ON THE SENSITIVITY ESTIMATION OF THE SYMMETRIC MATRIX RICCATI DIFFERENTIAL EQUATION

Vera Angelova, Todor Balabanov, Ivan Popchev

Abstract

In this paper, we consider three already known perturbation estimates of the solution to the symmetric matrix Riccati differential equation. The aim of the paper is to analyse experimentally the effectiveness of the bounds considered, to compare their sharpness for problems with increasing conditioning, and to specify the area of application of the bounds analysed. The analytical solution to the scalar Riccati differential equation of one of the experimental models is proved in theorem. The results suggest a new field of research.

Key words: symmetric differential matrix Riccati equation, perturbation bounds, control theory

2020 Mathematics Subject Classification: 15A24, 37N35

1. Introduction. The initial value problem for a symmetric matrix Riccati differential equation (RDE) is

\[ \dot{X}(t) = A^T X(t) + X(t)A + B - X(t)C X(t), \quad X(0) = X_0, \]

on the time interval \( T = [0, t_1], \ t_1 > 0, \) with coefficient matrices \( A, B, C \in \mathbb{R}^{n \times n}, \) and a solution \( X(t) \in \mathbb{R}^{n \times n}. \) The coefficient matrices \( C, B \) and the initial condition \( X_0 \) are symmetric and nonnegative. This problem arises in optimal control of linear systems, optimal filtering, game theory, \( H_\infty \) control of linear time-varying systems, etc. A large variety of approaches to compute the solution to the RDE are developed (see [1] and the literature therein). Linear [2-3] and
nonlinear\[^4,5\] perturbation bounds are derived. The experimental analysis in\[^5,6\]\ of the effectiveness in terms of closeness to the estimated value and size of the domain of validity, based on a reference model, shows that with the deterioration of the conditioning of the matrix $A$, the nonlinear bound of Konstantinov and Pelova\[^4\] and the nonlinear bound of Konstantinov and Angelova\[^5\] are superior to the linear bound of Kenney and Hewer\[^2\] and are alternative. It is interesting to compare the effectiveness of the bounds of Konstantinov and Pelova\[^4\] and of Konstantinov and Angelova\[^5\] – winners of the comparative analysis in\[^5,6\]\, to the bound, proposed by Weng and Phoa in\[^3\]\, which demonstrates an impressive sharpness.

Throughout the paper we use the notations: $\mathbb{R}^{m \times n}$ for the space of $m \times n$ real matrices; $I_n$ is the identity $n \times n$ matrix; $A^\top$ is the transpose of the matrix $A \in \mathbb{R}^{m \times n}$; $\|\cdot\|$ is the the Euclidean vector or the spectral matrix norm $\|A\| = (\lambda_{\text{max}}(A^\top A))^{1/2}$, where $\lambda_{\text{max}}(B)$ is the maximum eigenvalue of the symmetric matrix $B$; $\|\cdot\|_F$ is the Frobenius norm; $\text{vec}(A) = [a_1^\top \ a_2^\top \ \ldots \ a_n^\top]^\top \in \mathbb{C}^{n^2}$ is the column-wise vector representation of the matrix $A = [a_1, a_2, \ldots, a_n] \in \mathbb{C}^{n \times n}$, $a_j \in \mathbb{C}^n$, where $\mathbb{C}^n = \mathbb{C}^{n \times 1}$; $A \otimes B = [A(k,l)B]$ stands for the Kronecker product of the matrices $A = [A(k,l)]$ and $B$. The notation ‘:=’ stands for ‘equal by definition’.

2. Statement of the problem. For a given problem

\[ F(X, S) = 0, \]

where $X$ is the solution of (2) and $S$ denotes the collection of data matrices, for the purposes of the sensitivity analysis, the uncertainties and the perturbations in the data, as well as the effect of the errors due to the finite precision machine arithmetic and the computed solution of the problem (round-off errors, errors of stopping iterative procedure) are presented as equivalent perturbations in the data $\delta S$, which result in perturbation error $\delta X$ in the solution $X$:

\[ S \rightarrow S + \delta S, \quad \text{and} \quad X \rightarrow X + \delta X. \]

The sensitivity analysis derives perturbation bounds for the norm of the error $\delta X$ in the solution as a linear or non-linear function of the norm of the perturbations in the data $\delta S$\[^7\text{-}12\].

