Appl. Math. J. Chinese Univ.
2021, 36(4): 537-553

Refined rigorous perturbation bounds for the SR decomposition

Mahvish Samar[1,*] Aamir Farooq[2]

Abstract. In this article, some new rigorous perturbation bounds for the SR decomposition under normwise or componentwise perturbations for a given matrix are derived. Also, the explicit expressions for the mixed and componentwise condition numbers are presented by utilizing the block matrix-vector equation approach. Hypothetical and trial results demonstrate that these new bounds are constantly more tightly than the comparing ones in the literature.

§1 Introduction

Let $\mathbb{R}^{m\times n}$ be the arrangement of $m\times n$ real matrices, $\mathbb{R}_r^{m\times n}$ be the subset of $\mathbb{R}^{m\times n}$ comprising of marices with rank r, A^T be the transpose of matrix A and I_r be the identity matrix of order r. Assume $A\in\mathbb{R}^{2n\times 2n}$, and $P=[e_1,e_3,...,e_{2n-1},e_2,e_4,...,e_{2n}]$ with e_k speaking to the k-th unit vector. If all even leading principal submatrices of PA^TJAP^T are nonsingular, Bunse-Gerstner [1] demonstrated that A has the accompanying SR decomposition

$$A=SR=\begin{bmatrix} S_{11} & S_{12}\\ S_{21} & S_{22}\end{bmatrix}\begin{bmatrix} R_{11} & R_{12}\\ R_{21} & R_{22}\end{bmatrix}, \tag{1.1}$$

where $S\in\mathbb{R}^{2n\times 2n}$ is a symplectic matrix, i.e., it fulfills $S^TJS=S$,

$$J=\begin{bmatrix} 0 & I_n\\ -I_n & 0\end{bmatrix}\in\mathbb{R}^{2n\times 2n},$$

$R_{ij}(i,j=1,2)$ are upper triangular, and $\mathrm{diag}(R_{21})=0$. Further, if

$$\mathrm{diag}(R_{11})=|\mathrm{diag}(R_{22})| \quad\text{and}\quad \mathrm{diag}(R_{12})=0, \tag{1.2}$$

then the SR decomposition is unique [2]. Throughout this article, we always assume that the factor R satisfies (1.2).

The SR decomposition is a helpful instrument in the calculation of some optimal control problems (e.g., [3-5]). Also the SR decomposition is a key step for constructing structure-

Received: 2020-03-29. Revised: 2020-11-07.
MR Subject Classification: 65F35, 15A23, 15A57.
Keywords: SR decomposition, Rigorous perturbation bound, Lyapunov majorant function, Banach fixed point theorem, Mixed and componentwise condition numbers.
Digital Object Identifier(DOI): https://doi.org/10.1007/s11766-021-4086-x.
The work is supported by the National Natural Science Foundation of China (Grant No. 11771265).
*Corresponding author.

preserving methods in order to solve the eigenproblem of an important class of structured matrices [2,6-8]. For more points of interest, see for instance [9,10-13]. Henceforth, it is essential to see how perturbations in the original matrix influence consequence of such a decomposition.

The perturbation analysis for the SR decomposition was first introduced by Bhatia [9]. Then Chang [2] also considered this problem and its variants. They both gave first-order perturbation bounds for the SR decomposition. Later, the acquired first-order bounds for SR decomposition was enhanced by Xie et. al [14] and they also presented the rigorous normwise perturbation bounds. However, this bound can severely overestimate the true effect of a perturbation.

In this paper, we investigate some new rigorous perturbation bounds for the SR decomposition under normwise or componentwise perturbations. In addition, the obtained first-order bounds [14] are optimal, which lead to the normwise condition numbers for SR decomposition. However, we know that the normwise condition number may overestimate the ill-posedness of the problem because they ignore the structure of coefficient matrices regarding sparsity or scaling [15,16]. So, it is essential to study the mixed and componentwise condition numbers of SR decomposition. The mixed condition numbers measure the errors in the input data using componentwise error analysis and the output using the normwise error analysis. The componentwise condition numbers measure the errors using the componentwise for both input and output data. Inspired by this, we attempt to present the explicit expressions of the mixed and componentwise condition numbers for the factors S and R.

This paper is organized as follows. In Section 3, we combine the block matrix-vector equation approach, the method of Lyapunov majorant function (e.g., [17, Chapter 5]), and the Banach fixed point theorem (e.g., [17 Appendix 5]) to study rigorous perturbation bounds for the SR decomposition when the original matrix has the normwise or componentwise perturbations. In Section 4, we give the explicit expressions for mixed and componentwise condition numbers for the this decomposition. In addition, in Section 2, we provide some notations and preliminaries and in Section 5, we give some numerical experiments.

§2 Notations and preliminaries

Some notation can be endorsed from [18,19] to make the presentation apparent. We can still illustrate them here to make easier for readers.

For the given matrix $A = (a_{ij}) \in \mathbb{R}^{m\times n}$, its spectral norm and Frobenius norm are betoken by $\|A\|_2$ and $\|A\|_F$, respectively. For these two matrix norms, the following inequalities clasp (see [20, pp.80]):

$$\|XYZ\|_2 \le \|X\|_2 \|Y\|_2 \|Z\|_2, \quad \|XYZ\|_F \le \|X\|_2 \|Y\|_F \|Z\|_2, \tag{2.1}$$

whenever the matrix product XYZ is well-defined. For the matrix $A = (A_{ij}) \in \mathbb{R}^{2n\times 2n}$, where $A_{ij} \in \mathbb{R}^{2\times 2}$, $i, j = 1, 2, \cdots, n$, we define the accompanying operators:

$$\mathrm{upb}(A) = \begin{bmatrix} \frac{1}{2}A_{11} & A_{12} & \cdots & A_{1n} \\ 0 & \frac{1}{2}A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{2}A_{nn} \end{bmatrix}, \quad \mathrm{utb}(A) = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ 0 & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{nn} \end{bmatrix},$$

$$\text{uvecb}(A) = \left[\begin{array}{c} \text{vec}(A_{11}) \\ \vdots \\ \text{vec}(A_{1n}) \\ \hline \text{vec}(A_{22}) \\ \vdots \\ \text{vec}(A_{2n}) \\ \hline \vdots \\ \hline \text{vec}\left(A_{(n-1)(n-1)}\right) \\ \text{vec}\left(A_{(n-1)n}\right) \\ \hline \text{vec}(A_{nn}) \end{array}\right] \in \mathbb{R}^{v_1}, \quad \text{vecb}(A) = \left[\begin{array}{c} \text{vec}(A_{11}) \\ \vdots \\ \text{vec}(A_{1n}) \\ \hline \vdots \\ \hline \text{vec}(A_{n1}) \\ \vdots \\ \text{vec}(A_{nn}) \end{array}\right] \in \mathbb{R}^{4n^2},$$

where $v_1 = 2n(n+1)$. Making use of the structures of the operators defined above, we have

$$\begin{gathered} \text{uvecb}(A) = M_{\text{uvecb}}\text{vecb}(A), \quad \text{vecb}(\text{utb}(A)) = M_{\text{utb}}\text{vecb}(A), \\ \text{vecb}(\text{upb}(A)) = M_{\text{upb}}\text{vecb}(A), \end{gathered} \tag{2.2}$$

where

$$M_{\text{uvecb}} = \text{diag}(S_1, S_2, \cdots, S_n) \in \mathbb{R}^{v_1 \times 4n^2}, \ \ S_i = \left[0_{4(n-i+1)\times 4(i-1)}, I_{4(n-i+1)}\right] \in \mathbb{R}^{4(n-i+1)\times 4n},$$

$$M_{\text{utb}} = \text{diag}\left(\hat{S}_1, \hat{S}_2, \cdots, \hat{S}_n\right) \in \mathbb{R}^{4n^2\times 4n^2}, \ \ \hat{S}_i = \text{diag}\left(0_{4(i-1)\times 4(i-1)}, I_{4(n-i+1)}\right) \in \mathbb{R}^{4n\times 4n},$$

$$M_{\text{upb}} = \text{diag}\left(\tilde{S}_1, \tilde{S}_2, \cdots, \tilde{S}_n\right) \in \mathbb{R}^{4n^2\times 4n^2}, \ \ \tilde{S}_i = \text{diag}\left(0_{4(i-1)\times 4(i-1)}, 1/2I_4, I_{4(n-i)}\right) \in \mathbb{R}^{4n\times 4n}.$$

Moreover,

$$M_{\text{uvecb}}M_{\text{uvecb}}^T = I_{v_1}, \quad M_{\text{uvecb}}^T M_{\text{uvecb}} = M_{\text{utb}}. \tag{2.3}$$

Thus, letting $\text{uvecb}^\dagger : \mathbb{R}^{v_1} \to \mathbb{R}^{2n\times 2n}$ be the right inverse of the operator 'uvecb' such that $\text{uvecb}\cdot\text{uvecb}^\dagger = 1_{v_1\times v_1}$ and $\text{uvecb}^\dagger\cdot\text{uvecb} = \text{utb}$. Then the matrix of the operator 'uvecb' is M_{uvecb}^T. That is, $\text{uvecb}^\dagger(A) = M_{\text{uvecb}}^T\text{vecb}(A)$.

