

ON OPTIMAL SYMMETRIC ORTHOGONALISATION AND SQUARE ROOTS OF A NORMAL MATRIX

NAGWA SHERIF

It is well known that the factors in the polar decomposition of a full rank real $m \times n$ matrix, $m \geq n$ possess best approximation properties. We propose an iterative technique to compute the polar factors based on these best approximation properties. For normal matrices, the polar decomposition is useful. It is applied to compute the principal square roots of real and complex normal matrices.

1. INTRODUCTION

In view of the best approximation properties of the polar factors of a matrix, techniques for computing the polar decomposition are of interest. While the factors in the polar decomposition can be obtained via the singular value decomposition, this approach is not always the most efficient or the most convenient.

In this paper, optimal properties of the polar factors are described in Section 2. In Section 3, we present and analyse a one-sided Jacobi type method for computing the polar factors which involves only matrix multiplication. Generally, the algorithm consists of minimising the Frobenius norm of $A - V$ at each stage, where A is a given matrix and V is an orthogonal matrix. We illustrate the performance of the algorithm by some numerical examples. We end this section by reporting other available methods for computing the polar decomposition.

In Section 4, we propose a procedure to compute the principal square root of a real normal matrix, if one exists, in real arithmetic. This will be an application of the algorithm developed in the previous section. Finally, we show how to modify this procedure to compute a principal square root of a normal complex matrix A .

2. POLAR DECOMPOSITION.

In this section, we describe the polar decomposition of a matrix. Let $A \in \mathbb{R}^{m \times n}$, $m \geq n$. Then there exists a matrix $B \in \mathbb{R}^{m \times n}$, and a unique symmetric positive semidefinite matrix $H \in \mathbb{R}^{n \times n}$, such that

$$(2.1) \quad A = BH \quad \text{and} \quad B^T B = I.$$

Received 9 March 1992

Copyright Clearance Centre, Inc. Serial-fee code: 0004-9729/93 \$A2.00+0.00.

If A has a full rank, then H is positive definite and B is uniquely determined. The above decomposition is well known factorisation of the matrix A , [6]. This factorisation can be derived from the singular value decomposition of A . The singular value decomposition insures that there exists an $m \times n$ orthonormal matrix Q , an $n \times n$ orthogonal matrix P and an $n \times n$ diagonal matrix D with non-negative diagonal elements, called singular values of A , such that

$$(2.2) \quad A = Q\Sigma P^T, \quad \Sigma = [D, 0]^T$$

where

$$D = \text{diag}(S_1(A), S_2(A), \dots, S_n(A)) \quad \text{and} \quad S_1(A) \geq S_2(A) \geq \dots \geq S_n(A) \geq 0.$$

The fact that (2.1) can be derived from (2.2) follows by partitioning Q , as $Q = [Q_1, Q_2]$, $Q_1 \in \mathbb{R}^{m \times n}$, $Q_1^T Q_1 = I$. Now let

$$(2.3) \quad B = Q_1 P^T \quad \text{and} \quad H = P D P^T.$$

Then we have $BH = Q_1 D P^T = [Q_1, Q_2][D, 0]^T P^T = Q\Sigma P^T = A$.

In the sequel, the following notation will be used: $|A| = (A^T A)^{1/2}$ denotes the unique positive semidefinite square root of $A^T A$. If A is an $n \times n$ matrix, the eigenvalues of A will be denoted by $\lambda_j(A)$ ($j = 1, 2, 3, \dots, n$) with multiplicity counted, $\sigma(A)$ denotes the set of eigenvalues of A . The singular values of A (the eigenvalues of $|A|$) will always be counted with multiplicities and numbered in decreasing order: $S_1(A) \geq S_2(A) \dots \dots S_n(A) \geq 0$. A norm on $\mathbb{R}^{m \times n}$ which satisfies the additional condition $\|QAP\| = \|A\|$, where $Q \in \mathbb{R}^{m \times m}$ and $P \in \mathbb{R}^{n \times n}$ are orthogonal matrices, is called a unitarily invariant norm. Examples of unitarily invariant norms are Ky Fan k -norms, defined for $k = 1, 2, 3, \dots, n$ by

