

NEW PERTURBATION BOUNDS FOR THE UNITARY POLAR FACTOR*

REN-CANG LI†

Abstract. Let A be an $m \times n$ ($m \geq n$) complex matrix. It is known that there is a unique polar decomposition $A = QH$, where $Q^*Q = I$, the $n \times n$ identity matrix, and H is positive definite, provided A has full column rank. This note addresses the following question: How much may Q change if A is perturbed? For the square case $m = n$ our bound, which is valid for any unitarily invariant norm, is sharper and simpler than that of Mathias [*SIAM J. Matrix Anal. Appl.*, 14 (1993), pp. 588-597]. For the nonsquare case, a bound is also established for unitarily invariant norm, which has not been done in the literature.

Key words. polar decomposition, perturbation bound, unitarily invariant norm

AMS subject classifications. 15A12, 15A18, 15A23, 15A45

Let A be an $m \times n$ ($m \geq n$) complex matrix. It is known that there are Q with orthonormal column vectors, i.e., $Q^*Q = I$, and a unique positive semidefinite H such that

$$(1) \quad A = QH.$$

Hereafter I denotes an identity matrix with appropriate dimensions that are either specified or that are clear from the context. The decomposition (1) is called the polar decomposition of A . If, in addition, A has full column rank, then Q is uniquely determined also. In fact,

$$(2) \quad H = (A^*A)^{1/2}, \quad Q = A(A^*A)^{-1/2},$$

where the superscript $*$ denotes conjugate transpose. The decomposition (1) can also be computed from the singular value decomposition (SVD) $A = U\Sigma V^*$ by

$$(3) \quad H = V\Sigma_1V^*, \quad Q = U_1V^*,$$

where $U = (U_1, U_2)$ and V are unitary, U_1 is $m \times n$, $\Sigma = \begin{pmatrix} \Sigma_1 \\ 0 \end{pmatrix}$ and $\Sigma_1 = \text{diag}(\sigma_1, \dots, \sigma_n)$ is nonnegative.

There are published bounds stating how much the two factor matrices Q and H may change if entries of A are perturbed. Among the papers written on this subject are [1], [3], [4], [6]-[10], the perturbation bounds for Q when $m = n$ proved by Mathias [9], cover every unitarily invariant norm, while others cover the Frobenius norm only. Chen and Sun [6], [3] and Li [8] also deal with the case $m \geq n$ as we do here. A surprise is how heavily the sensitivity of the Q factor depends upon whether the working number field is real or complex [1], [7], [9].

In this paper, we obtain some bounds for the perturbations of Q , assuming A is complex. Our bound for the case $m = n$ is achievable and improves on that of Mathias slightly for small perturbations and significantly for big ones.

For the sake of convenience in our presentation, we use A and \tilde{A} for two matrices having full column rank, one of which is a perturbation of the other. Let

$$(4) \quad A = QH, \quad \tilde{A} = \tilde{Q}\tilde{H}$$

* Received by the editors September 7, 1993; accepted for publication (in revised form) by N. J. Higham, November 9, 1993.

† Department of Mathematics, University of California at Berkeley, Berkeley, California 94720 (li@math.berkeley.edu).

be the polar decompositions of A and \tilde{A} , respectively, and let

$$(5) \quad A = U\Sigma V^*, \quad \tilde{A} = \tilde{U}\tilde{\Sigma}\tilde{V}^*$$

be the SVDs of A and \tilde{A} , respectively, where $\tilde{U} = (\tilde{U}_1, \tilde{U}_2)$, \tilde{U}_1 is $m \times n$, and $\tilde{\Sigma} = \begin{pmatrix} \tilde{\Sigma}_1 \\ 0 \end{pmatrix}$ and $\tilde{\Sigma}_1 = \text{diag}(\tilde{\sigma}_1, \dots, \tilde{\sigma}_n)$. Assume as usual that

$$(6) \quad \sigma_1 \geq \dots \geq \sigma_n > 0 \quad \text{and} \quad \tilde{\sigma}_1 \geq \dots \geq \tilde{\sigma}_n > 0.$$