We focus our experimental analysis on the properties of the perturbation bounds of the solution to the symmetric matrix Riccati differential equation (1). The nonlinear bounds proposed by Konstantinov and Pelova in\[^4\] and by Konstantinov and Angelova in\[^5\], and the linear bound of Weng and Phoa from\[^3\]\ are considered. Both previous comparative analyses\[^5,6\]\ show that the linear bound of Kenney and Hewer from\[^2\]\ is more pessimistic than the nonlinear bounds proposed in\[^4\] and\[^5\]\, that is why we exclude it from our analysis.

\[ C. \ R. \ Acad. \ Bulg. \ Sci., \ 75, \ No \ 11, \ 2022 \]
3. Perturbation bounds. Denote by \( S := \{A, B, C, X_0\} \) the collection of data matrices in (1), by \( \delta S := \{\delta A, \delta B, \delta C, \delta X_0\} \) the collection of perturbations in the data matrices from \( S \), and by \( \delta := [\delta A, \delta B, \delta C, \delta X_0] \), the vector of the norms of the perturbations in the data \( \delta Z := \|\delta Z\|_2, Z \in S \).

3.1. The estimate of Konstantinov and Pelova [1]. Let \( \alpha := \max \{\|\delta Z\| \} \), \( Z \in S, \beta := 1/(2\sqrt{\gamma^2 + \mu/v}), \gamma := \sigma + \|B\|_2(\mu + \nu \sigma^2), \sigma := 1 + \|X(t)\|_2, \mu := \max \{\|\Phi(0,t)\|_2^2 : t \in T\} \), where \( \Phi(0,t) \) is the fundamental matrix of the equation \( x(t) = (A - CX(t))x(t), \Phi(t, t) = I \).

If the condition

\[
\alpha < \beta, \text{ is satisfied, then }
\]

\[
\|\delta X(t)\|_2 \leq f_1(t, \alpha) = \frac{g(\alpha)}{2\nu(\|B\|_2 + \alpha)},
\]

\[
g(\alpha) = 1 - 2\nu \sigma \alpha - \sqrt{1 - 4\nu \gamma \alpha - 4\nu \mu \alpha^2}.
\]

3.2. The estimate of Konstantinov and Angelova [5]. According to Radon’s Lemma [1], any matrix Riccati differential equation (1) is locally equivalent to the linear differential system \( \dot{\Psi}(t) = M\Psi(t), \Psi(0) = I_{2n}, M := \begin{bmatrix} -A & C \\ B & A^T \end{bmatrix}, \)

with a solution \( \Psi(t) = e^{Mt}; e^{Mt} := \begin{bmatrix} \Psi_{11}(t) & \Psi_{12}(t) \\ \Psi_{21}(t) & \Psi_{22}(t) \end{bmatrix} \). If the inverse of the matrix \( \mathcal{J}(t) := \Psi_{11}(t) + \Psi_{12}(t)X_0 \) exists, then the solution of the initial value problem (1) is represented as \( X(t) = (\Psi_{21}(t) + \Psi_{22}(t)X_0)(\Psi_{11}(t) + \Psi_{12}(t)X_0)^{-1} := \mathcal{H}(t)\mathcal{J}^{-1}(t) \). Denote \( r := \text{rank}(M), \delta \Psi(t) = e^{(M + \delta M)t} - e^{Mt}, \omega_1(t, \delta) := \sqrt{r}(1 + \|X_0\|_2 + \|\delta X_0\|_2 + \|\Psi_{12}(t)\|_2 \delta X_0, \omega_2(t, \delta) := \sqrt{r}(1 + \|X_0\|_2 + \|\delta X_0\|_2 \|\delta \Psi(t)\|_2 + \|\mathcal{J}^{-1}(t)\|_2) \), \( \tau := \sup \{t \in T : \omega_1(t, \delta)\mathcal{J}^{-1}(t)\|_2 < 1\} \). For

\[
t \in [0, \tau)
\]

for the spectral norm of the perturbation \( \delta X(t) \) in the computed solution to (1), the nonlinear perturbation bound is valid

\[
\|\delta X(t)\|_2 \leq f_2(t, \delta) := \frac{(\omega_2(t, \delta) + \omega_1(t, \delta)\|X(t)\|_2)\|\mathcal{J}^{-1}(t)\|_2}{1 - \omega_1(t, \delta)\|\mathcal{J}^{-1}(t)\|_2}.
\]

