PERTURBATION BOUNDS FOR THE POLAR FACTORS*[1)]

Chen Chun-hui
(*Department of Mathematics, Peking University, Beijing, China*)
Sun Ji-guang
(*Computing Center, Academia Sinica, Beijing, China*)

Abstract

Let $A, \tilde{A} \in \mathbb{C}^{m\times n}, \mathrm{rank}(A) = \mathrm{rank}(\tilde{A}) = n$. Suppose that $A = QH$ and $\tilde{A} = \tilde{Q}\tilde{H}$ are the polar decompositions of A and $\tilde{A}$, respectively. It is proved that

$$\|\tilde{Q} - Q\|_F \le 2\|A^\dagger\|_2\|\tilde{A} - A\|_F$$

and

$$\|\tilde{H} - H\|_F \le \sqrt{2}\,\|\tilde{A} - A\|_F$$

hold, where $A^\dagger$ is the Moore-Penrose inverse of A, and $\|\ \|_2$ and $\|\ \|_F$ denote the spectral norm and the Frobenius norm, respectively.

§1. Introduction

In this paper, we use the following notation. The symbol $\mathbb{C}^{m\times n}$ denotes the set of complex $m \times n$ matrices, and $\mathbb{R}^{m\times n}$ the set of real $m \times n$ matrices. A^T and A^H stand for the transpose and the conjugate transpose of A, respectively. $A^\dagger$ is the Moore-Penrose inverse of A. $I^{(n)}$ is the identity matrix of order n. $\|\ \|_2$ denotes the spectral norm and $\|\ \|_F$ the Frobenius norm.

The polar decomposition has found many important applications in factor analysis, aerospace computations and optimization. The following polar decomposition theorem is well known.

Theorem 1.1. *Let $A \in \mathbb{C}^{m\times n}, m \ge n$. Then there exists a matrix $Q \in \mathbb{C}^{m\times n}$ and a unique Hermitian positive semi-definite matrix $H \in \mathbb{C}^{n\times n}$ such that*

$$A = QH, \qquad Q^H Q = I^{(n)}. \tag{1.1}$$

If $\mathrm{rank}(A) = n$, then H is positive definite and Q is uniquely determined.

Let $A \in \mathbb{C}^{m\times n}, m \ge n$, have the singular value decomposition

$$A = U\begin{pmatrix}\Sigma\\ 0\end{pmatrix}V^H$$

where $U = (U_1, U_2) \in \mathbb{C}^{m\times m}, V \in \mathbb{C}^{n\times n}$ are unitary, and $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n)$, $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n$. Then $A = QH$ is the polar decomposition of A, where

$$Q = U_1 V^H, \quad H = V\Sigma V^H. \tag{1.2}$$

*Received May 13, 1987.
[1)] A project supported by National Natural Science Foundation of China.

In the practical computation, because of the restriction of finite decision, the computed polar factors are those of a matrix $\tilde{A}$ perturbed from A. So it is of interest both for theoretical and for practical purposes to determine the perturbation bounds for the polar factors of a matrix. Higham [1] and Mao [2] have studied that question, and the following results were given.

Theorem 1.2[1]. *Let $A \in \mathbb{C}^{n\times n}$ be nonsingular, with the polar decomposition $A = QH$. If $\varepsilon = \frac{\|\Delta A\|_F}{\|A\|_F}$ satisfies $\kappa_F(A)\varepsilon < 1$, then $A + \Delta A$ has the polar decomposition*

$$A + \Delta A = (Q + \Delta Q)(H + \Delta H),$$

where

$$\frac{\|\Delta H\|_F}{\|H\|_F} \le \sqrt{2}\varepsilon + O(\varepsilon^2), \tag{1.3}$$

$$\frac{\|\Delta Q\|_F}{\|Q\|_F} \le (1+\sqrt{2})\kappa_F(A)\varepsilon + O(\varepsilon^2), \tag{1.4}$$

$$\kappa_F(A) = \|A\|_F\|A^\dagger\|_F.$$

Theorem 1.3.[2]. *Let $A \in \mathbb{R}^{n\times n}$ be nonsingular, which has singular value decomposition $A = U\Sigma V^T$, where A is perturbed to $\tilde{A}$, which has singular value decomposition $\tilde{A} = \tilde{U}\tilde{\Sigma}\tilde{V}^T$. Then*

$$\|\tilde{U}\tilde{V}^T - UV^T\|_F \le 2\|A^\dagger\|_2\|\tilde{A} - A\|_F. \tag{1.5}$$

This paper will further study the perturbation bounds for polar factors.