It follows from (2) and (5) that

$$Q = U_1V^*, \quad \tilde{Q} = \tilde{U}_1\tilde{V}^*.$$

In what follows, $\|X\|_2$ denotes the spectral norm which is the biggest singular value of X and $\|X\|_F$ the Frobenius norm which is the square root of the trace of X^*X . We shall use $\|\cdot\|$ to denote a general unitarily invariant norm [5], [11]. Two particular ones are $\|\cdot\|_2$ and $\|\cdot\|_F$. Consider

$$(7) \quad \begin{aligned} \|A - \tilde{A}\| &= \|U^*(A - \tilde{A})\tilde{V}\| = \|\Sigma V^*\tilde{V} - U^*\tilde{U}\tilde{\Sigma}\| \\ (8) \quad &= \|\tilde{U}^*(\tilde{A} - A)V\| = \|\tilde{\Sigma}\tilde{V}^*V - \tilde{U}^*U\Sigma\|. \end{aligned}$$

Define

$$(9) \quad E \stackrel{\text{def}}{=} \Sigma V^*\tilde{V} - U^*\tilde{U}\tilde{\Sigma} \quad \text{and}$$

$$(10) \quad \tilde{E} \stackrel{\text{def}}{=} \tilde{\Sigma}\tilde{V}^*V - \tilde{U}^*U\Sigma$$

to infer from (7) and (8) that

$$(11) \quad \|E\| = \|\tilde{E}\| = \|A - \tilde{A}\|.$$

Notice that by (9) and (10)

$$\begin{aligned} (I, 0)E &= \Sigma_1 V^*\tilde{V} - U_1^*\tilde{U}_1\tilde{\Sigma}_1 \quad \text{and} \\ (I, 0)\tilde{E} &= \tilde{\Sigma}_1\tilde{V}^*V - \tilde{U}_1^*U_1\Sigma_1, \end{aligned}$$

where I is $n \times n$. Adding the conjugate transpose of the second equation to the first yields

$$(12) \quad \Sigma_1(V^*\tilde{V} - U_1^*\tilde{U}_1) + (V^*\tilde{V} - U_1^*\tilde{U}_1)\tilde{\Sigma}_1 = (I, 0)E + \tilde{E}^* \begin{pmatrix} I \\ 0 \end{pmatrix}.$$

This is our perturbation equation to derive our perturbation bounds for Q because for any unitarily invariant norm $\|\cdot\|$,

$$(13) \quad \|V^*\tilde{V} - U_1^*\tilde{U}_1\| = \|I - VU_1^*\tilde{U}_1\tilde{V}^*\| = \|I - Q^*\tilde{Q}\|.$$

We shall use Lemma 1, which is a special case of Davis and Kahan [2, Thm. 5.2].

LEMMA 1. *Let M and N be two Hermitian matrices and let S be a complex matrix with suitable dimensions. Suppose there are two disjoint intervals separated by a gap of width at least η , one of which contains the spectrum of M and the other*

contains that of N . If $\eta > 0$, then there is a unique solution X to the matrix equation $MX - XN = S$, and moreover $\|X\| \leq \frac{1}{\eta} \|S\|$ for every unitarily invariant norm $\|\cdot\|$.

Applying this lemma to (11), (12), and (13) with $M = \Sigma_1$, $N = -\tilde{\Sigma}_1$ and $X = V^*\tilde{V} - U_1^*\tilde{U}_1$ yields Lemma 2.