3.3. The estimate of Weng and Phoa [3]. Denote \( A_g(t) := A^T - X(t)C \). Let \( \Phi_g(t) \) satisfy the differential equation \( \dot{\Phi}_g(t) = A_g(t)\Phi_g(t), \Phi_g(0) = I_n \) for \( t \in T \). Denote \( \Omega^{-1}(Z) := \int_0^t \Phi_g(t)\Phi_g^{-1}(s)Z\Phi_g^{-1}(s)\Phi_g(s) ds \),

\[
\Theta(Z) := \Omega^{-1}(Z^TX(t) + X(t)Z), m_1 := \|\Phi_g(t_1 - t)\|_F^2\|X_0\| + \|\Theta\|\|A\|,
\]

\[
\varepsilon := \max \left\{ \frac{\|\delta A\|_F}{\|A\|_F}, \frac{\|\delta X_0\|_F}{\|X_0\|_F} \right\}.
\]

V. Angelova, T. Balabanov, I. Popechov
The linear perturbation bound
\begin{equation}
\|\delta X(t)\|_2 \leq f_3(t, \delta) := \varepsilon m_1
\end{equation}
of the solution to the RDE (1) is valid.

4. Main results. Numerical examples. To compare the effectiveness of
the perturbation bounds, listed in Section 3, we consider symmetric matrix Riccati
differential equations of type (1) on a time interval \( T = [0, 1] \), for different matrix
coefficients and for size of 3 and 2. The tests are performed with GNU Octave,
version 5.2.0 and SymPy version 1.5.1, and then confirmed with Matlab.

Numerical example 1. Consider the symmetric matrix Riccati differential
equation of type (1), for matrix coefficients
\begin{equation}
Z = VZ_0V,
\end{equation}
where \( Z \in S \) and \( A_0 = \text{diag}(-1*10^{-k}, -2, -3*10^{-k}) \), \( X(0)_0 = \text{diag}(0, 0, 0) \),
\( C_0 = \text{diag}(10^{-k}, 1, 10^k) \), \( B_0 = \text{diag}(3*10^{-k}, 5, 7*10^{-k}) \), and a solution \( X(t) = V X_0(t) V \), where \( V \) is the elementary reflection
\( V = I_3 - 2uv^\top/3 \), \( v := [1, 1, 1]^\top \).
The exponent \( k \) is a positive integer. With increasing of \( k \), the conditioning of the
matrix \( A \) deteriorates and the sensitivity of the equation increases. This scheme
is used in [3] and [9] to analyze the effectiveness of the perturbation bounds for the
RDE. The advantages of this scheme is that it allows to obtain in an explicit form
the analytical solution to the exact and to the perturbed Riccati equation (1).
As the perturbation bound of Weng and Phoa from [3] considers perturbations
in the matrix \( A \) and in the initial condition \( X_0 \), the perturbations in the data
are taken as: \( \delta Z = V \delta Z_0 V \), \( Z \in S \), where \( \delta A_0 = \text{diag}(3*10^k, 2, 10^k)10^{-j} \), for
\( j = 10, 8, \ldots, 2 \). and \( \delta X(0)_0 = \delta C_0 = \delta B_0 = 0 \).
The diagonal form of the matrix coefficients \( Z_0 \), \( Z \in S \) and the solution
\( X_0(t) \) before their multiplication by the elementary reflection \( V \) allows the
symmetric matrix Riccati differential equation (1), (8), to be transformed into a scalar
equation for every of the 3 nonzero diagonal elements
\begin{equation}
\dot{x}_i(t) = 2a_i x_i(t) + b_i - c_i x_i(t)^2, \quad x_i(0) = 0,
\end{equation}
where \( x_i(t) = X(i, i)(t), z_i = Z(i, i), Z \in S \), for \( i = 1, 2, 3 \).

Theorem 1. Consider the symmetric matrix Riccati differential equation (1)
with matrix coefficients (8). This equation has an analytical solution \( X(t) = VX_0(t)V \), with \( X_0(t) = \text{diag}(x_1(t), x_2(t), x_3(t)) \), where
\begin{equation}
x_i(t) = \frac{e^{ti \sigma_2(a_i + c_i)}}{2 \sigma_2} - \frac{e^{ti \sigma_2(a_i + c_i + 3c_i, \sigma_2)}}{2 \sigma_2} + 1,
\end{equation}
with \( \sigma_1 = 2b_ic_i + 2a_i \sigma_2 + 2a_i^2 \) and \( \sigma_2 = \sqrt{a_i^2 + b_i c_i} \).
Proof. Let \( x_{i,1}(t) \) be a known particular solution to equation (9). The change of variable \( v_i(t) = x_i(t) - x_{i,1}(t) \) transforms the Riccati equation (9) into a Bernoulli equation of order \( m = 2 \). Then, the substitution \( v_i(t) = w_i(t)^{1-m} \) reduces the Bernoulli equation to an ordinary differential equation.