$$\|A\|_k = \sum_{j=1}^k S_j(A).$$

Let $X, Y \in \mathbb{R}^{n \times n}$. It is well known [12] that a sufficient and a necessary condition for $\|X\| \leq \|Y\|$ in any unitarily invariant norm is

$$(2.4) \quad \|X\|_k \leq \|Y\|_k, \quad k = 1, 2, 3, \dots, n.$$

An extension to the case $X, Y \in \mathbb{R}^{m \times n}$ is straight forward. A particular unitarily invariant norm is the Frobenius norm

$$\|A\|_F = (\text{tr}(A^T A))^{1/2} = \left(\sum_{j=1}^n (S_j(A))^2 \right)^{1/2} = \left(\sum_{j=1}^n \sum_{i=1}^m a_{ij}^2 \right)^{1/2}.$$

For normal matrices the polar factors possess additional properties, which characterise the normality of the matrix.

THEOREM 2.1. *Let $A \in \mathbb{R}^{n \times n}$ have a polar decomposition $A = BH$. Then the following statements are equivalent:*

- (a) A is normal.
- (b) $BH = HB$.
- (c) $AB = BA$.
- (d) $AH = HA$.
- (e) There exists a permutation i_1, i_2, \dots, i_n of the numbers $1, 2, 3, \dots, n$ such that

$$\lambda_j(A) = \lambda_j(B)\lambda_{i_j}(H) \quad \text{for } j = 1, 2, 3, \dots, n.$$

- (f) modulus $(\sigma(A)) = \{\lambda_1(H), \dots, \lambda_n(H)\}$.
- (g) argument $[\sigma(A)] = \{\lambda_1(B), \dots, \lambda_n(B)\}$, (as long as A is non-singular).

PROOF: Since A is normal, we have $A^T A = A A^T$, which implies that B commutes with H^2 . This implies that B commutes with H , since H is positive semidefinite. If $BH = HB$, then A is normal, being the product of two commuting normal matrices [7]. This establishes the equivalence of (a) and (b). Clearly, (b) is equivalent to each of the conditions (c) and (d). A being normal implies (e). This obviously implies both (f) and (g). That (g) implies (a) appears in [7]. □

The optimality of the polar factors is described by the following theorem.

THEOREM 2.2. *Let $A \in \mathbb{R}^{m \times n}$ have a polar decomposition $A = BH$. Then the following statements are true:*

- (a) $\|A - B\| = \min_{V \in \mathcal{V}} \|A - V\|$, where \mathcal{V} is the subset of orthonormal matrices of $\mathbb{R}^{m \times n}$. The result is true in any unitarily invariant norm.
- (b) If $A = B_1 H_1$, B_1 is orthonormal and H_1 is symmetric then $H_1 \leq H$.

PROOF: (a) The case $m = n$ was proved in [13]. An extension to the case $m > n$ goes in similar fashion. Our aim to show that for $k = 1, 2, 3, \dots, n$

$$(2.5) \quad \|A - B\|_k \leq \|A - V\|_k$$

hence the result will follow because of (2.4). Now $S_j(A - B) = S_j(H - I) = |\lambda_j(H - I)|$. Let $\mu_1 \geq \dots \geq \mu_n$ be the eigenvalues of H with multiplicity counted. Then by [12] we have

$$(2.6) \quad \sum_{j=1}^k S_j(A - B) = \max_{j_1 < j_2 < \dots < j_k} \sum_{i=1}^k |\mu_{j_i} - 1|.$$

We define for any matrix $X \in \mathbb{R}^{m \times n}$ the symmetric matrix $\tilde{X} = \begin{bmatrix} 0 & X \\ X^T & 0 \end{bmatrix}$ of order $m + n$. It is easy to see that the eigenvalues of \tilde{X} are precisely the singular values of