For the operator 'upb', the following properties are needed later in this paper. Let $\mathbb{D}_{2n} \in \mathbb{R}^{2n\times 2n}$ denote the set of diagonal positive definite matrices with 2×2 main diagonal blocks s_iI_2, where $s_i > 0,\ i = 1, 2, \cdots, n$. Then, for any $D_{2n} \in \mathbb{D}_{2n}$,

$$\text{upb}(AD_{2n}) = \text{upb}(A)D_{2n}, \quad D_{2n}\text{upb}(A) = D_{2n}\text{upb}(A). \tag{2.4}$$

Furthermore, from [2],

$$\left\|\text{upb}(A) - D_{2n}^{-1}\text{upb}\left(A^T\right)D_{2n}\right\|_F \le \sqrt{1+\varsigma_{D_{2n}}^2}\|A\|_F, \tag{2.5}$$

where $\varsigma_{D_{2n}} = \max\limits_{1\le i<j\le n}\{s_j/s_i\}$.

To give the definitions of mixed and componentwise condition numbers, the following form of relative distance function will be useful (see [21] for detail). For two vectors $a = [a_1, \cdots, a_p]^T$ and $b = [b_1, \cdots, b_p]^T \in \mathbb{R}^p$, we define the entry-wise division with

$$c_{i_0} = \begin{cases} \frac{a_{i_0}}{b_{i_0}}, & \text{if } b_{i_0} \neq 0, \\ a_{i_0}, & \text{if } b_{i_0} = 0. \end{cases}$$

Then we define the componentwise distance between a and b by

$$d(a,b) = \left\|\frac{a-b}{b}\right\|_\infty = \max_{i=1,\cdots,p}\left\{\frac{|a_{i_0}-b_{i_0}|}{|b_{i_0}|}\right\} = \begin{cases} \frac{|a_{i_0}-b_{i_0}|}{|b_{i_0}|}, & \text{if } b_{i_0} \neq 0, \\ |a_{i_0}|, & \text{if } b_{i_0} = 0. \end{cases}$$

Note that when $b_{i_0} \neq 0$, $d(a,b)$ will give the relative distance from a to b with respect to b, while the absolute distance for $b_{i_0} = 0$. For the distance between the matrices $A, B \in \mathbb{R}^{n\times n}$, we define

$$d(A,B) = d(\mathrm{vec}(A), \mathrm{vec}(B)).$$

In addition, for every $\epsilon > 0$, we have $B(a,\epsilon) = \{x \in \mathbb{R}^p || x_i - a_i| \leq \epsilon |a_i|, i = 1, \cdots, p\}$, and denote the domain of definition of function $F : \mathbb{R}^p \to \mathbb{R}^q$ as $\mathrm{Dom}(F)$.

Definition 2.1. [22]: Let $F : \mathbb{R}^p \to \mathbb{R}^q$ be a continuous mapping defined on an open set $\mathrm{Dom}(F) \subset \mathbb{R}^p$, and $a \in \mathrm{Dom}(F)$, $a \neq 0$ such that $F(a) \neq 0$.
(i) The mixed condition number of F at a is defined by

$$m(F,a) = \lim_{\varepsilon\to 0} \sup_{\substack{x\in B(a,\epsilon)\\ x\neq a}} \frac{\|F(x) - F(a)\|_\infty}{\|F(a)\|_\infty} \frac{1}{d(x,a)}.$$

(ii) The componentwise condition number of F at a is defined by

$$c(F,a) = \lim_{\varepsilon\to 0} \sup_{\substack{x\in B(a,\epsilon)\\ x\neq a}} \frac{d\big(F(x), F(a)\big)}{d(x,a)}.$$

When the map F in Definition 2.1 is Fréchet differentiable, the following lemma given in [22] makes the computation of mixed and componentwise condition number easier.

Lemma 2.2. *With the same assumptions as in Definition 2.1, and supposing that F is Frèchet differentiable at a, we have*

$$m(F,a) = \frac{\||DF(a)||a|\|_\infty}{\|F(a)\|_\infty},\ c(F,a) = \left\| \frac{|DF(a)||a|}{|F(a)|} \right\|_\infty.$$

where $DF(a)$ is the Fréchet derivative of F at a.

Let $A = (A_{ij}) \in \mathbb{R}^{2m\times 2n}$ with $A_{ij} \in \mathbb{R}^{2\times 2}$, $i = 1, 2, \cdots, m$, $j = 1, 2, \cdots, n$. Like the result for the regular operator 'vec', the following result holds for 'vecb':

$$\hat{\Pi}_{m,n}\mathrm{vecb}(A) = \mathrm{vecb}(A^T), \tag{2.6}$$

where $\hat{\Pi}_{m,n} = (\Pi_{m,n} \otimes \Pi_{2,2}) \in \mathbb{R}^{4mn\times 4mn}$ with $\Pi_{m,n} = \sum_{i=1}^{m}\sum_{j=1}^{n}(E_{ij} \otimes E_{ij}^T)$. In these expressions, $\otimes$ denotes the Kronecker product [23] and the matrix $E_{ij} \in \mathbb{R}^{m\times n}$ has in the (i,j)-th position and zeros elsewhere. Given another matrix B, the block Kronecker product between B and A is defined by

$$B \boxtimes A = \begin{bmatrix} B\otimes A_{11} & B\otimes A_{12} & \cdots & B\otimes A_{1n} \\ B\otimes A_{21} & B\otimes A_{22} & \cdots & B\otimes A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ B\otimes A_{m1} & B\otimes A_{m2} & \cdots & B\otimes A_{mn} \end{bmatrix}.$$

For the block Kronecker product, the following results hold [24]

$$\mathrm{vecb}\,(ACB) = (B^T \boxtimes A)\mathrm{vecb}\,(C)\,, \tag{2.7}$$

$$\|B \boxtimes A\|_2 = \|B\|_2 \|A\|_2, \tag{2.8}$$

$$(B \boxtimes A)(C \boxtimes G) = (BC \boxtimes AG), \tag{2.9}$$

$$(B \boxtimes A)^{-1} = B^{-1} \boxtimes A^{-1}, \text{ if } B \text{ and } A \text{ are nonsingular.} \tag{2.10}$$

Here, the matrices C and G are of suitable orders and are partitioned appropriately.

§3 Rigorous perturbation bounds

However, it is well known that the first order perturbation bound may lead to erroneous conclusions because they neglect the high order terms. So it is necessary to discuss the rigorous perturbation bound. Next we consider the rigorous perturbation bounds for the factors R and S when the original matrix has normwise or componentwise perturbations in the given matrix. First, we give a unique decomposition theorem for a perturbed matrix.

Theorem 3.1. *Given $A \in \mathbb{R}^{2n\times 2n}$, consider that all even leading submatrices of PA^TJAP^T are non singular and A has the unique SR decomposition. If*

$$\|S^TJ\Delta AR^{-1}\|_2 < \sqrt{2}-1, \tag{3.1}$$

then $A+\Delta A$ has the following unique SR decomposition

$$A+\Delta A = (R+\Delta R)(S+\Delta S). \tag{3.2}$$

Proof: Note that $J^{-1} = J^T = -J$ and S is nonsingular. Then, left multiplying and right multiplying $S^TJS = J$ by S^TJ and $S^{-1}J^T$, respectively gives $SJ^TS^T = J^T$, i.e. $SJS^T = J$. Now,

$$\begin{aligned}
&P(A+\Delta A)^TJ(A+\Delta A)P^T\\
&= P(R^TJR + R^TS^TJ\Delta A + (\Delta A)^TJSR + (\Delta A)^TJ\Delta A)P^T\\
&= PR^T(J + S^TJ\Delta AR^{-1} + R^{-T}(\Delta A)^TJS + R^{-T}(\Delta A)^TJ\Delta AR^{-1})RP^T\\
&= PR^T(J+K)RP^T,
\end{aligned} \tag{3.3}$$

where $K = S^TJ\Delta AR^{-1} + R^{-T}(\Delta A)^TJS + R^{-T}(\Delta A)^TJ\Delta AR^{-1}$. Taking the spectral norm on K and using $J^TJJ = J$

$$\begin{aligned}
\|K\|_2 &\le 2\|S^TJ\Delta AR^{-1}\|_2 + \|R^{-T}(\Delta A)^TJ\Delta AR^{-1}\|_2\\
&\le 2\|S^TJ\Delta AR^{-1}\|_2 + \|R^{-T}(\Delta A)^TJ^TSJS^TJ\Delta AR^{-1}\|_2\\
&\le 2\|S^TJ\Delta AR^{-1}\|_2 + \|S^TJ\Delta AR^{-1}\|_2^2
\end{aligned}$$

by (3.1) we have $\|K\|_2 < 1$. Using result from [7] we get, $\|K_{(2i)}\|_2 < 1$ for $i = 1, 2, ..., n$. Furthermore, obviously, $\|J_{(2i)}K_{(2i)}\|_2 < 1$ also holds. Then $I_{2i} - J_{(2i)}K_{(2i)}$ is nonsingular, which together with $(J+K)_{2i} = J_{(2i)}(I_{2i} - J_{(2i)}K_{(2i)})$ shows that $(J+K)_{(2i)}$ is nonsingular. Noting the structure of R, we can proof that $(R^T(J+K)R)_{2i} = R_{2i}^T(J+K)_{2i}R_{2i}$. Since R_{2i} is nonsingular, shows that $(R^T(J+K)R)_{2i}$ is also nonsingular. Thus, observing (3.3), we can get that all even leading principal submatrices of $P(A+\Delta A)^TJ(A+\Delta A)P^T$ are nonsingular. Thus, $A+\Delta A$ has unique SR decomposition.

