigg\}, \tag{4.6}$$

we have

$$\sigma_1<\frac{4b}{\mathrm{sep}_2(T)}\ \mu(A,\ B). \tag{4.7}$$

Proof. From the definition of $\mu(A,\ B)$, there exist $\delta A\in\mathbb{C}^{n\times n}$ and $\delta B\in\mathbb{C}^{n\times m}$ satisfying $\|(\delta A,\ \delta B)\|_2=\mu(A,\ B)$ such that $(A+\delta A,\ B+\delta B)$ is unstabilizable. Let

$$\delta M=\begin{pmatrix}\delta A & (\delta B)B^H+B(\delta B)^H+(\delta B)(\delta B)^H\\ 0 & -(\delta A)^H\end{pmatrix}.$$

Then the E_{ij} defined by (3.4) satisfies

$$\|E_{ij}\|_2\leqslant\|\delta M\|_2\leqslant\|\delta A\|_2+(2\|B\|_2+\|\delta B\|_2)\|\delta B\|_2\equiv\Delta_2.$$

By (4.6) we have

$$\Delta_2=\|\delta A\|_2+(2\|B\|_2+\|\delta B\|_2)\|\delta B\|_2\leqslant(1+2\|B\|_2+\|\delta B\|_2)\cdot\max\{\|\delta A\|_2,\ \|\delta B\|_2\}$$
$$\leqslant(1+2\|B\|_2+\mu(A,\ B))\mu(A,\ B)\leqslant\left(1+2\|B\|_2+\frac{1}{b}\right)\mu(A,\ B)\leqslant b\mu(A,\ B)$$
$$<\frac{\mathrm{sep}_2^2(T)}{4(\|R\|_2+\mathrm{sep}_2(T))}.$$

In a manner similar to the proof of Theorem 3.3, we know that there is a Hermitian matrix $P\in\mathbb{C}^{n\times n}$ satisfying

$$\|P\|_2<\frac{2\Delta_2}{\operatorname{sep}_2(T)-2\Delta_2}$$

such that (3.8) is true. It can also be proved that $\tilde{T}$ is stable.

Let

$$Q=\begin{pmatrix} Q_{11} & Q_{12} \\ -Q_{12} & Q_{11} \end{pmatrix}, \quad \tilde{Q}=\begin{pmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} \\ -\tilde{Q}_{12} & \tilde{Q}_{11} \end{pmatrix}.$$

We can prove that $\tilde{Q}_{11}$ is singular. In fact, if $\tilde{Q}_{11}$ is nonsingular, it follows from (3.8) that $(A+\delta A)-(B+\delta B)(B+\delta B)^H\tilde{Q}_{12}\tilde{Q}_{11}^{-1}=\tilde{Q}_{11}\tilde{T}\tilde{Q}_{11}^{-1}$. Thus $(A+\delta A,\ B+\delta B)$ is stabilizable. That contradicts the selection of δA and δB. So we have

$$\sigma_1=\min_{\det(Q_{11}+E)=0}\|E\|_2\leqslant\|\tilde{Q}_{11}-Q_{11}\|_2\leqslant\left\|\begin{pmatrix} \tilde{Q}_{11}-Q_{11} \\ -(\tilde{Q}_{12}-Q_{12}) \end{pmatrix}\right\|_2.$$

From the proof of Theorem 3.3 and (4.6), we get

$$\sigma_1\leqslant\left\|\begin{pmatrix} Q'_{11} \\ -Q'_{12} \end{pmatrix}\right\|_2\leqslant\|P\|_2<\frac{2\Delta_2}{\operatorname{sep}_2(T)-2\Delta_2}\leqslant\frac{2b\mu(A,\ B)}{\frac{1}{2}\operatorname{sep}_2(T)}=\frac{4b}{\operatorname{sep}_2(T)}\ \mu(A,\ B).$$

Thus (4.7) is valid.

§ 5. Concluding Remarks

In summary, the sensitivity of the positive semi-definite solution of the algebraic Riccati equation (1.1) to perturbations in data is determined by σ_1 and $\operatorname{sep}_2(T)$. When $\sigma_1\operatorname{sep}_2(T)$ is relatively large, the solution of (1.1) is insensitive to perturbations in data. That $\operatorname{sep}_2(T)$ is relatively large is equivalent to that $\min_{1\leqslant i\leqslant n}|\operatorname{Re}\lambda_i(T)|$ is relatively large, i.e., the distance between the spectrum of the matrix T (or M) and the imaginary axis is relatively large. When $\operatorname{sep}_2(T)$ is not very small, if $(A,\ B)$ is very near to an unstabilizable system, then σ_1 is very small provided that $\|B\|_2$ is not vary large.

Acknowledgment. The author expresses her sincere thanks to Prof. Sun Ji-guang under whose direction this paper was finished.

References

[1] Boley, D. L.; Lu Wu-sheng: Measuring how far a controllable system is from an uncontrollable one, *IEEE Trans. On Auto. Contr.*, AC—**31**:3 (1986), 249—251.

[2] Kahan, W. M.; Parlett, B. N.; Jiang Er-Xiong: Residual bounds on approximate eigensystems of nonnormal matrices, *SIAM J. Numer. Anal.*, **19** (1982), 470—484.

[3] Laub, J.: A Schur method for solving algebraic Riccati equations, *IEEE Trans. on Auto. Contr.*, AC-**24**: 6 (1979), 913—921.

[4] Paige, C; Van Loan, C. A Schur decomposition for Hamiltonian matrices, *LAA*, **41** (1981), 11—32.

[5] Petkov, P. Hr.; Konstantinov, M. M.; Christov, N. D.; Computional algorithms for linear control systems: a brief survey, *Int. J. Systems Sci.*, **16**: 4 (1985), 465—477.

[6] Stewart, G. W.: Error and perturbation bounds for subspaces associated with certain eigenvalue problems, *SIAM Review*, **15**:4 (1973), 727—764.

[7] Chu Da-xun: On computation for a class of symplectic-invariant matrices and solutions of algebraic Riccati equations, Fudan University, 1986.