X , with their negatives and $m - n$ zeros. Using singular value decomposition (2.2) of A we have $\|A - V\| = \|\Sigma - U\|$, $U = Q^TVP$. Let $\tilde{X} = \tilde{\Sigma} - \tilde{U}$. The eigenvalues of \tilde{X} , $\tilde{\Sigma}$ and \tilde{U} are

$$S_1(A - V), \dots, S_n(A - V), 0, \dots, 0, -S_n(A - V), \dots, -S_1(A - V), \\ \mu_1, \mu_2, \dots, \mu_n, 0, 0, \dots, 0, -\mu_n, \dots, -\mu_1, 1, 1, \dots, 1, 0, \dots, 0, -1, -1, \dots, -1.$$

respectively. Again from [12] we have

$$(2.7) \quad \max_{j_1 < j_2 < \dots < j_k} \sum_{i=1}^k |\mu_{j_i} - 1| \leq \sum_{i=1}^k S_i(A - V).$$

Now (2.6) and (2.7) imply (2.4).

(b) This is immediate, since $H_1^2 = H^2$, and hence $|H_1| = H$. But it is always true that $H_1 \leq |H_1|$, so $H_1 \leq H$. □

REMARK 2.1. The symmetric polar factor has best approximation properties in the particular case when $A \in \mathbb{R}^{n \times n}$ is symmetric [8, 9].

In the next section, an algorithm for computing the polar decomposition of $A \in \mathbb{R}^{n \times n}$ is described. The general case of $A \in \mathbb{R}^{m \times n}$ can be reduced to this case by performing a QR factorisation step and applying the algorithm to the factor R .

3. POLAR DECOMPOSITION ALGORITHM

Let $A \in \mathbb{R}^{n \times n}$ be a non-singular matrix having polar decomposition $A = BH$. We construct the one-sided Jacobi type method to compute the polar factors B and H . The technique is based on Theorem (2.2) (a), and we call it the polar decomposition algorithm. Let J_n denote the plane rotations. For integers i, j ($1 \leq i < j \leq n$) and real θ define the orthogonal matrix $R(i, j, \theta) = [R_{rs}]$;

$$R_{rs} = \begin{cases} \cos \theta & \text{if } (r, s) = (i, i) \text{ or } (j, j) \\ \sin \theta & \text{if } (r, s) = (i, j) \\ -\sin \theta & \text{if } (r, s) = (j, i) \\ \delta_{rs} & \text{for all other } (r, s), \delta_{rs} \text{ the Kronecker delta.} \end{cases}$$

In order to implement the minimising procedure (Theorem (2.2) (a)) we construct a sequence of orthogonal matrices $\{B_k\}$ and a sequence of matrices $\{H_k\}$, where

$$(3.1) \quad H_{k+1} = R_k H_k, H_0 = A, B_{k+1} = B_k R_k^T, B_0 = I, R_k = R_k(i_k, j_k, \theta_k).$$

Here $R_k \in J_n$ ($k = 0, 1, \dots$), with parameters to be determined. Let us write

$$H_k = [h_{[1]}^{(k)}, h_{[2]}^{(k)}, \dots, h_{[n]}^{(k)}]^T$$

where $h_{[i]}^{(k)}$ denotes the i th row of H_k . Let $\{\sigma_k\} = \{(i_k, j_k)\}$. This is called the sequence of indexed pairs. We drop the subscript from the pair and write $\sigma_k = (i, j)$, and it is understood that (i, j) is to be determined in the k th step by some strategy considered later. In order to determine $R(i, j, \theta_k)$ at the k th step in our minimising procedure let us define

$$(3.2) \quad \Omega(i, j, \theta) = \|A - B_k\|_F^2 - \|A - B_{k+1}\|_F^2 = 2tr(H_{k+1} - H_k) = 2tr((R_k - I)H_k).$$