LEMMA 2. *It holds that*

$$(14) \quad \left\| I - Q^*\tilde{Q} \right\| \leq \frac{2}{\sigma_n + \tilde{\sigma}_n} \left\| A - \tilde{A} \right\|.$$

When $m = n$, both Q and \tilde{Q} are unitary. Thus $\|I - Q^*\tilde{Q}\| = \|Q - \tilde{Q}\|$, and Lemma 2 yields the following theorem.¹

THEOREM 1. *Let A and \tilde{A} be two $n \times n$ nonsingular complex matrices whose polar decompositions are given by (4), and let σ_n and $\tilde{\sigma}_n$ be the smallest singular values of A and \tilde{A} , respectively. Then*

$$(15) \quad \left\| Q - \tilde{Q} \right\| \leq \frac{2}{\sigma_n + \tilde{\sigma}_n} \left\| A - \tilde{A} \right\|.$$

If, however, $m > n$, then it follows from (9) and (10) that

$$\begin{aligned} (0, I)E &= -U_2^*\tilde{U}_1\tilde{\Sigma}_1 \quad \text{and} \\ (0, I)\tilde{E} &= -\tilde{U}_2^*U_1\Sigma_1, \end{aligned}$$

where I is $(m - n) \times (m - n)$. Therefore

$$\left\| U_2^*\tilde{U}_1 \right\| \leq \left\| -U_2^*\tilde{U}_1\tilde{\Sigma}_1 \right\| \left\| \tilde{\Sigma}_1^{-1} \right\|_2 \leq \frac{\left\| (0, I)E \right\|}{\tilde{\sigma}_n} \leq \frac{\left\| A - \tilde{A} \right\|}{\tilde{\sigma}_n}.$$

Similarly,

$$\left\| \tilde{U}_2^*U_1 \right\| \leq \frac{\left\| (0, I)\tilde{E} \right\|}{\sigma_n} \leq \frac{\left\| A - \tilde{A} \right\|}{\sigma_n}.$$

Notice that $(U_1V^*, U_2) = (Q, U_2)$ and $(\tilde{U}_1\tilde{V}^*, \tilde{U}_2) = (\tilde{Q}, \tilde{U}_2)$ are unitary. Hence $U_2^*Q = 0$ and

$$\begin{aligned} (16) \quad \left\| Q - \tilde{Q} \right\| &= \left\| (Q, U_2)^*(Q - \tilde{Q}) \right\| = \left\| \begin{pmatrix} I - Q^*\tilde{Q} \\ -U_2^*\tilde{Q} \end{pmatrix} \right\| \\ &\leq \left\| I - Q^*\tilde{Q} \right\| + \left\| -U_2^*\tilde{U}_1\tilde{V}^* \right\| \\ &= \left\| I - Q^*\tilde{Q} \right\| + \left\| U_2^*\tilde{U}_1 \right\| \\ &\leq \left(\frac{2}{\sigma_n + \tilde{\sigma}_n} + \frac{1}{\tilde{\sigma}_n} \right) \left\| A - \tilde{A} \right\|. \end{aligned}$$

¹ Professor R. Bhatia kindly pointed out to me that Theorem 1 would be true in infinite dimensions. That is because of the infinite dimensional version of Lemma 1 in [2]. In the infinite dimensional version of the inequality (15), σ_n and $\tilde{\sigma}_n$ should be replaced by $\|A^{-1}\|^{-1}$ and $\|\tilde{A}^{-1}\|^{-1}$, respectively, where $\|\cdot\|$ is the operator norm in the Hilbert space where A and \tilde{A} live.

Similarly, we can prove

$$(17) \quad \|Q - \tilde{Q}\| \leq \left(\frac{2}{\sigma_n + \tilde{\sigma}_n} + \frac{1}{\sigma_n} \right) \|A - \tilde{A}\|.$$

Therefore, generally, we have Theorem 2.

THEOREM 2. *Let A and \tilde{A} be two $m \times n$ ($m > n$) complex matrices having full column rank and with the polar decompositions (4), and let σ_n and $\tilde{\sigma}_n$ be the smallest singular values of A and \tilde{A} , respectively. Then*

$$(18) \quad \|Q - \tilde{Q}\| \leq \left(\frac{2}{\sigma_n + \tilde{\sigma}_n} + \frac{1}{\max\{\sigma_n, \tilde{\sigma}_n\}} \right) \|A - \tilde{A}\|.$$