To determine a particular solution \( x_{i,1}(t) \), from the constant term \( b_i \) of equation (9), it can be seen that the particular solution is of the form \( x_{i,1}(t) = q \), with derivative \( \dot{x}_{i,1}(t) = 0 \). By substituting \( x_{i,1}(t) = q \) in equation (9), two values for the coefficient \( q \) are obtained \( q_{1,2} = \frac{a_i \pm \sigma_2}{c_i} \), where \( \sigma_2 := \sqrt{a_i^2 + b_i c_i} \).

The particular solution to (9) is chosen as \( x_{i,1}(t) = \frac{a_i + \sigma_2}{c_i} \). The change of variable \( v_i(t) = x_i(t) - x_{i,1}(t) \) reduces the Riccati equation (9) to a Bernoulli equation \( \dot{v}_i(t) = -c_i v_i(t)^2 - 2\sigma_2 v_i(t) \) of order \( m = 2 \). To solve the Bernoulli equation, the substitution \( v_i(t) = w_i(t)^{1-m} = w_i(t)^{1-2} = \frac{1}{v_i(t)} \), with derivative \( v_i(t) = -w_i(t) \frac{\dot{w}_i(t)}{w_i(t)^2} \), is applied to transform the Bernoulli differential equation to ordinary differential equation. Finally, the differential Riccati equation (9) is transformed to the ordinary differential equation \( \dot{w}_i(t) = 2\sigma_2 w_i(t) + c_i \) with solution \( w_i(t) = C \varphi(t) + \psi(t) \), where \( \varphi(t) := e^{\int 2\sigma_2 \, dt} \) and \( \psi(t) := \varphi(t) \int c_i \varphi(t) \, dt \).

Reverse the substitutions \( w_i(t) = \frac{1}{v_i(t)} \) and \( x_i(t) = v_i(t) + x_{i,1}(t) \), so the solution to (9) is

\[
x_i(t) = \frac{e^{\int \sigma_2 (a_i + \sigma_2) \, dt} + C e^{2\int \sigma_2 (a_i + \sigma_2) \, dt}}{C e^{2\int \sigma_2 \, dt} + c_i e^{\int \sigma_2 \, dt}} + \frac{1}{\frac{a_i c_i + 3 c_i \sigma_2}{2 b_i c_i + 2 a_i \sigma_2 + 2 a_i^2}}.
\]

Using the initial condition \( x_i(0) = 0 \) to determine \( C \) from (11), the following expression for \( C \) is obtained: \( C = -\frac{a_i c_i + 3 c_i \sigma_2}{2 b_i c_i + 2 a_i \sigma_2 + 2 a_i^2} \). Substituting \( C \) in (11), one obtains (10). \( \square \)

The bounds of Konstantinov and Pelova \( f_1(t, \alpha) \), given in (3), (4), and Konstantinov and Angelova \( f_2(t, \delta) \), given in (5), (6) are nonlinear bounds. The nonlinear bound is more pessimistic than the linear bound due to the higher-order terms of the perturbations in the data included. But the nonlinear bound is valid for perturbations belonging to an a priori prescribed domain, which guarantees the existence of the solution to the RDE. The linear bound excludes terms of second and higher order. It is easy to compute, sharper, and is valid for small perturbations in the data. The inconvenience is that no one knows how small the perturbations must be. The effectiveness of the nonlinear error bounds \( f_1(t, \alpha) \) (4) of Konstantinov and Pelova [4] with condition (3), and \( f_2(t, \delta) \) (6) of Konstantinov and Angelova [5] with condition (5), and the linear bound \( f_3(t, \delta) \) (7)
of Weng and Phoa [3] are compared. The ratios $r_i(t) := \frac{f_i(t, \cdot)}{\|\delta X(t)\|_2}$, $i = 1, 2, 3$ of the error bound $f_i(t, \cdot)$ to the norm of the perturbation in the solution $\|\delta X(t)\|_2$ are visualized in Fig. 1 for different values of the parameters $k = 0, 1, 2, 3$ and $j = 3$, and for $t \in T$.

The error bound $f_3(t, \delta)$ (7) of Weng and Phoa [3] is superior to the other two methods with respect to closeness to the estimated quantity. As it is a linear bound, the size of the domain of its validity is not specified. The bound estimates the solution to RDE at all chosen values of the parameters $k = 0, 1, 2, 3$ and $j = 10, 8, \ldots, 2$ of the experimental model, but no one knows if the perturbations are small enough in order for the estimate to be valid. The nonlinear perturbation bound $f_2(t, \delta)$ from Konstantinov and Angelova [5] is more pessimistic than the other two bounds. But its domain of validity is wider than this of the other
nonlinear bound $f_1(t, \alpha)$ (4) of Konstantinov and Pelova from [4] and has the advantage of not needing the solution of the RDE and hence, to be related to the problems with possible loss of symmetry and divergence of the computational procedure.