We define

$$(3.3) \quad \Omega_{ij}^{(k)} = \Omega(i, j, \theta_k) = \max_{\theta} \Omega(i, j, \theta),$$

where the index pair $\sigma_k(i, j)$ is already chosen. In the k th step, θ_k will be called the angle of rotation, $\sigma_k = (i, j)$ is the index pair for rotation. Since H_k is premultiplied by R_k , only the i th and j th rows are affected, indeed

$$(3.4) \quad h_{[i]}^{(k+1)} = \cos \theta h_{[i]}^{(k)} + \sin \theta h_{[j]}^{(k)}, \quad h_{[j]}^{(k+1)} = -\sin \theta h_{[i]}^{(k)} + \cos \theta h_{[j]}^{(k)}.$$

Let us write

$$\mu_{ij}^{(k)} = h_{ii}^{(k)} + h_{jj}^{(k)} \quad \text{and} \quad \nu_{ij}^{(k)} = h_{ji}^{(k)} - h_{ij}^{(k)}.$$

Consequently, it follows from (3.2) that

$$(3.5) \quad \Omega(i, j, \theta) = 2\{(\cos \theta - 1)\mu_{ij}^{(k)} + \sin \theta \nu_{ij}^{(k)}\}.$$

It can be checked that θ_k is defined by the following equations,

$$(3.6) \quad \eta_{ij}^{(k)} \cos \theta_k = \mu_{ij}^{(k)}, \quad \eta_{ij}^{(k)} \sin \theta_k = \nu_{ij}^{(k)}, \quad \eta_{ij}^{(k)} = \sqrt{[\mu_{ij}^{(k)}]^2 + [\nu_{ij}^{(k)}]^2}.$$

The following relations are easy to check,

$$(3.7) \quad \Omega_{ij}^{(k)} = 4\eta_{ij}^{(k)} \sin^2 \left(\frac{1}{2} \theta_k \right), \quad 4 \left| \nu_{ij}^{(k)} \right| = \Omega_{ij}^{(k)} [\Omega_{ij}^{(k)} + 4\mu_{ij}^{(k)}].$$

Let

$$(3.8) \quad \xi_k = tr(H_k).$$

It follows directly from equations (3.5) and (3.7) that $\{\xi_k\}$ is a monotone increasing sequence. Also, we have

$$(3.9) \quad \left\| h_{[i]}^{(k+1)} \right\| = \left\| h_{[i]}^{(k)} \right\|, \quad 1 \leq i \leq n.$$

Now we consider different strategies to choose the index pair $\sigma_k = (i, j)$ at each step k .

- (I) Process of maximising ν_{ij} : The index pair $\sigma_k = (i, j)$ is chosen from the condition

$$(3.10) \quad \left| \nu_{ij}^{(k)} \right| = \max_{r,s} \{ |\nu_{rs}| \}.$$

- (II) Process of maximising Ω_{ij} : The index pair $\sigma_k = (i, j)$ is chosen from the condition

$$(3.11) \quad \Omega_{ij}^{(k)} = \max_{r,s} \{ \Omega_{rs}^{(k)} \}.$$

- (III) Cyclic Process: A procedure is called cyclic if in every segment of $n(n - 1)/2$ consecutive elements of the sequence $\{\sigma_k\}$ every index pair (i, j) , $1 \leq i < j \leq n$, occurs exactly once. Such a segment of $n(n - 1)/2$ steps is called a sweep. In particular, we can define a cyclic process of row-wise scanning. This is one that will be implemented in the numerical examples given in this work.

Under most practical conditions, the sequence $\{B_k\}$ converges to B while $\{H_k\}$ converges to H . The next theorem discusses the convergence properties of the algorithm.