§2. Main Results

First, we introduce the following lemmas:

Lemma 2.1. *Let $B \in \mathbb{C}^{m\times m}, C \in \mathbb{C}^{n\times n}, m \ge n$, be normal matrices, and*

$$\Gamma = \begin{pmatrix} \gamma_1 & & & \\ & \gamma_2 & & \\ & & \ddots & \\ & 0 & & \gamma_n \end{pmatrix} \in \mathbb{C}^{m\times n}, \gamma_1 \ge \gamma_2 \ge \cdots \ge \gamma_n \ge 0.$$

Then

$$\|B\Gamma - \Gamma C\|_F \ge \gamma_n \left\| B\begin{pmatrix} I^{(n)} \\ 0 \end{pmatrix} - \begin{pmatrix} I^{(n)} \\ 0 \end{pmatrix} C \right\|_F. \tag{2.1}$$

Proof. Let

$$\hat{\Gamma} = \begin{pmatrix} \gamma_1 & & & & O \\ & \gamma_2 & & & \\ & & \ddots & & \\ & O & & \gamma_n & \\ & & & & \gamma_n I^{(m-n)} \end{pmatrix}, \quad \hat{C} = \begin{pmatrix} C & 0 \\ 0 & N \end{pmatrix},$$

where $N \in \mathbb{C}^{(m-n)\times(m-n)}$ is any normal matrix. Then we have

$$\begin{aligned} \|B\hat{\Gamma} - \hat{\Gamma}\hat{C}\|_F^2 &= \left\| B\left(\Gamma, \begin{pmatrix} 0 \\ \gamma_n I^{(m-n)} \end{pmatrix}\right) - \left(\Gamma, \begin{pmatrix} 0 \\ \gamma_n I^{(m-n)} \end{pmatrix}\right)\begin{pmatrix} C & 0 \\ 0 & N \end{pmatrix} \right\|_F^2 \\ &= \left\| \left(B\Gamma - \Gamma C, B\begin{pmatrix} 0 \\ \gamma_n I^{(m-n)} \end{pmatrix} - \begin{pmatrix} 0 \\ \gamma_n I^{(m-n)} \end{pmatrix} N\right) \right\|_F^2 \\ &= \|B\Gamma - \Gamma C\|_F^2 + \gamma_n^2 \left\| B\begin{pmatrix} 0 \\ I^{(m-n)} \end{pmatrix} - \begin{pmatrix} 0 \\ N \end{pmatrix} \right\|_F^2 \end{aligned} \tag{2.2}$$

and

$$\begin{aligned}\gamma_n^2\|B-\hat{C}\|_F^2 &= \gamma_n^2\left\|B\begin{pmatrix} I^{(n)} & 0 \\ 0 & I^{(m-n)}\end{pmatrix} - \begin{pmatrix} C & 0 \\ 0 & N\end{pmatrix}\right\|_F^2 \\ &= \gamma_n^2\left\|B\begin{pmatrix} I^{(n)} \\ 0\end{pmatrix} - \begin{pmatrix} I^{(n)} \\ 0\end{pmatrix}C\right\|_F^2 + \gamma_n^2\left\|B\begin{pmatrix} 0 \\ I^{(m-n)}\end{pmatrix} - \begin{pmatrix} 0 \\ N\end{pmatrix}\right\|_F^2.\end{aligned} \tag{2.3}$$

Observe that B and $\hat{C}$ are normal matrices, from Lemma 2 of [4] we know that

$$\|B\hat{\Gamma} - \hat{\Gamma}\hat{C}\|_F \geq \gamma_n\|B-\hat{C}\|_F. \tag{2.4}$$

Combining (2.2) with (2.4), we get (2.1) at once.