Remark 3.2. Here we provide the condition under which the perturbed matrix always has the unique SR factorization, while in [2,14] they only simply assume that is true.

3.1. Refined normwise perturbation bound

Assume that the perturbed SR decomposition defined in (3.2) and using the fact that

$$(A+\Delta A)^TJ(A+\Delta A) = (R+\Delta R)^TJ(R+\Delta R).$$

Using (1.1) and little simplification we will get

$$\begin{aligned} &S^TJ\Delta AR^{-1}+R^{-T}(\Delta A)^TJS+R^{-T}(\Delta A)^TJ(\Delta A)R^{-1}-R^{-T}(\Delta R)^TJ(\Delta R)R^{-1}\\ &= J\Delta RR^{-1}+(J\Delta RR^{-1})^T. \end{aligned} \tag{3.4}$$

In the above equation $J\Delta RR^{-1}$ is upper triangular, we have

$$\begin{aligned} &J\Delta RR^{-1}\\ &= \mathrm{upb}\left(S^TJ\Delta AR^{-1}+R^{-T}(\Delta A)^TJS\right)+\mathrm{upb}\left(R^{-T}(\Delta A)^TJ(\Delta A)R^{-1}-R^{-T}(\Delta R)^TJ(\Delta R)R^{-1}\right). \end{aligned} \tag{3.5}$$

Applying the operator 'vecb' to (3.5) and (2.7), (2.6) and (2.3) gives

$$\begin{aligned} \left(R^{-T}\boxtimes J\right)\mathrm{vecb}\left(\Delta R\right) &= M_{\mathrm{upb}}\left((R^{-T}\boxtimes S^TJ)+(S^TJ\boxtimes R^{-T}J)\Pi_{n,n}\right)\mathrm{vecb}(\Delta A)\\ &\quad+M_{\mathrm{upb}}\left(R^{-T}\boxtimes R^{-T}\right)\mathrm{vecb}\left((\Delta A)^TJ\Delta A-(\Delta R)^TJ\Delta R\right). \end{aligned}$$

As done in [25,26] and (2.10), we can obtain

$$\begin{aligned} \mathrm{vecb}\left(\Delta R\right) &= \left(R^T\boxtimes J^{-1}\right)M_{\mathrm{upb}}\left((R^{-T}\boxtimes S^TJ)+(S^TJ\boxtimes R^{-T}J)\Pi_{n,n}\right)\mathrm{vecb}(\Delta A)\\ &\quad+\left(R^T\boxtimes J^{-1}\right)M_{\mathrm{upb}}\left(R^{-T}\boxtimes R^{-T}\right)\mathrm{vecb}\left((\Delta A)^TJ\Delta A-(\Delta R)^TJ\Delta R\right). \end{aligned} \tag{3.6}$$

and show that Eq. (3.6) is equivalent to

$$\begin{aligned} \mathrm{uvecb}\left(\Delta R\right) &= M_{\mathrm{uvecb}}\left(R^T\boxtimes J^{-1}\right)M_{\mathrm{upb}}\left((R^{-T}\boxtimes S^TJ)+(S^TJ\boxtimes R^{-T}J)\Pi_{n,n}\right)\mathrm{vecb}(\Delta A)\\ &\quad+M_{\mathrm{uvecb}}\left(R^T\boxtimes J^{-1}\right)M_{\mathrm{upb}}\left(R^{-T}\boxtimes R^{-T}\right)\mathrm{vecb}\left((\Delta A)^TJ\Delta A-(\Delta R)^TJ\Delta R\right). \end{aligned} \tag{3.7}$$

For simplicity, let

$$G_R = M_{\mathrm{uvecb}}\left(R^T\boxtimes J^{-1}\right)M_{\mathrm{upb}}\left((R^{-T}\boxtimes S^TJ)+(S^TJ\boxtimes R^{-T}J)\Pi_{n,n}\right), \tag{3.8}$$

$$H_R = M_{\mathrm{uvecb}}\left(R^T\boxtimes J^{-1}\right)M_{\mathrm{upb}}\left(R^{-T}\boxtimes R^{-T}\right). \tag{3.9}$$

Then (3.7) becomes

$$\mathrm{uvecb}\left(\Delta R\right) = G_R\mathrm{vecb}(\Delta A)+H_R\mathrm{vecb}\left((\Delta A)^TJ\Delta A-(\Delta R)^TJ\Delta R\right). \tag{3.10}$$

Thus, applying the operator 'uvecb†' to (3.10) leads

$$\Delta R = \mathrm{uvecb}^\dagger\left[G_R\mathrm{vecb}\left(\Delta A\right)+H_R\mathrm{uvecb}\left((\Delta A)^TJ\Delta A-(\Delta R)^TJ\Delta R\right)\right].$$

The above equation can be written as an operator equation for ΔR:

$$\begin{aligned} \Delta R &= \Phi\left(\Delta R,\Delta A\right),\\ &= \mathrm{uvecb}^\dagger\left[G_R\mathrm{vecb}\left(\Delta A\right)+H_R\mathrm{uvecb}\left((\Delta A)^TJ\Delta A-(\Delta R)^TJ\Delta R\right)\right]. \end{aligned} \tag{3.11}$$

We will execute the technique of Lyapunov majorant function and the Banach fixed point principle to probe the rigorous perturbation bounds for ΔR based on the operator equation (3.11) as done in [19]. To make it easy and clear for readers and for plenum of the method we comprehend the detail process here through some steps which are same as in [19]. Assume that $Z\in\mathbb{R}^{(2n)\times(2n)}$ is upper triangular with the same structure as that of ΔR, $\|Z\|_F\le\rho$ for some $\rho\ge 0$, and $\|\Delta A\|_F=\delta$. Then it follows from the definition of the operator 'uvecb†' and (2.1) that

$$\|\Phi\left(Z,\Delta K\right)\|_F \le \|G_R\|_2\delta+\|H_R\|_2\left(\delta^2+\rho^2\right). \tag{3.12}$$

From (3.12), we have the Lyapunov majorant function of the operator equation (3.13)

$$h\left(\rho,\delta\right) = \|G_R\|_2\delta+\|H_R\|_2\left(\delta^2+\rho^2\right),$$

and the Lyapunov majorant equation

$$h(\rho,\delta)=\rho,\quad \text{i.e.,}\quad \|G_R\|_2\delta+\|H_R\|_2\big(\delta^2+\rho^2\big)=\rho. \tag{3.13}$$

Assume that $\delta\in\Omega=\delta\geq 0:1-4\|H_R\|_2\left(\|G_R\|_2\delta+\|H_R\|_2\delta^2\right)\geq 0$. Then, the Lyapunov majorant equation (3.8) has two nonnegative roots: $\rho_1(\delta)\leq\rho_2(\delta)$ with

$$\rho_1(\delta)=f(\delta)=\frac{2\left(\|G_R\|_2\delta+\|H_R\|_2\delta^2\right)}{1+\sqrt{1-4\|H_R\|_2\left(\|G_R\|_2\delta+\|H_R\|_2\delta^2\right)}}.$$

Let the set $B(\delta)$ be

$$B(\delta)=\{Z\in\mathbb{R}^{(2n)\times(2n)}:\text{Having the same structure as that of }\Delta R\text{ and }\|Z\|_F\leq f(\delta)\},$$

which is closed and convex. We can check that the operator $\Phi(\cdot,\Delta A)$ maps the set $B(\delta)$ into itself and for $Z,\tilde{Z}\in B(\delta)$,

$$\left\|\Phi(Z,\Delta A)-\Phi\left(\tilde{Z},\Delta A\right)\right\|_F\leq h'_\rho\left(f(\delta),\delta\right)\left\|Z-\tilde{Z}\right\|_F.$$

Since the derivative of the function $h(\rho,\delta)$ relative to ρ at $f(\delta)$ satisfies $h'_\rho\left(f(\delta),\delta\right)=1-\sqrt{1-4\|H_R\|_2\left(\|G_R\|_2\delta+\|H_R\|_2\delta^2\right)}<1$ when
$\delta\in\Omega_1=\left\{\delta\geq 0:1-4\|H_R\|_2\left(\|G_R\|_2\delta+\|H_R\|_2\delta^2\right)>0\right\}$. Then the operator $\Phi(\cdot,\Delta A)$ is contractive on the set $B(\delta)$ for $\delta\in\Omega_1$. Thus, from the Banach fixed point principle, we have that the operator equation (3.11), i.e., the matrix equation (3.4), has a unique solution in the set $B(\delta)$. As a result, $\|\Delta R\|_F\leq f(\delta)$ for $\delta\in\Omega_1$. In summary, we have the following main theorem.