THEOREM 3.1. *Let the sequence of matrices $\{H_k\}$ and $\{B_k\}$ be generated by applying the polar decomposition algorithm to the matrix A , where A is non-singular. Let $H_k = S_k + L_k$, where S_k is symmetric and L_k is skew symmetric. Then for any strategy $L_k \rightarrow 0$ as $k \rightarrow \infty$.*

PROOF: The sequence $\{\xi_k\}$ defined in (3.8) is monotone increasing and bounded, hence convergent. That $\{\xi_k\}$ is bounded, can be seen from the inequality

$$\xi_k = \sum_{i=1}^n \lambda_i(H_k) \leq \sum_{i=1}^n |\lambda_i(H_k)| \leq \sum_{i=1}^n S_i(H_k);$$

for the last inequality, see [5, p.41]. Now $S_i(H_k) = S_i(A)$, $1 \leq i \leq n$, hence $\xi_k \leq \sum_{i=1}^n S_i(A) = \xi$. From equation (3.2) it follows that

$$(3.12) \quad \Omega_{ij}^{(k)} \rightarrow 0 \quad \text{as } k \rightarrow \infty,$$

for any process of choosing $\{\sigma_k\}$. The sequence $\{\mu_{ij}^{(k)}\}$ is bounded. Then for any strategy, (3.7) together with (3.12) gives

$$(3.13) \quad \nu_{ij}^{(k)} \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

To show that $L_k \rightarrow 0$, we prove it for the second strategy, that is, according to (3.11).

We define

$$(3.14) \quad \left| \nu_{uv}^{(k)} \right| = \max_{r,s} \left| \nu_{r,s}^{(k)} \right|.$$

Then we have

$$\mu_{uv}^{(k)} \leq \left\| h_{[u]}^{(k)} \right\| + \left\| h_{[v]}^{(k)} \right\| \leq 2 \|H_0\| = \zeta.$$

Equation (3.7) together with the above inequality give

$$\Omega_{ij}^{(k)} \geq \Omega_{uv}^{(k)} = 2 \left[\sqrt{[\mu_{uv}^{(k)}]^2 + [\nu_{uv}^{(k)}]^2} - \mu_{uv}^{(k)} \right] \geq 2 \left[\sqrt{\zeta^2 + [\nu_{uv}^{(k)}]^2} - \zeta \right].$$

Consequently, from (3.12) it follows that $\nu_{uv}^{(k)} \rightarrow 0$. By equation (3.14) we have $L_k \rightarrow 0$ as $k \rightarrow \infty$. □

REMARK 3.1. The sequence $\{H_k\}$ is bounded (equation (3.9)) and satisfies

$$\|H_{k+1} - H_k\| \leq 2S_1(A) \left| \sin \frac{1}{2} \theta_k \right|$$

where $\| \cdot \|$ denotes the operator matrix norm. It is now clear, using (3.7) and (3.12), that

$$(3.15) \quad \|H_{k+1} - H_k\| \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

Also we have $H_k^2 - A^T A = (H_k - H_k^T) H_k = 2L_k H_k$, hence

$$(3.16) \quad H_k^2 \rightarrow A^T A \quad \text{as } k \rightarrow \infty.$$

REMARK 3.2. Bounded sequences with property (3.15) were investigated in [1]. Applying that argument, it follows that $\{H_k\}'$, the derived set of $\{H_k\}$, is a compact connected set. In fact, $\{H_k\}'$ either consists of only one single point, and we have convergence in the usual sense, or it has uncountably many points. In case $\{H_k\}'$ consists of only one point, (3.16) implies that this limit point is a symmetric square root of $A^T A$. In all our numerical experiments, the sequence $\{H_k\}$ converges to the positive definite square root H .

The following lemma shows that the convergence of $\{H_k\}$ will be guaranteed under some assumptions on A .

LEMMA 3.1. Let $A \in \mathbb{R}^{n \times n}$ be non-singular such that $S_n(A) \geq \alpha > 0$. Let $\|A - \beta I\|_F < \sqrt{2}(\alpha + \beta)$, where $\beta = S_1(A)$. If the polar decomposition algorithm is applied to A , then the derived set of the sequence $\{H_k\}$ consists of the positive definite matrix H .