Estimates (16) and (17) can be sharpened a little bit when $\|\cdot\| = \|\cdot\|_F$. As a matter of fact, we shall have

$$\begin{aligned} \|Q - \tilde{Q}\|_F &= \sqrt{\|I - Q^* \tilde{Q}\|_F^2 + \|U_2^* \tilde{U}_1\|_F^2} \\ &\leq \sqrt{\left(\frac{2}{\sigma_n + \tilde{\sigma}_n} \right)^2 + \frac{1}{\tilde{\sigma}_n^2}} \|A - \tilde{A}\|_F \quad \text{and} \\ \|Q - \tilde{Q}\|_F &\leq \sqrt{\left(\frac{2}{\sigma_n + \tilde{\sigma}_n} \right)^2 + \frac{1}{\sigma_n^2}} \|A - \tilde{A}\|_F. \end{aligned}$$

A consequence of these two inequalities is the following theorem.

THEOREM 3. *Under the conditions of Theorem 2,*

$$(19) \quad \|Q - \tilde{Q}\|_F \leq \sqrt{\left(\frac{2}{\sigma_n + \tilde{\sigma}_n} \right)^2 + \left(\frac{1}{\max\{\sigma_n, \tilde{\sigma}_n\}} \right)^2} \|A - \tilde{A}\|_F.$$

We conclude this paper with a few remarks.

Remark 1. The bound in (15) is the best possible, in the sense that the equality can be achieved. Take the following case for an example: Both A and \tilde{A} are $n \times n$ unitary matrices. Thus $\sigma_n = \tilde{\sigma}_n = 1$, $Q = A$, $\tilde{Q} = \tilde{A}$, and

$$\|Q - \tilde{Q}\| = \frac{2}{\sigma_n + \tilde{\sigma}_n} \|A - \tilde{A}\|.$$

It is even achievable in the real number field by taking A and \tilde{A} to be two $n \times n$ orthogonal matrices although, as we know, Q behaves quite differently in the real number field (Remark 5). All previously published bounds do not achieve this!

Remark 2. Bounds (15), (18), and (19) involve both σ_n and $\tilde{\sigma}_n$. To obtain bounds involving σ_n alone, one can weaken them by utilizing the following fact:

$$\|A - \tilde{A}\|_2 \geq |\sigma_n - \tilde{\sigma}_n|.$$

For example, (15) yields

$$(20) \quad \|Q - \tilde{Q}\| \leq \frac{2}{2\sigma_n - \|A - \tilde{A}\|_2} \|A - \tilde{A}\|,$$

provided $\|A - \tilde{A}\|_2 < 2\sigma_n$.

Remark 3. Mathias [9] proved that for $m = n$ if $\|A - \tilde{A}\|_2 < \sigma_n$, then

$$(21) \quad \|Q - \tilde{Q}\| \leq -\frac{\|A - \tilde{A}\|}{\|A - \tilde{A}\|_2} \times \ln \left(1 - \frac{\|A - \tilde{A}\|_2}{\sigma_n} \right).$$

Although his bound uses slightly different information than ours, it is always a bigger, sometimes much bigger, bound than (15) (since the left-hand side of (21) could blow up). To see why this is true, we claim that even (20), the weakened form of (15), is still no weaker than that of Mathias because their ratio (his/ours) is

$$-\frac{\ln(1-x)}{x} \cdot \left(1 - \frac{x}{2}\right) = 1 + \sum_{j=2}^{\infty} \left(\frac{1}{j+1} - \frac{1}{2j}\right) x^j > 1$$

for $0 < x = \|A - \tilde{A}\|_2 / \sigma_n < 1$.

Remark 4. Chen and Sun [6] studied the case $m > n$, also. But only the Frobenius norm was considered. They proved

$$(22) \quad \|Q - \tilde{Q}\|_F \leq \frac{2}{\sigma_n} \|A - \tilde{A}\|_F.$$

Without loss of generality, assume $\tilde{\sigma}_n \leq \sigma_n$. Then it is easy to see our bound (19) is sharper than (22) when

$$\tilde{\sigma}_n \leq \sigma_n \leq \frac{\sqrt{3}}{2 - \sqrt{3}} \tilde{\sigma}_n \approx 6.5 \tilde{\sigma}_n;$$