Theorem 3.3. *With the same assumptions as in Theorem 3.1, if*

$$\|H_R\|_2\left(\|G_R\|_2\delta+\|H_R\|_2\delta^2\right)<\frac{1}{4}, \tag{3.14}$$

then $A+\Delta A$ has the unique SR factorization (3.2) *and*

$$\|\Delta R\|_2\leq\frac{2\left(\|G_R\|_2\delta+\|H_R\|_2\delta^2\right)}{1+\sqrt{1-4\|H_R\|_2\left(\|G_R\|_2\delta+\|H_R\|_2\delta^2\right)}} \tag{3.15}$$

$$\leq 2\left(\|G_R\|_2\delta+\|H_R\|_2\delta^2\right) \tag{3.16}$$

$$<(1+2\|G_R\|_2)\|\Delta A\|_F. \tag{3.17}$$

Proof: It is easy to see that the condition (3.15) is the same as the one in Ω_1. Thus, from the discussions before Theorem 3.1, it suffices to obtain the bound (3.17). This can be done by noting (3.14) and the fact

$$2\|H_R\|_2\|\Delta A\|_F\leq\sqrt{1+\|G_R\|_2^2}-\|G_R\|_2<1,$$

which can be derived from (3.16).

Remark 3.4. From (3.15), by omitting the higher order terms, we can get the first order perturbation bound of R factor

$$\|\Delta R\|_F\leq\|G_R\|_2\|\Delta A\|_F+O\left(\|\Delta A\|_F^2\right). \tag{3.18}$$

In [14], the authors presented the following optimal first order bound for R

$$\|\Delta R\|_F\leq\|(R^T\boxtimes J^T)\mathfrak{D}K_{SR}\|_2\|\Delta A\|_F+O\left(\|\Delta A\|_F^2\right), \tag{3.19}$$

where

$$\mathfrak{D}\equiv\operatorname{diag}(\mathfrak{D}_1,\mathfrak{D}_2,\mathfrak{D}_2,\mathfrak{D}_1)\in\mathbb{R}^{(4n^2)\times(4n^2)}$$

for $\mathfrak{D}_1$ and $\mathfrak{D}_2$ see [14] and

$$K_{SR} \equiv (R^{-T} \boxtimes S^T J) - (S^T J \boxtimes R^{-T})\Pi. \tag{3.20}$$

Now we show that the bound (3.18) is the same as (3.19). In fact, according to the above definition and the definition of the operator 'upb', we can check that for any matrix $X \in \mathbb{R}^{2n\times 2n}$,

$$\mathfrak{D}\mathrm{vecb}(X) = \mathrm{vecb}(\mathrm{upb}(\mathrm{X})).$$

Thus, for any matrix $X \in \mathbb{R}^{2n\times 2n}$, using (2.7) and (2.2), we have

$$\begin{aligned}
&\|(R^T \boxtimes J^T)\mathfrak{D}K_{SR}\mathrm{vecb}(W)\|_2\\
&= \|(R^T \boxtimes J^T)\mathfrak{D}\mathrm{vecb}(S^T JWR^{-1} - R^{-T}W^T J^T S)\|_2\\
&= \|(R^T \boxtimes J^T)\mathrm{vecb}(\mathrm{upb}(S^T JWR^{-1} - R^{-T}W^T J^T S))\|_2\\
&= \|\mathrm{vecb}(J^T\mathrm{upb}(S^T JWR^{-1} - R^{-T}W^T J^T S)R)\|_2\\
&= \|(R^T \boxtimes J^T)\mathrm{M}_{\mathrm{upb}}(\mathrm{upb}(S^T JWR^{-1} - R^{-T}W^T J^T S))\|_2\\
&= \|(R^T \boxtimes J^T)\mathrm{M}_{\mathrm{upb}}((R^{-T} \boxtimes S^T J) - (S^T J \boxtimes R^{-T})\Pi)\mathrm{vecb}(\mathrm{W})\|_2.
\end{aligned} \tag{3.21}$$

From the definitions of the matrices '$\mathrm{M}_{\mathrm{utb}}$' and '$\mathrm{M}_{\mathrm{upb}}$', we can verify that

$$\mathrm{M}_{\mathrm{utb}}(R^T \boxtimes J^T)\mathrm{M}_{\mathrm{upb}} = (R^T \boxtimes J^T)\mathrm{M}_{\mathrm{upb}},$$

which together with (3.21) and (2.3) gives

$$\|(R^T \boxtimes J^T)\mathfrak{D}K_{SR}\mathrm{vecb}(W)\|_2 = \|M^T_{\mathrm{uvecb}}G_R\mathrm{vecb}(W)\|_2 = \|G_R\mathrm{vecb}(W)\|_2.$$

Thus, from the definition of spectral norm we get

$$\begin{aligned}
\|(R^T \boxtimes J^T)\mathfrak{D}K_{SR}\|_2 &= \max_{\|\mathrm{vecb}(W)\|_2=1} \|(R^T \boxtimes J^T)\mathfrak{D}K_{SR}\mathrm{vecb}(W)\|_2\\
&= \max_{\|\mathrm{vecb}(W)\|_2=1} \|G_R\mathrm{vecb}(W)\|_2\\
&= \|G_R\|_2.
\end{aligned}$$

So the bounds (3.18) and (3.19) are the same. Therefore, the rigorous bounds in Theorem 3.3 can be regarded as the rigorous versions of the optimal first order perturbation bound given in [14].

Remark 3.5. In [14] the following rigorous perturbation bounds were derived by a combination of the classic and refine matrix equation approaches,

$$\|\Delta R\|_F \leqslant \sqrt{2}\kappa(R^{-1})\|S\|_2\|\Delta A\|_F + (3\sqrt{2} + 2\sqrt{3})\kappa(R^{-1})\|S\|_2^2\|\Delta A\|_F^2, \tag{3.22}$$

under the condition

$$\|S^T J\Delta A R^{-1}\|_F < 1/\sqrt{6} + 2.$$

In the above bounds, for a non singular matrix Z, $\kappa(Z)$ denotes its condition number and is defined as $\kappa(Z) = \|Z\|_2\|Z^{-1}\|_2$ the set of $n \times n$ positive diagonal matrices.
Now we will show that our bounds (3.17) is sharper than (3.22)

$$\begin{aligned}
\|G_R\|_2 &= \left\|M_{\mathrm{uvecb}}\left(R^T \boxtimes J^{-1}\right)M_{\mathrm{upb}}\left((R^{-T} \boxtimes S^T J) + (S^T J \boxtimes R^{-T}J)\Pi_{n,n}\right)\right\|_2\\
&\leq \left\|R^T\right\|_2\left\|M_{\mathrm{upb}}\left((R^{-T} \boxtimes S^T J) + (S^T J \boxtimes R^{-T}J)\right)\Pi_{n,n}\right\|_2\\
&= \left\|R^T\right\|_2 \max_{\|\mathrm{vecb}(X)\|_2=1}\left\|M_{\mathrm{upb}}(\left((R^{-T} \boxtimes S^T J) + (S^T J \boxtimes R^{-T}J)\right)\Pi_{n,n})\mathrm{vecb}\left(X\right)\right\|_2
\end{aligned}$$

$$
\begin{aligned}
&= \left\|R^T\right\|_2 \max_{\|\mathrm{vecb}(X)\|_2=1} \left\|M_{\mathrm{upb}}\mathrm{vecb}\left(S^TJXR^{-1} - (S^TJXR^{-1})^T\right)\right\|_2 \\
&= \left\|R^T\right\|_2 \max_{\|\mathrm{vecb}(X)\|_2=1} \left\|\mathrm{vecb}\left(\mathrm{upb}\left(S^TJXR^{-1} - (S^TJXR^{-1})^T\right)\right)\right\|_2 \\
&= \left\|R^T\right\|_2 \max_{\|X\|_F=1} \left\|\mathrm{upb}\left(S^TJXR^{-1} - (S^TJXR^{-1})^T\right)\right\|_F \\
&\leq \left\|R^T\right\|_2 \max_{\|X\|_F=1} \sqrt{2}\left\|S^TJXR^{-1}\right\|_F \\
&\leq \sqrt{2}\kappa(R^{-1})\left\|S\right\|_2 .
\end{aligned}
$$