PROOF: It follows from (3.1) that

$$(3.17) \quad \|H_k - \beta I\|_F \leq \|H_0 - \beta I\|_F = \|A - \beta I\|_F.$$

Then, for $j = 1, \dots, n$, $|\lambda_j(H_k)| \geq S_n(H_k) = S_n(A) \geq \alpha$. Let H be a member of the derived set of $\{H_k\}$. Since $\det(H)$ is positive, the number of negative eigenvalues of H is even, and $\lambda_j(H)$ satisfies

$$|\lambda_j(H - \beta I)| \geq \alpha + \beta, \quad j = 1, 2, \dots, 2r.$$

Hence $\|H - \beta I\|_F \geq \left\{ \sum_{j=1}^{2r} |\lambda_j(H - \beta I)|^2 \right\}^{1/2} \geq \sqrt{2}(\alpha + \beta)\sqrt{r}$.

This contradicts (3.17). Then $r = 0$ and hence H is positive definite. □

The above method belongs to the one-sided Jacobi-type techniques. These techniques are appropriate to implement on the hyper cube or a linear array of processors [4]. Applications of the polar decomposition can be found in [15, 18]. In particular [9] gives a satisfactory presentation of the polar decomposition and its applications. In Section 4 we present another application.

The polar decomposition algorithm is stable, that is, rounding errors made in the actual multiplications do not lead to numerical instability. This is a well known behaviour of multiplications by rotations [21, 3 Section 24]. The cost of the polar decomposition algorithm is $2n^3$ per sweep. Further, the algorithm is easy to code and it does not require library routines (important in a microcomputer environment). Tests have shown that convergence is reached between 2-10 sweeps for well conditioned matrices. The algorithm converges when A has real negative eigenvalues with even multiplicities. However; if A has real negative eigenvalues with odd multiplicities, (that is, $\det(A) < 0$), then we have to modify the initial values. In this case we set $B_0 = D$ and $H_0 = DA$, where $D = \text{diag}(-1, 1, \dots)$. This will guarantee that $\det(H_k) > 0$ for all k , and consequently, $\{H_k\}$ will converge to the positive definite square root H .

As compared to [9, 15], the polar decomposition algorithm has the additional feature that it can be applied to singular matrices. However, for such matrices the convergence is generally slow as illustrated by the following example. Let

$$A = \begin{bmatrix} 1.50000 & 0.50000 & -0.50000 & -0.50000 \\ -0.50000 & 0.50000 & -0.50000 & -0.50000 \\ 0.50000 & -0.50000 & 0.50000 & 0.50000 \\ -0.50000 & 0.50000 & -0.50000 & 1.50000 \end{bmatrix}$$

After 8 sweeps we have the residual $E = A^T A - H^2$, with $\max |e_{ij}| = 1.19 E - 6$. The symmetric polar factor H is

$$H = \begin{bmatrix} 1.70711 & 0.00000 & 0.00000 & -0.29289 \\ 0.00000 & 0.70711 & -0.70711 & 0.00000 \\ 0.00000 & -0.70711 & 0.70711 & 0.00000 \\ -0.29289 & 0.00000 & 0.00000 & 1.70711 \end{bmatrix}$$

with $\sigma(H) = \{0, 1.414214, 2\}$, the required positive semidefinite polar factor.

The polar decomposition can be obtained via the singular value decomposition equations (2.2), (2.3). This is readily available in library routines such as LINPACK; [3]. It is numerically stable, and it requires approximately $12n^3$ flops. In [9, 15] quadratically convergent Newton-type iterations were suggested, with restrictions for numerical stability.

4. THE PRINCIPAL SQUARE ROOT OF A REAL NORMAL MATRIX

If A is a matrix, then a matrix X is said to be a square root of A in the case that $X^2 = A$. No useful necessary and sufficient condition for the existence of X is known. Even if X exists, it is not guaranteed to be a function of A . For the definition of a matrix function, see [14, 17]. In particular, an excellent investigation of real square roots functions of a real matrix A is presented in [11]. In this section, the existence of normal real square roots functions of a real normal matrix is investigated (Theorem 4.1). Then we propose an algorithm to compute the principal square root of a real normal matrix, if it exists, in real arithmetic. This algorithm does not require the computation of eigenvalues or eigenvectors. It is based on the application of the polar decomposition algorithm described in Section 3.