For $k = 0$, the RDE is of good conditioning, and all of the three bounds give results for $j$ from 10 to 2. The sharper bound is $f_3(t, \delta)$ (7) of Weng and Phoa from [3]. When $k = 1, j = 3$ from $t = 0.533$ to $t = 1$, the bound $f_1(t, \alpha)$ is not valid because of violated condition (3). For $k = 2$ the conditioning deteriorates. The nonlinear bound $f_1(t, \alpha)$ does not work for $j = 4$ and $j = 3$ because of violated condition (3). For $j = 3$ condition (5) is violated when $t$ is bigger than 0.6667, and the nonlinear bound $f_2(t, \delta)$ (6) does not work. The linear bound $f_3(t, \delta)$ (7) of Weng and Phoa [3] still gives estimate, but when $t = 1$ the ratio $r_3(t) = f_3(t, \delta)/\|\delta X(t)\|_2 = 0.9969$, which means that the error bound does not reach the estimated value. This is more visible for $k = 3, j = 3$, when after $t = 0.2667$, the curve of the ratio $r_3(t)$ passes below the line $1(t)$. The bound does reach the estimated value which is not allowed for an upper error bound.

The results show that with the deterioration of the conditioning of the system (i.e. of the matrix $A$) the linear estimate $f_3(t, \delta)$ from (7) is superior to the two nonlinear bounds with respect to closeness to the estimated quantity, but after certain conditions, the linear bound does not reach the estimated value. The nonlinear bound $f_1(t, \alpha)$ (4) is superior to the nonlinear bound $f_2(t, \delta)$ (6) in sharpness. The nonlinear perturbation bound $f_2(t, \delta)$ (6) has a wider domain for validity than the bound $f_1(t, \alpha)$ and the advantage that it is not related to the solution of the RDE and hence to the problem of possible loss of symmetry and divergence of the numerical procedure. The results are available in a detailed tabular form in [13].

**Numerical example 2.** The example is constructed on the simulation scheme of Example 1, chosen by Weng and Phoa in [3] to demonstrate the effectiveness of the bounds, proposed in the paper. A second order matrix Riccati differential equation of type (1), with matrices: $A = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}$, $B = GR^{-1}G^\top$, $G = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$, $R = I$, $C = H^\top H$, $H = \begin{bmatrix} 1 & 0 \end{bmatrix}$, is considered. The perturbations in the data are taken as $\delta Z = 10^{-3} \ast \text{rand}(2, 2)$, for $Z = \{A, X_0\}$. The experiment is performed 30 times. The average values of the ratios $r_i(t) := \frac{f_i(t, \ast)}{\|\delta X(t)\|_2}$, $i = 1, 2, 3$ of the error bound $f_i(t, \ast)$ and the norm of the perturbation in the solution $\|\delta X(t)\|_2$ are visualized in Fig. 2. All the three considered bounds give estimates. The sharper bound is the linear bound $f_3(t, \delta)$ (7) of Weng and Phoa [3]. The nonlinear bound $f_1(t, \alpha)$ (4) of Konstantinov and Pelova [4] is better than the nonlinear bound $f_2(t, \delta)$ (6) of Konstantinov and Angelova [5].
Conclusions. Three upper perturbation bounds for the solution to the symmetric matrix Riccati differential equation are considered. The effectiveness of the bounds with respect to closeness to the estimated value and size of the domain of validity are analyzed for two experimental schemes. The experimental analysis shows that for well-conditioned equation and small perturbations in the data, the linear bound of Weng and Phoa [3] is sharper than the nonlinear bound of Konstantinov and Pelova [4] and the nonlinear bound of Konstantinov and Angelova [5]. When the conditioning of the equation deteriorates and the perturbations in the data are not small enough, the two nonlinear bounds do not give results because of violated conditions of the valid size of the data perturbations. The linear bound of Weng and Phoa [3] still gives an estimate, but the value of the estimate is below the estimated quantity, which makes it unreliable for practical usage. Maybe some terms omitted must be included in order to be, indeed, an upper perturbation bound. The demonstrated advantages and disadvantages of the three perturbation bounds considered in this experimental analysis show that the bounds are alternatives. The obtained results open new horizons for research.

REFERENCES