Remark 3.6. From (3.2) we have,

$$\Delta A = S\Delta R + \Delta SR + \Delta S\Delta R. \tag{3.23}$$

Postmultiplying by R^{-1} gives

$$\Delta S = \Delta AR^{-1} - S\Delta RR^{-1} - \Delta S\Delta RR^{-1}. \tag{3.24}$$

Applying the operator 'vecb' to (3.24) and using (2.7), (2.6) and (2.3), we have

$$\mathrm{vecb}(\Delta S) \approx (R^{-T} \boxtimes I_{2n})\mathrm{vecb}(\Delta A) - (R^{-T} \boxtimes S)\mathrm{vecb}(\Delta R) - (R^{-T} \boxtimes I_{2n})\mathrm{vecb}(\Delta S\Delta R). \tag{3.25}$$

with the help of above equation and (3.6), we have

$$
\begin{aligned}
\mathrm{vecb}(\Delta S) \approx \big((R^{-T} \boxtimes I_{2n}) - (I_{2n} \boxtimes SJ^{-1})M_{\mathrm{upb}}\big((R^{-T} \boxtimes S^TJ) \\
+ (S^TJ \boxtimes R^{-T}J)\Pi_{n,n}\big)\big)\mathrm{vecb}(\Delta A) - (R^{-T} \boxtimes I_{2n})\mathrm{vecb}(\Delta S\Delta R).
\end{aligned} \tag{3.26}
$$

For simplicity, let

$$G_S = ((R^{-T} \boxtimes I_{2n}) - (I_{2n} \boxtimes SJ^{-1})M_{\mathrm{upb}}\left((R^{-T} \boxtimes S^TJ) + (S^TJ \boxtimes R^{-T}J)\Pi_{n,n}\right)). \tag{3.27}$$

Then (3.26) becomes

$$\mathrm{vecb}(\Delta S) \approx G_S\mathrm{vecb}(\Delta A) - (R^{-T} \boxtimes I_{2n})\mathrm{vecb}(\Delta S\Delta R). \tag{3.28}$$

Facilitate more, utilizing (3.24) and the outcomes on R consider given Theorem 3.3, we can get the rigorous perturbation bounds for S factor. The rigorous perturbation bounds for factor S is bigger than the one given in [14]. So we don't talk about their detailed derivation. From (3.26), by omitting the higher order terms, we can get the first order perturbation bound of S factor

$$\|\Delta S\|_F \leq \|G_S\|_2\|\Delta A\|_F + O\left(\|\Delta A\|_F^2\right). \tag{3.29}$$

Similar to the discussions in Remark 4.2, we can confirm that

$$\|(R^{-T} \boxtimes I_{2n}) - (I_{2n} \boxtimes SJ^{-1})\mathfrak{D}K_{SR}\|_2 = \|G_S\|_2. \tag{3.30}$$

So the bound (3.29) is same as the optimal one in [14].

3.2. Componentwise perturbation

As done in [27], in the following, we consider the componentwise perturbation in the given matrix with the following perturbation

$$|\Delta A| \leq \epsilon L|A|, \quad L = (l_{ij}) \in \mathbb{R}^{2n\times 2n}, \quad 0 \leq l_{ij} \leq 1, \tag{3.31}$$

where ΔA is the perturbation matrix and $\epsilon \geq 0$ is a small scalar. Now we consider compenent-wise rigorous perturbation bounds for S and R factors by using matrix equation approach. The

derivations of these bounds are similar to the ones in [14].

Theorem 3.7. *Given $A \in \mathbb{R}^{2n\times 2n}$, consider that all even leading submatrices of PA^TJAP^T are non singular and A has the unique SR factorization and ΔA satisfies* (3.31)*. If*

$$\||S^T||J||L||R^{-1}|\|_F \mathrm{cond}(R)\epsilon < \frac{1}{\sqrt{6}+2}, \tag{3.32}$$

then $A+\Delta A$ has the following unique SR factorization and

$$\begin{aligned}&\|\Delta R\|_F \qquad (3.33)\\ &\le \frac{\inf\limits_{D\in\mathbb{D}_{2n}} \||R||R^{-1}|D\|_2\|D^{-1}R\|_2\big(\sqrt{2+2\zeta_D^2}\||S^T||J||L||R^{-1}|\|_F + (\sqrt{3}-\sqrt{2})\||S^T||J||L||R^{-1}|\|_F\big)}{\sqrt{2}-1}\epsilon\end{aligned}$$

$$\|\Delta R\|_F \le (\sqrt{6}+\sqrt{3})\big(\inf_{D\in\mathbb{D}_{2n}} \||R||R^{-1}|D\|_2\|D^{-1}R\|_2\big)\||S^T||J||L||R^{-1}|\|_F\epsilon \tag{3.34}$$

$$\|\Delta S\|_F \le (\sqrt{6}+2\sqrt{3}+2+2\sqrt{2})\||S^T||J||L||S|\|_F\|S\|_2\mathrm{cond}(R)\epsilon, \tag{3.35}$$

where $\mathrm{cond}(R) = \||R||R^{-1}|\|_2$.

Proof: Note that J is skew-symmetric. Then, (3.5) equation can be rewritten as

$$\begin{aligned}J\Delta RR^{-1} &= \mathrm{upb}\left(S^TJ\Delta AR^{-1}\text{–}(S^TJ\Delta AR^{-1})^T\right)\\ &\quad + \mathrm{upb}\left(R^{-T}(\Delta A)^TJ(\Delta A)R^{-1}\text{–}R^{-T}(\Delta R)^TJ(\Delta R)R^{-1}\right).\end{aligned} \tag{3.36}$$

Taking the Frobenius norm on (3.36) and using (2.5) by putting $D_{2n} = I_{2n}$ yields

$$\begin{aligned}\|\Delta RR^{-1}\|_F &\le \sqrt{2}\|S^TJ\Delta AR^{-1}\|_F + \frac{1}{\sqrt{2}}\|R^{-T}(\Delta A)^TJ^TSJS^TJ(\Delta A)R^{-1}\|_F\\ &\quad + \frac{1}{\sqrt{2}}\|R^{-T}(\Delta R)^TJ(\Delta R)R^{-1}\|_F\\ \|\Delta RR^{-1}\|_F &\le \sqrt{2}\|S^TJ\Delta AR^{-1}\|_F + \frac{1}{\sqrt{2}}\|S^TJ(\Delta A)R^{-1}\|_F^2 + \frac{1}{\sqrt{2}}\|\Delta RR^{-1}\|_F^2.\end{aligned}$$

Therfore,

$$\frac{1}{\sqrt{2}}\|\Delta RR^{-1}\|_F^2 - \|\Delta RR^{-1}\|_F + \sqrt{2}\|S^TJ\Delta AR^{-1}\|_F + \frac{1}{\sqrt{2}}\|S^TJ(\Delta A)R^{-1}\|_F^2 \ge 0. \tag{3.37}$$

The inequality (3.37) can be considered as a quadratic inequality on $\|\Delta RR^{-1}\|_F$. By the assumption (3.32), we have

$$\begin{aligned}\Upsilon &\equiv (-1)^2 - 4\times\frac{1}{\sqrt{2}}\times\left(\sqrt{2}\|S^TJ\Delta AR^{-1}\|_F + \frac{1}{\sqrt{2}}\|S^TJ(\Delta A)R^{-1}\|_F^2\right)\\ &= 1 - 4\|S^TJ\Delta AR^{-1}\|_F - 2\|S^TJ(\Delta A)R^{-1}\|_F^2 \ge 0.\end{aligned}$$

Hence $\|\Delta RR^{-1}\|_F \le \frac{1}{\sqrt{2}}(1-\sqrt{\Upsilon})$ or $\|\Delta RR^{-1}\|_F \ge \frac{1}{\sqrt{2}}(1+\sqrt{\Upsilon})$ Since $\frac{1}{\sqrt{2}}(1-\sqrt{\Upsilon})$ or $\frac{1}{\sqrt{2}}(1+\sqrt{\Upsilon})$ and $\|\Delta RR^{-1}\|_F$ are all continuous functions of the ΔA, and $\Delta R\to 0$ as $\Delta A\to 0$, we must have

$$\|\Delta RR^{-1}\|_F \le \frac{1}{\sqrt{2}}\left(1-\sqrt{1-4\|S^TJ(\Delta A)R^{-1}\|_F - 2\|S^TJ(\Delta A)R^{-1}\|_F^2}\right) < \frac{1}{\sqrt{2}}.$$