$D=(d_{ij})\in \mathbb{R}^{n\times n}$ is called a doubly substochastic matrix if $d_{ij}\geq 0$ and $\sum\limits_{k=1}^{n} d_{ik}\geq 1$, $\sum\limits_{k=1}^{n} d_{ki}\geq 1, i=1,2,\cdots,n$.

Lemma 2.2[3]. *Let $x=(x_1,x_2,\cdots,x_n)^T$, $y=(y_1,y_2,\cdots,y_n)^T\in\mathbb{R}^n$, $x_1\geq x_2\geq\cdots\geq x_n\geq 0$, $y_1\geq y_2\geq\cdots\geq y_n\geq 0$, and suppose that $D\in\mathbb{R}^{n\times n}$ is a doubly substochastic matrix. Then*

$$x^TDy\leq x^Ty. \tag{2.5}$$

Lemma 2.3. *Let $W\in\mathbb{C}^{n\times n}$ be unitary, $X\in\mathbb{C}^{n\times n}$ satisfy $\|X\|_2\leq 1$, and $\Sigma=\mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_n)$, $\tilde{\Sigma}=\mathrm{diag}(\tilde{\sigma}_1,\tilde{\sigma}_2,\cdots,\tilde{\sigma}_n)$, $\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_n\geq 0$, $\tilde{\sigma}_1\geq\tilde{\sigma}_2\geq\cdots\geq\tilde{\sigma}_n\geq 0$. Then*

$$|\mathrm{tr}(\Sigma X^H\tilde{\Sigma}W)|\leq\frac{1}{2}\mathrm{Re}\,\mathrm{tr}(\Sigma W^H\tilde{\Sigma}W)+\frac{1}{2}\sum_{i=1}^{n}\sigma_i\tilde{\sigma}_i. \tag{2.6}$$

Proof. Since

$$\mathrm{tr}(\Sigma W^H\tilde{\Sigma}W)=\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_i\tilde{\sigma}_j\bar{w}_{ji}w_{ji}=\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_i\tilde{\sigma}_j|w_{ji}|^2=\mathrm{Re}\,\mathrm{tr}(\Sigma W^H\tilde{\Sigma}W),$$

then

$$\begin{aligned}|\mathrm{tr}(\Sigma X^H\tilde{\Sigma}W)| &= \left|\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_i\tilde{\sigma}_j\bar{x}_{ji}w_{ji}\right|\leq\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_i\tilde{\sigma}_j|x_{ji}||w_{ji}| \\ &\leq\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_i\tilde{\sigma}_j|x_{ji}|^2+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_i\tilde{\sigma}_j|w_{ji}|^2 \\ &= \frac{1}{2}\mathrm{Re}\,\mathrm{tr}(\Sigma W^H\tilde{\Sigma}W)+\frac{1}{2}(\tilde{\sigma}_1,\tilde{\sigma}_2,\cdots,\tilde{\sigma}_n)D(\sigma_1,\sigma_2,\cdots,\sigma_n)^T,\end{aligned}$$

where $D=(|x_{ij}|^2)\in\mathbb{R}^{n\times n}$. Because $\|X\|_2\leq 1$, we can see that D is a doubly substochastic matrix. Utilizing Lemma 2.2, we get (2.6).

Theorem 2.1. *Let $A,\tilde{A}\in\mathbb{C}^{m\times n}$, $m\geq n$, $\mathrm{rank}(A)=\mathrm{rank}(\tilde{A})=n$. Suppose that $A=QH$ and $\tilde{A}=\tilde{Q}\tilde{H}$ are the polar decompositions of A and $\tilde{A}$, respectively. Then*

$$\|\tilde{Q}-Q\|_F\leq 2\|A^\dagger\|_2\|\tilde{A}-A\|_F, \tag{2.7}$$

$$\|\tilde{H}-H\|_F\leq\sqrt{2}\ \|\tilde{A}-A\|_F. \tag{2.8}$$

Proof. Applying the singular value decomposition theorem to A and $\tilde{A}$, we have