otherwise (22) is a little sharper because

$$\sqrt{\left(\frac{2}{\sigma_n + \tilde{\sigma}_n}\right)^2 + \left(\frac{1}{\sigma_n}\right)^2} \leq \frac{\sqrt{5}}{\sigma_n} \approx \frac{2.2}{\sigma_n}$$

always. More generally, Sun and Chen [3] and Li [8] treated the cases when A and \tilde{A} do not necessarily have full column rank. Applied to our full column rank case here, the perturbation bound for the polar factor in [3] reads exactly the same as (22), and in [8] reads

$$\|Q - \tilde{Q}\|_F \leq \frac{1}{\min\{\sigma_n, \tilde{\sigma}_n\}} \|A - \tilde{A}\|_F,$$

which is clearly sharper than (19) and (22) when $\sigma_n \approx \tilde{\sigma}_n$. However, it may be very bad if one of σ_n and $\tilde{\sigma}_n$ is much smaller than the other.

Remark 5. Perturbation bounds for the Q factor in polar decomposition illustrate that the change in Q is proportional to the reciprocal of the smallest singular value of A when $m = n$ and when the working number field is complex. However, it was discovered by Barrlund [1], Kenney and Laub [7], and Mathias [9] that for the real case the change in Q is proportional to the reciprocal of the sum of the two smallest singular values of A if $m = n$, which means Q is (much) less sensitive to perturbations in A

in the real case than in the complex case. The above derivation of the perturbation bound (15) for the complex case is very elementary while giving the best among the derivations that have been published. However the author was unable to extend this derivation to perform for the real case. It is worth stating (as pointed out by one of the anonymous referees) that even in the real number field when $m > n$, the change in Q is not proportional to $1/(\sigma_{n-1} + \sigma_n)$ instead of $1/\sigma_n$. The following example offered by the referee makes this point very clear:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0.8 \times 10^{-6} \\ 0 & 0 \end{pmatrix}, \quad Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix};$$

$$\tilde{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0.8 \times 10^{-6} \\ 0 & 0.6 \times 10^{-6} \end{pmatrix}, \quad \tilde{Q} = \begin{pmatrix} 1 & 0 \\ 0 & 0.8 \\ 0 & 0.6 \end{pmatrix}.$$

Acknowledgments. The author is grateful for the encouragement of Professors W. Kahan and J. Demmel. He thanks Dr. N. J. Higham for his valuable comments concerning the manuscript. He is indebted to the referees for their many helpful suggestions, especially the last part of Remark 5 which he had not previously addressed.

REFERENCES

- [1] A. BARRLUND, *Perturbation bounds on the polar decomposition*, BIT, 31 (1990), pp. 101–113.
- [2] C. DAVIS AND W. M. KAHAN, *The rotation of eigenvectors by a perturbation. iii*, SIAM J. Numer. Anal., 1 (1970), pp. 1–46.
- [3] J.-G. SUN AND C. HUI CHEN, *Generalized polar decomposition*, Math. Numer. Sinica, 11 (1989), pp. 262–273. (In Chinese.)
- [4] N. J. HIGHAM, *Computing the polar decomposition—with applications*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 1160–1174.
- [5] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.
- [6] C. HUI CHEN AND J. GUANG SUN, *Perturbation bounds for the polar factors*, J. Comput. Math., 7 (1989), pp. 397–401.
- [7] C. KENNEY AND A. J. LAUB, *Polar decomposition and matrix sign function condition estimates*, SIAM J. Sci. Statist. Comput., 12 (1991), pp. 488–504.
- [8] R.-C. LI, *A perturbation bound for the generalized polar decomposition*, BIT, 34 (1993), pp. 304–308.
- [9] R. MATHIAS, *Perturbation bounds for the polar decomposition*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 588–597.
- [10] J. QIN MAO, *The perturbation analysis of the product of singular vector matrices uv^h* , J. Comput. Math., 4 (1986), pp. 245–248.
- [11] G. W. STEWART AND J.-G. SUN, *Matrix Perturbation Theory*, Academic Press, New York, 1990.