Postmultiplying (3.36) by D, where $D\in\mathbb{D}_{2n}$ and noting $R = D\bar{R}$

$$\begin{aligned}J\Delta R\bar{R}^{-1} &= \mathrm{upb}\left(S^TJ\Delta A\bar{R}^{-1} - (S^TJ\Delta A\bar{R}^{-1})^T\right)\\ &\quad + \mathrm{upb}\left(R^{-T}(\Delta A)^TJ(\Delta A)\bar{R}^{-1} - R^{-T}(\Delta R)^TJ(\Delta R)\bar{R}^{-1}\right).\end{aligned}$$

Taking Frobinous norm and using (2.5), we get

$$\begin{aligned}\|\Delta R\bar{R}^{-1}\|_F &\le \sqrt{1+\zeta_D^2}\|S^TJ\Delta A\bar{R}^{-1}\|_F + \|R^{-T}(\Delta A)^TJ^TSJS^TJ(\Delta A)\bar{R}^{-1}\|_F \\ &\quad + \|R^{-T}(\Delta R)^TJ(\Delta R)\bar{R}^{-1}\|_F \\ &\le \sqrt{1+\zeta_D^2}\|S^TJ\Delta A\bar{R}^{-1}\|_F + \|S^TJ(\Delta A)R^{-1}\|_F\|S^TJ(\Delta A)\bar{R}^{-1}\|_F \\ &\quad + \|\Delta RR^{-1}\|_F\|\Delta R\bar{R}^{-1}\|_F,\end{aligned}$$

using $\|\Delta RR^{-1}\|_F < 1/\sqrt{2}$, we have

$$\|\Delta R\bar{R}^{-1}\|_F \le \frac{\sqrt{2+2\zeta_D^2}\|S^TJ\Delta A\bar{R}^{-1}\|_F + \sqrt{2}\|S^TJ(\Delta A)R^{-1}\|_F\|S^TJ(\Delta A)\bar{R}^{-1}\|_F}{\sqrt{2}-1}.$$

With help of this $\|S^TJ(\Delta A)\bar{R}^{-1}\|_F \le \|R||\bar{R}^{-1}|\|_2\||S^T||J||L||R^{-1}|\|_F\epsilon$, $\sqrt{1+\zeta_D^2} \ge 1$, which further reduces to

$$\begin{aligned}&\|\Delta R\bar{R}^{-1}\|_F \\ &\le \frac{\sqrt{2+2\zeta_D^2}\||R||\bar{R}^{-1}|\|_2\||S^T||J||L||R^{-1}|\|_F\epsilon + (\sqrt{3}-\sqrt{2})\||R||\bar{R}^{-1}|\|_2\||S^T||J||L||R^{-1}|\|_F\epsilon}{\sqrt{2}-1} \\ &\le (\sqrt{3}+\sqrt{6})\sqrt{1+\zeta_D^2}\||R||\bar{R}^{-1}|\|_2\||S^T||J||L||R^{-1}|\|_F\epsilon.\end{aligned}$$

Noting the fact that

$$\|\Delta R\|_F = \|\Delta R\bar{R}^{-1}\bar{R}\|_F \le \|\Delta R\bar{R}^{-1}\|_F\|\bar{R}^{-1}\|_2,$$

we will get (3.34). Next we prove (3.35). For this, premultiplying (3.24) by S^TJ and noting $S^TJS = J$ leads to

$$S^TJ\Delta S = S^TJ\Delta AR^{-1} - J\Delta RR^{-1} - S^TJ\Delta S\Delta RR^{-1}.$$

With the help of (3.36)

$$\begin{aligned}S^TJ\Delta S &= S^TJ\Delta AR^{-1} - \mathrm{upb}\left(S^TJ\Delta AR^{-1} - R^{-T}(\Delta A)^TJS\right) \\ &\quad - \mathrm{upb}\left(R^{-T}(\Delta A)^TJ(\Delta A)R^{-1} + R^{-T}(\Delta R)^TJ(\Delta R)R^{-1}\right) - S^TJ\Delta S\Delta RR^{-1}.\end{aligned}$$

The above equation can rewritten as

$$\begin{aligned}S^TJ\Delta S &= \mathrm{lowb}(S^TJ\Delta AR^{-1}) + (\mathrm{lowb}(S^TJ\Delta AR^{-1}))^T \\ &\quad - \mathrm{upb}\left(R^{-T}(\Delta A)^TJ^TSJS^T(\Delta A)R^{-1} + R^{-T}(\Delta R)^TJ(\Delta R)R^{-1}\right) \\ &\quad - S^TJ\Delta S\Delta RR^{-1}.\end{aligned}$$

Taking the Frobenius norm and using (2.5) yields

$$\begin{aligned}\|S^TJ\Delta S\|_F &\le \sqrt{2}\|S^TJ\Delta AR^{-1}\|_F + \frac{1}{\sqrt{2}}\|S^TJ\Delta AR^{-1}\|_F^2 + \frac{1}{\sqrt{2}}\|\Delta RR^{-1}\|_F^2 \\ &\quad + \|S^TJ\Delta S\|_F\|\Delta RR^{-1}\|_F. \qquad (3.38)\end{aligned}$$

Now,

$$\|\Delta RR^{-1}\|_F \le \frac{1}{\sqrt{2}}(4\|S^TJ(\Delta A)R^{-1}\|_F + 2\|S^TJ(\Delta A)R^{-1}\|_F^2),$$

which combined with $\|S^TJ(\Delta A)R^{-1}\|_F \le \sqrt{3/2} - 1$ gives

$$\|\Delta RR^{-1}\|_F^2 \le (5+2\sqrt{6})\|S^TJ(\Delta A)R^{-1}\|_F^2. \qquad (3.39)$$

Substituting (3.39) into (3.38) and using $\|\Delta RR^{-1}\|_F < 1/\sqrt{2}$, we have

$$\|S^TJ\Delta S\|_F \leq \frac{\sqrt{2}\|S^TJ\Delta AR^{-1}\|_F + (3\sqrt{2}+2\sqrt{3})\|S^TJ\Delta AR^{-1}\|_F^2}{1-\|\Delta RR^{-1}\|_F}$$

$$\|S^TJ\Delta S\|_F \leq \frac{(\sqrt{2}+(3\sqrt{2}+2\sqrt{3})\|S^TJ\Delta AR^{-1}\|_F)\|S^TJ\Delta AR^{-1}\|_F}{1-\frac{1}{\sqrt{2}}}$$

$$\|S^TJ\Delta S\|_F \leq \frac{2+\sqrt{6}}{\sqrt{2}-1}\|S^TJ\Delta AR^{-1}\|_F. \tag{3.40}$$

Substituting (3.39), (3.40) in (3.38) we have

$$\|S^TJ\Delta S\|_F \leq \sqrt{2}\|S^TJ\Delta AR^{-1}\|_F + (3\sqrt{2}+2\sqrt{3})\|S^TJ\Delta AR^{-1}\|_F^2 + \frac{(2+\sqrt{6})(\sqrt{3}+\sqrt{2})}{\sqrt{2}-1}\|S^TJ\Delta AR^{-1}\|_F^2. \tag{3.41}$$

By (3.31), we have

$$\|S^TJ\Delta AR^{-1}\|_F \leq \||S^TJ\Delta AR^{-1}|\|_F \leq \||S^T||J||L||S||R||R^{-1}|\|_F\epsilon, \tag{3.42}$$

Now using, $\|S^TJ(\Delta A)R^{-1}\|_F \leq \sqrt{3/2}-1$ and (3.42) the (3.41) becomes

$$\|S^TJ\Delta S\|_F \leq (2\sqrt{2}+2\sqrt{3}+2+\sqrt{6})\||S^T||J||L||S||R||R^{-1}|\|_F\epsilon.$$

Since $\|\Delta S\|_F = \|SJ^TS^TJ\Delta S\|_F \leq \|S\|_2\|S^TJ\Delta S\|_F$. So, we will get (3.35).

Utilizing the methodology of block matrix equation, we can likewise acquire the rigorous perturbation bounds for S and R factors with componentwise perturbation i.e. (3.31) however we can't demonstrate that the acquired bounds are always tighter than the (3.33), (3.34) and (3.35). Along these lines, these outcomes are not displayed in this article.