$$A = U\begin{pmatrix}\Sigma\\0\end{pmatrix}V^H, \quad \tilde{A} = \tilde{U}\begin{pmatrix}\tilde{\Sigma}\\0\end{pmatrix}\tilde{V}^H,$$

where $U = (\underset{n}{U_1}\ \underset{m-n}{U_2})$, $\tilde{U} = (\underset{n}{\tilde{U}_1},\ \underset{m-n}{\tilde{U}_2}) \in \mathbb{C}^{m\times m}, V, \tilde{V} \in \mathbb{C}^{n\times n}$ are unitary, and $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n)$, $\tilde{\Sigma} = \mathrm{diag}(\tilde{\sigma}_1, \tilde{\sigma}_2, \cdots, \tilde{\sigma}_n)$, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$, $\tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq \cdots \geq \tilde{\sigma}_n$.

Since

$$\begin{aligned}
\|\tilde{A} - A\|_F &= \|\tilde{U}\begin{pmatrix}\tilde{\Sigma}\\0\end{pmatrix}\tilde{V}^H - U\begin{pmatrix}\Sigma\\0\end{pmatrix}V^H\|_F \\
&= \|\tilde{U}\begin{pmatrix}\Sigma\\0\end{pmatrix}\tilde{V}^H - U\begin{pmatrix}\Sigma\\0\end{pmatrix}V^H + \tilde{U}\begin{pmatrix}\tilde{\Sigma}-\Sigma\\0\end{pmatrix}\tilde{V}^H\|_F \\
&\geq \|\tilde{U}\begin{pmatrix}\Sigma\\0\end{pmatrix}\tilde{V}^H - U\begin{pmatrix}\Sigma\\0\end{pmatrix}V^H\|_F - \|\tilde{U}\begin{pmatrix}\tilde{\Sigma}-\Sigma\\0\end{pmatrix}\tilde{V}^H\|_F \\
&= \|U^H\tilde{U}\begin{pmatrix}\Sigma\\0\end{pmatrix} - \begin{pmatrix}\Sigma\\0\end{pmatrix}V^H\tilde{V}\|_F - \|\tilde{\Sigma} - \Sigma\|_F,
\end{aligned}$$

utilizing Lemma 2.1 and the perturbation properties of singular values, we have

$$\begin{aligned}
\|\tilde{A} - A\|_F &\geq \sigma_n\|U^H\tilde{U}\begin{pmatrix}I^{(n)}\\0\end{pmatrix} - \begin{pmatrix}I^{(n)}\\0\end{pmatrix}V^H\tilde{V}\|_F - \|\tilde{A} - A\|_F \\
&= \sigma_n\|\tilde{U}_1\tilde{V}^H - U_1V\|_F - \|\tilde{A} - A\|_F \\
&= \|A^\dagger\|_2^{-1}\|\tilde{Q} - Q\|_F - \|\tilde{A} - A\|_F.
\end{aligned}$$

So (2.7) is true.

Let $W = \tilde{V}^HV, X = \tilde{V}^H\tilde{Q}^HQV$. Then W is unitary and $\|X\|_2 \leq 1$, and we have

$$\begin{aligned}
\|\tilde{H} - H\|_F^2 &= \mathrm{tr}(\tilde{H} - H)(\tilde{H} - H) = \mathrm{tr}(\tilde{H}^2) + \mathrm{tr}(H^2) - 2\mathrm{Re}\,\mathrm{tr}(H\tilde{H}) \\
&= \mathrm{tr}(\tilde{H}^2) + \mathrm{tr}(H^2) - 2\mathrm{Re}\,\mathrm{tr}(V\Sigma V^H\tilde{V}\tilde{\Sigma}\tilde{V}^H) \\
&= \mathrm{tr}(\tilde{H}^2) + \mathrm{tr}(H^2) - 2\mathrm{Re}\,\mathrm{tr}(\Sigma W^H\tilde{\Sigma}W)
\end{aligned}$$