The existence of a real matrix X such that $X^2 = A$ can be seen from the real Schur decomposition of the matrix A , as given by the following Lemma [19].

LEMMA 4.1. *Let $A \in \mathbb{R}^{n \times n}$. There is an orthogonal matrix U such that $R = U^T A U$ is quasi-triangular. Moreover, U may be chosen so that any 2×2 diagonal block of R has only complex eigenvalues (which must therefore be conjugate).*

If in addition, A is normal with no real negative eigenvalues, then the quasi-triangular matrix R is block diagonal with diagonal blocks R_{ii} , $1 \leq i \leq r$, of the form

$$(4.1) \quad R_{ii} = \begin{bmatrix} a_i & -b_i \\ b_i & a_i \end{bmatrix}, \quad b_i \neq 0 \quad \text{and/or} \quad R_{ii} = f_i, \quad f_i \geq 0.$$

Now let S be the block diagonal matrix with diagonal blocks S_{ii} , $1 \leq i \leq r$,

$$S_{ii} = \pm \begin{bmatrix} c_i & -d_i \\ d_i & c_i \end{bmatrix}, \quad \text{and/or} \quad S_{ii} = \pm \sqrt{f_i}, \quad d_i = \left[\frac{b_i^2}{2[a_i^2 + \sqrt{a_i^2 + b_i^2}]} \right]^{1/2}, \quad c_i = \frac{1}{2d_i}.$$

It can be checked that each S_{ii} is a square root function of R_{ii} . So $S^2 = R$, and clearly S is normal. Setting $X = USU^T$ we have $X^2 = A$. So X is a square root function of A and is necessarily normal. We sum up these results in the following theorem. This theorem is still valid for a general real matrix $A \in \mathbb{R}^{n \times n}$, [11].

THEOREM 4.1. *Let $A \in \mathbb{R}^{n \times n}$ be a non-singular normal matrix. If A has real negative eigenvalues, then A has no real square roots which are functions of A . If A has no real negative eigenvalues, then there are precisely 2^{q+p} real square roots of A which are functions of A , where q is the number of distinct real eigenvalues, and p is the number of distinct complex eigenvalue pairs.*

We remark that the above theorem emphasises only the existence of square roots that are functions of A . If A has negative eigenvalues with even multiplicity, then A will have a real square root which is not a function of A . For example, $X = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, is a real normal square root of A , $A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$, but X is obviously not a function of A . For the matrix A described in Theorem 4.1, a particular real normal square root, which is computationally stable and optimal in some sense, is the principal square root of A .

DEFINITION: A real matrix N is said to be a principal square root of A if

- (a) $N^2 = A$.
- (b) $\text{Re } N = (N + N^T)/2$ is positive definite.

The following lemma characterises a principal square root of A . The proof follows from Putnam’s result [16].

LEMMA 4.2. *Let $A \in \mathbb{R}^{n \times n}$. There is a principal square root of A if and only if $N^2 = A$ and $\text{Re } N > 0$.*

The next result describes the optimality of a principal square root among other real normal square roots of A .

THEOREM 4.2. *Let $A \in \mathbb{R}^{n \times n}$ be a non-singular normal matrix with no real negative eigenvalues. Let N be a principal square root of A . Then the following statements are true:*

- (a) N is unique.
- (b) *If M is a real normal square root of A , then $MN = NM$ and $\|A - N\| \leq \|A - M\|$ in any unitarily invariant norm.*

PROOF: (a) Suppose L is a principal square root of A , then

$$(L + L^T)L = L^2 + L^T L = A + |A| = N^2 + N^T N = (N + N^T)N,$$

moreover,

$$(L + L^T)^2 = L^2 + 2LL^T + (L^T)^2 = A + 2|A| + A^T = (N + N^T)^2.$$

But $A + 2|A| + A^T$ is positive definite. The last equation implies $L + L^T = N + N^T$, since the positive definite square root of a positive definite matrix is unique. Consequently $L = N$.