§4 Mixed and componentwise condition numbers

In this section, we present the explicit expressions for the mixed and componentwise condition numbers. To obtain the explicit expressions for mixed and componentwise condition numbers of SR decomposition (1.1), we need to define the following mappings:

$$\Psi_R : \text{vecb}(A) \to \text{uvecb}(R), \tag{4.1}$$

$$\Psi_S : \text{vecb}(A) \to \text{vecb}(S). \tag{4.2}$$

and use the following two first-order approximations:

$$\text{uvecb}(\Delta R) = G_R(\Delta A) + O\left(\|\Delta A\|_F^2\right), \tag{4.3}$$

$$\text{vecb}(\Delta S) = G_S(\Delta A) + O\left(\|\Delta A\|_F^2\right). \tag{4.4}$$

which are from (3.10) and (3.28), respectively

Lemma 4.1. *Given $A \in \mathbb{R}^{2n\times 2n}$, consider that all even leading submatrices of PA^TJAP^T are non singular and A has the unique SR decomposition. The mapping defined by* (4.1) *and* (4.2) *are Fréchet differentiable, and the derivative of Ψ_R and Ψ_S at $a = (\text{vecb}(A)^T)^T$ is given by*

$$D\Psi_R(a) = G_R, \tag{4.5}$$

$$D\Psi_S(a) = G_S. \tag{4.6}$$

where G_R and G_S are given in (3.8) *and* (3.27).

Proof: Utilizing the meaning in (4.3), (4.4) and by the definition of Fréchet derivative, we

have the desired results. The definitions for the mixed and componentwise condition numbers for the factors R and S are first given as follows

$$m_R(\Psi_R, a) = \lim_{\epsilon\to 0} \sup_{|\Delta A|\le\epsilon|A|} \frac{\|\Delta R\|_{\max}}{d(a+\Delta a, a)\|R\|_{\max}} = m_R,$$

$$c_R(\Psi_R, a) = \lim_{\epsilon\to 0} \sup_{|\Delta A|\le\epsilon|A|} \frac{1}{d(a+\Delta a, a)}\|\Delta R/R\|_{\max} = c_R,$$

$$m_S(\Psi_S, a) = \lim_{\epsilon\to 0} \sup_{|\Delta A|\le\epsilon|A|} \frac{\|\Delta S\|_{\max}}{d(a+\Delta a, a)\|S\|_{\max}} = m_S,$$

$$c_S(\Psi_S, a) = \lim_{\epsilon\to 0} \sup_{|\Delta A|\le\epsilon|A|} \frac{1}{d(a+\Delta a, a)}\|\Delta S/S\|_{\max} = c_S.$$

where $a = (\mathrm{vecb}(A)^T)^T$, and for a matrix A, $\|A\|_{\max} = \|\mathrm{vecb}(A)\|_\infty = \max\limits_{i,j}|a_{ij}|$. Thus, considering Lemma 2.2 and Lemma 4.1, we have the accompanying theorem which gives the explicit expressions of mixed and componentwise condition numbers for the factors R and S.

Theorem 4.2. *Given $A \in \mathbb{R}^{2n\times 2n}$, consider that all even leading submatrices of PA^TJAP^T are non singular and A has the unique SR decomposition. Then mixed and componentwise condition numbers for the factors R and S are given by*

$$m_R = \frac{\||G_R|\mathrm{vecb}(|A|)\|_\infty}{\|R\|_{\max}}, \quad c_R = \left\|\frac{|G_R|\mathrm{vecb}(|A|)}{|R|}\right\|_\infty, \tag{4.7}$$

$$m_S = \frac{\||G_S|\mathrm{vecb}(|A|)\|_\infty}{\|S\|_{\max}}, \quad c_S = \left\|\frac{|G_S|\mathrm{vecb}(|A|)}{|S|}\right\|_\infty. \tag{4.8}$$

where G_R and G_S are given in (3.8) *and* (3.27).

In accompanying, we give the upper bounds for the above two condition numbers because the matrices in the exact expressions are very large and will be expensive to compute them.

Corollary 4.3. *With the same assumptions as in Theorem 4.2, we have*

$$m_R \le \frac{\|\mathrm{upb}\left((|R^{-T}||A^T||J||S|) + (|S^T||J||A||R^{-1}|)\right)|R|\|_{max}}{\|R\|_{\max}} = m_R^{upp}(A), \tag{4.9}$$

$$c_R \le \left\|\mathrm{upb}\left((|R^{-T}||A^T||J||S|) + (|S^T||J||A||R^{-1}|)\right)\right\|_{\max} = c_R^{upp}(A), \tag{4.10}$$

$$m_S \le \frac{\||R^{-1}|A|| + |S||J^{-1}|\mathrm{upb}\left((|R^{-T}||A^T||J||S|) + (|S^T||J||A||R^{-1}|)\right)\|_{\max}}{\|S\|_{\max}} = m_S^{upp}(A), \tag{4.11}$$

$$c_S \le \left\|\frac{|R^{-1}|A|| + |S||J^{-1}|\mathrm{upb}\left((|R^{-T}||A^T||J||S|) + (|S^T||J||A||R^{-1}|)\right)}{|S|}\right\|_{\max} = c_S^{upp}(A). \tag{4.12}$$

Proof: As for (4.9), it can be obtained from (4.7) and

$$\begin{aligned}
&\||M_{\mathrm{uvecb}}\left(R^T \boxtimes J^{-1}\right) M_{\mathrm{upb}}\left((R^{-T}\boxtimes S^TJ) + (S^TJ\boxtimes R^{-T}J)\Pi_{n,n}\right)|\mathrm{vecb}(|A|)\|_\infty\\
&\le \|M_{\mathrm{uvecb}}\left(|R^T|\boxtimes|J^{-1}|\right) M_{\mathrm{upb}}\left((|R^{-T}|\boxtimes|S^TJ|) + (|S^TJ|\boxtimes|R^{-T}J|)\Pi_{n,n}\right)|\mathrm{vecb}(|A|)\|_\infty\\
&= \|\mathrm{uvecb}(|J^{-1}|\mathrm{upb}\left((|R^{-T}||A^T||J||S|) + (|S^T||J||A||R^{-1}|)\right)|R|)\|_\infty\\
&\le \|\mathrm{upb}\left((|R^{-T}||A^T||J||S|) + (|S^T||J||A||R^{-1}|)\right)|R|\|_{\max}.
\end{aligned}$$

Similarly, we can obtain (4.10), (4.11) and (4.12).

§5 Numerical Experiments

In this section, we will give some numerical precedents to show the outcomes got in Sections 3 and 4. All calculations are done in MATLAB R2016a, with accuracy 2.22×10^{-16}.

Example 5.1. In the first example, we compare the normwise, mixed and componentwise condition numbers and their upper bounds given in Theorem 4.2 and Corollary 4.3 respectively. In this example, we consider the Hamiltonian matrix

$$K = \begin{bmatrix} 0.1 & 0 & 1 & 0 \\ 0 & x & 0 & 0 \\ 1 & 1 & -0.1 & 0 \\ 1 & 1 & 0 & -x \end{bmatrix}.$$

Table 1. Comparison of condition numbers and their upper bounds for different values of x.

x	0.1	0.2	0.3	0.4	0.5	1
$\kappa_R(A)$	24.5283	11.9803	7.5375	5.6071	4.6918	4.2205
$m_R(A)$	2.3758	1.3727	0.9720	0.7690	0.6438	0.8316
$c_R(A)$	11.1317	7.0121	5.3901	4.6119	4.1606	4.0217
$m_R^{upp}(A)$	14.8171	9.9012	8.0688	7.2430	6.8133	6.3460
$c_R^{upp}(A)$	53.8659	36.8307	30.6957	28.1660	27.0700	27.8927
$\kappa_S(A)$	96.9234	26.4750	15.1380	11.6387	9.8911	7.5649
$m_S(A)$	13.5598	3.6199	2.8858	2.5300	2.3165	1.8851
$c_S(A)$	50.2804	11.9383	6.1466	5.8693	5.7145	6.0673
$m_S^{upp}(A)$	50.8248	29.2708	21.5888	17.8466	15.6537	11.4092
$c_S^{upp}(A)$	144.4159	83.4914	62.8635	59.6577	57.8969	54.4266

In Table 1 $\kappa_R(A)$ and $\kappa_S(A)$ denotes the normwise condition number for factors R and S individually. For their explicit expressions see [14]. Hence, upon calculations, we get the numerical outcomes recorded in Table 1, from which we can find that the mixed and componentwise condition numbers are tighter than the normwise one.

Example 5.2. Let $A = (a_{ij}) \in \mathbb{R}^{2n\times 2n}$ be a Hilbert matrix (i.e., $a_{ij} = 1/(i+j-1)$), a Frank matrix (i.e., $a_{ij} = 2n-j+1, i \le j; a_{i,i-1} = 2n-i+1; a_{ij} = 0, i > j+1$), a Pascal matrix (i.e., $a_{i1} = a_{1j} = 1, a_{ij} = a_{i-1,j} + a_{i,j-1}$) and a Random($2n$) matrix. From Table 2, we can find that the bound (3.17) is constantly more tightly at that point the bound (3.22) however the disadvantage is that it is somewhat costly by looking section set apart by $t_{(3.17)}$ and $t_{(3.22)}$.