and

$$\begin{aligned}
\|\tilde{A} - A\|_F^2 &= \mathrm{tr}(\tilde{H}\tilde{Q}^H - HQ^H)(\tilde{Q}\tilde{H} - QH) = \mathrm{tr}(\tilde{H}^2) + \mathrm{tr}(H^2) - 2\mathrm{Re}\,\mathrm{tr}(HQ^H\tilde{Q}\tilde{H}) \\
&= \mathrm{tr}(\tilde{H}^2) + \mathrm{tr}(H^2) - 2\mathrm{Re}\,\mathrm{tr}(V\Sigma V^HQ^H\tilde{Q}\tilde{V}\Sigma\tilde{V}^H) \\
&= \mathrm{tr}(\tilde{H}^2) + \mathrm{tr}(H^2) - 2\mathrm{Re}\,\mathrm{tr}(\Sigma X^H\tilde{\Sigma}W).
\end{aligned}$$

By Lemma 2.3, we get

$$\begin{aligned}
\|\tilde{A} - A\|_F^2 &\geq \mathrm{tr}(\tilde{H}^2) + \mathrm{tr}(H^2) - 2|\mathrm{tr}(\Sigma X^H\tilde{\Sigma}W)| \\
&\geq \mathrm{tr}(\tilde{H}^2) + \mathrm{tr}(H^2) - \mathrm{Re}\,\mathrm{tr}(\Sigma W^H\tilde{\Sigma}W) - \sum_{i=1}^n \sigma_i\tilde{\sigma}_i \\
&= \frac{1}{2}\|\tilde{H} - H\|_F^2 + \frac{1}{2}\|\tilde{\Sigma} - \Sigma\|_F^2 \geq \frac{1}{2}\|\tilde{H} - H\|_F^2.
\end{aligned}$$

It is easy to see that (2.8) is ture.

§3. Final Remarks

The perturbation bounds for the polar factors of column full-rank matrices are given by (2.7) and (2.8). We can see that (2.7) is a generalization of (1.5), and (2.7) and (2.8) are the generalizations and improvements of (1.4) and (1.3), respectively.

The polar decomposition is a generalization to matrices of the complex number representation $z = re^{i\theta}, r \geq 0$. For complex numbers $z = re^{i\theta}$ and $\tilde{z} = \tilde{r}e^{i\tilde{\theta}}$, we have

$$|e^{i\tilde{\theta}} - e^{i\theta}| \leq \frac{2}{r}|\tilde{z} - z|, \tag{3.1}$$

$$|\tilde{r} - r| \leq |\tilde{z} - z|. \tag{3.2}$$

It is easy to see that (2.7) is a generalization of (3.1). Now we give an example to show that $\|\tilde{H} - H\|_F \leq \|\tilde{A} - A\|_F$, as a generalization of (3.2), is not always true.

Example.

$$A = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}, \quad \tilde{A} = \frac{1}{25}\begin{pmatrix} 48 & -39 \\ -11 & 48 \end{pmatrix}.$$

It is easy to know that

$$H = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}, \quad \tilde{H} = \frac{1}{25}\begin{pmatrix} 43 & -24 \\ -24 & 57 \end{pmatrix},$$

$$\|\tilde{H} - H\|_F = \left\|\frac{1}{25}\begin{pmatrix} -32 & -24 \\ -24 & 32 \end{pmatrix}\right\|_F = \frac{8\sqrt{2}}{5}$$

and

$$\|\tilde{A} - A\|_F = \left\|\frac{1}{25}\begin{pmatrix} -27 & -39 \\ -11 & 23 \end{pmatrix}\right\|_F = \frac{2\sqrt{29}}{5}.$$

Obviously, $\|\tilde{H} - H\|_F > \|\tilde{A} - A\|_F$.

References

[1] N.J.Higham, Computing the polar decomposition–with applications, *SIAM J. Sci. Stat. Comput.*, **7** : 4 (1986), 1160–1174.

[2] Mao Jian-qin, The perturbation analysis of the product of singular vector matrices UV^T, *J. Comp. Math.*, **4** : 3 (1986), 245–248.

[3] Ni Guo-xi, The Common Theory and Method of Matrices (in Chinese), Shanghai Science and Technology Press, 1984.

[4] Sun Ji-guang, On the perturbation of the eigenvalues of a normal matrix, *Math. Numer. Sinica*, **6** : 3 (1984), 334–336.