Example 5.3. The test matrix A and L are set to be

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

$L = 10^{-l}\text{Random}(2n)$, where $A_{11} = \text{eye}(n)$; $A_{22} = \text{diag}((1.1)^{1/2}, (1.1)^{2/2}, ..., (1.1)^{n/2})$ and

$$A_{12} = \begin{bmatrix} 0.1 & & & & \\ 0.1 & 0.1 & & & \\ & \ddots & \ddots & & \\ & & & 0.1 & 0.1 \end{bmatrix}, \quad A_{21} = \begin{bmatrix} 0.1 & & & & \\ 0.01 & 0.1 & & & \\ & \ddots & \ddots & & \\ & & & 0.01 & 0.1 \end{bmatrix}.$$

Table 2. Comparison of rigorous normwise perturbation bounds (3.17) and (3.22).

n		$b_{(3.17)}$	$t_{(3.17)}$	$b_{(3.22)}$	$t_{(3.22)}$
2	Pascal	40.6367	0.111850	174.5033	0.015743
	Hilbert	1.2804e+03	0.135869	4.9493e+04	0.023751
	Frank	41.4635	0.132276	155.6911	0.016750
	Random	28.3318	0.085248	646.7755	0.030598
3	Pascal	601.7476	0.159316	4.8353e+03	0.025492
	Hilbert	1.0225e+05	0.140354	1.2674e+08	0.024408
	Frank	1.3537e+03	0.164466	5.2178e+04	0.115959
	Random	2.1857e+03	0.148285	4.9745e+04	0.025030
4	Pascal	3.1207e+07	0.163765	8.5424e+08	0.025447
	Hilbert	4.2752e+08	0.146814	1.7700e+12	0.025031
	Frank	2.7770e+05	0.153419	4.1753e+06	0.026466
	Random	4.8270e+03	0.229636	1.9533e+05	0.024839
5	Pascal	4.0686e+11	0.167354	8.2880e+13	0.028043
	Hilbert	9.4404e+11	0.215623	4.0918e+16	0.025678
	Frank	1.5548e+10	0.282169	9.8352e+12	0.027103
	Random	6.8858e+06	0.346150	6.7399e+08	0.025810
6	Pascal	2.7992e+14	0.318038	9.9907e+16	0.075091
	Hilbert	1.3927e+19	0.218312	1.3368e+24	0.028824
	Frank	3.4876e+13	0.393546	8.0500e+16	0.032341
	Random	7.8562e+09	0.437335	3.3139e+12	0.026591
8	Pascal	1.4795e+20	0.214910	6.0466e+22	0.075685
	Hilbert	2.9537e+32	0.250255	3.4004e+37	0.045839
	Frank	1.1544e+30	0.547057	3.5986e+35	0.037279
	Random	4.1896e+12	0.672344	4.1544e+15	0.042988
10	Pascal	9.7643e+20	0.452109	3.5216e+23	0.085685
	Hilbert	1.7852e+33	0.576208	7.0921e+38	0.056583
	Frank	8.9764e+30	0.862061	1.8352e+36	0.041279
	Random	6.8043e+13	0.663286	9.0526e+16	0.062988

Table 3. Rigorous componentwise perturbation bounds (3.34) and (3.35).

l	n	4	6	8	10	12	14	30	50
-1	$b_{(3.34)}$	8.2312	12.3658	20.3615	23.4827	30.2538	31.8950	43.6521	76.4218
	$b_{(3.35)}$	39.2562	55.5677	136.9707	141.922	203.3562	294.5619	487.2063	673.65431
-2	$b_{(3.34)}$	0.7990	1.1089	1.8772	2.3197	2.9432	3.3237	22.6706	37.4325
	$b_{(3.35)}$	3.6962	5.0394	12.6822	13.5271	20.1710	21.3739	55.9432	87.3452
-3	$b_{(3.34)}$	0.0822	0.1132	0.1886	0.2219	0.2929	0.3249	18.4312	29.7439
	$b_{(3.35)}$	0.3898	0.5063	1.2157	1.3213	2.0074	2.9681	42.5643	65.0945
-4	$b_{(3.34)}$	0.0066	0.0110	0.0186	0.0215	0.0290	0.0321	12.9812	15.6205
	$b_{(3.35)}$	0.0316	0.0503	0.1248	0.1260	0.1963	0.1989	19.5219	25.0645
-5	$b_{(3.34)}$	0.0005	0.0012	0.0185	0.0022	0.0028	0.0034	5.5409	9.3482
	$b_{(3.35)}$	0.0038	0.0050	0.1155	0.1270	0.1280	0.1310	11.8543	17.4762

From Table 3, we can see that the bound (3.34) is constantly more tightly at that point the bound (3.35).

References

[1] A Bunse-Gerstner. *Matrix factorization for symplectic QR-like methods*, Linear Algebra Appl, 1986, 83: 49-77.

[2] X W Chang. *On the sensitivity of the SR decomposition*, Linear Algebra Appl, 1998, 282(1-3): 297-310.

[3] A Bunse-Gerstner, V Mehrmann. *A symplectic QR-like algorithm for the solution of the real algebraic Riccati equation*, IEEE Trans Automat Control, 1986, 31(12): 1104-1113.

[4] H A Kandil, G Freiling, V Ionescu, G Jank. *Matrix Riccati Equations in Control and Systems Theory*, Birkhauser Verlag, Basel, 2003.

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

[6] C Paige, C Van Loan. *A Schur decomposition for Hamiltonian matrices*, Linear Algebra Appl, 1981, 41: 11-32.

[7] C Van Loan. *A symplectic method for approximating all the eigenvalues of a Hamiltonian matrix*, Linear Algebra Appl, 1984, 61: 233-251.

[8] D S Watkins, L Elsner. Self-similar flows, Linear Algebra Appl, 1983, 110: 213-242.

[9] R Bhatia. *Matrix factorizations and their perturbations*, Linear Algebra Appl, 1994, 197-198: 245-276.

[10] A Salam. *On theoretical and numerical aspects of symplectic Gram-Schmidt-like algorithms*, Numer. Algorithms, 2005, 39: 437-462.

[11] A Salam, E Al-Aidarous. *Error analysis and computational aspects of SR factorization via optimal symplectic householder transformations*, Electron Trans Numer Anal, 2009, 33: 189-206.

[12] A Salam, E Al-Aidarous, A El Farouk. *Optimal symplectic Householder transformations for SR decomposition*, Linear Algebra Appl, 2008, 429: 1334-1353.

[13] A Salam, A El Farouk, E Al-Aidarous. *Symplectic Householder transformations for a QR-like decomposition, a geometric and algebraic approaches*, J Comput Appl Math, 2008, 214: 533-548.

[14] Z Xie, W Li. *Sensitivity analysis for the SR decomposition*, Linear and Multilinear Algebra, 2015, 63: 222-234.

[15] N J Higham. *Accuracy and Stability of Numerical Algorithms*, second ed, SIAM, Philadelphia, 2002.

[16] J Sun. *Condition numbers of algebraic Riccati equations in the Frobenius norm*, Linear Algebra Appl, 2002, 350: 237-261.

[17] M Konstantinov, D Gu, V Mehrmann, P Petkov. *Perturbation Theory for Matrix Equations*, Elsevier, Amsterdam, 2003.

[18] X W Chang. *Perturbation analysis of some matrix factorization*, Ph D diss, McGill University, 1997.

[19] Li Hanyu, Y Wei, Y Yang. *New rigorous perturbation bounds for the Cholesky-like factorization of skew-symmetric matrix*, Linear Algebra Appl, 2016, 491: 83-100.

[20] G W Stewart, J Sun. *Matrix Perturbation Theory*, Academic Press, Boston, 1990.

[21] Z Xie, W Li, X Jin. *On condition numbers for the canonical generalized polar decomposition of real matrices*, Electron J Linear Algebra, 2013, 26: 842-857.

[22] F Cucker, H Diao, Y Wei. *On mixed and componentwise condition numbers for Moore-Penrose inverse and linear least squares problems*, Math Comp, 2007, 76: 947-963.

[23] R A Horn, C R Johnson. *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

[24] R H Koning, H Neudecker, T Wansbeek. *Block Kronecker products and the vecb operator*, Linear Algebra Appl, 1991, 149: 165-184.

[25] M Samar, A Farooq, H Li, C Mu. *New rigorous perturbation bounds for the generalized Cholesky factorization*, Appl Math Comput, 2019, 362: 124556.

[26] M Samar, A Farooq, C Mu. *Structured condition numbers and statistical condition estimation for the LDU factorization*, Appl Math J Chin Univ, 2020, 35: 332-348.

[27] S Singer, S Singer. *Rounding error and perturbation bounds for the symplectic QR factorization*, Linear Algebra Appl, 2003, 358: 255-279.

[1]Department of Mathematics, Shantou University, Shantou 515063, China.
Email: mahvishsamar@hotmail.com
[2]College of Mathematics and Statistics, Chongqing University, Chongqing 401331, China.