(b) Let M be a real normal square root of A , then $MA = M^3 = AM$, hence, $p(A)M = Mp(A)$ for any polynomial $p(x)$. But N being the principal square root is represented by such a polynomial in A , hence $MN = NM$. To prove the required inequality, by equation (2.4) it is enough to prove it for the Ky Fan k -norms. Now A and M commute and consequently they have a common set of eigenvectors, hence

$$S_j(A - M) = |\lambda_j(A - M)| = |\lambda_j(A) - \lambda_j(M)|, \quad S_j(A - N) = |\lambda_j(A) - \lambda_j(N)|.$$

Since N is the principal square root of A , it follows that $S_j(A - N) \leq S_j(A - M)$, $1 \leq j \leq n$, and consequently $\|A - N\|_k \leq \|A - M\|_k$. □

The next theorem establishes the theoretical basis to compute the principal square root of a non-singular normal matrix.

THEOREM 4.3. *Let $A \in \mathbb{R}^{n \times n}$ be a non-singular normal matrix. Assume that A has no real negative eigenvalues and let $A = BH$ be the polar decomposition of A . Then the following statements are true:*

- (a) $I + B$ is non-singular; the orthogonal polar factor B_1 of $I + B$ is the principal square root of B .
- (b) If $H = LL^T$ is the Cholesky decomposition of H , and if $L^T = B_2H_2$ is the polar decomposition of L^T , then H_2 is the principal square root of H .
- (c) $N = B_1H_2$ is the principal square root of A .

PROOF: (a) If $A = BH$, then argument $\sigma(A) = \sigma(B)$ (Theorem 2.2), hence $-1 \notin \sigma(B)$, since A has no real negative eigenvalues. It follows that B is a rotation and $I + B$ is non-singular. Let $I + B$ have the polar decomposition, $I + B = B_1H_1$, where B_1 commutes with H_1 , since $I + B$ is normal. Now we have

$$(I + B)^2 = (B_1H_1)^2 = B_1^2H_1^2 = B_1^2(2I + B + B^T).$$

On the other hand

$$(I + B)^2 = 1 + 2B + B^2 = B(2I + B + B^T),$$

hence $(B_1^2 - B)(2I + B + B^T) = 0$. Consequently $B_1^2 - B$, since $2I + B + B^T$ is positive definite ($-1 \notin \sigma(B)$). Further, since $(B_1 + B_1^T)H_1 = 2I + B + B^T$, we know that $B_1 + B_1^T$ is positive definite. So, indeed B_1 is the principal square root of B .

(b) Clearly, $H_2^2 = LL^T = H$, H_2 is positive definite hence H_2 is the principal square root of H .

(c) We note that both B_1 and H_2 are functions of A , so they commute. Consequently, it follows that $N^2 = (B_1H_2)^2 = B_1^2H_1^2 = BH = A$. So, it is left to show that $N + N^T$ is positive definite. Since $N + N^T = (B_1 + B_1^T)H_2$, it follows that $N + N^T$ is positive definite. □

Now we summarise the computational procedure to compute the principal square root of a normal non-singular A .

ALGORITHM. Let $A \in \mathbb{R}^{n \times n}$ be as mentioned in Theorem 4.3. The following steps compute the principal square root N of A :

1. Compute the polar factors B and H of A .
2. Compute the orthogonal polar factor B_1 of $I + B$.
3. Compute the Cholesky decomposition, $H = LL^T$.
4. Let H_2 be the symmetric polar factor of L^T .
5. Set $N = B_1H_2$.

In our implementation of the above algorithm, we set $H = (H + H^T)/2$ to ensure that H is symmetric. We remark that, if the input matrix A is positive definite, we need to compute step 3 and step 4 only, and set $N = H_2$. Also, if the input matrix A is a rotation then we carry only step 2 and set $N = B_1$.

REMARK 4.1. The assumption that A is non-singular was made by many authors [2, 10, 20] to compute a square root of A . In many instances, however, a square root of a singular matrix does exist. This is true for normal singular matrices, producing a square root N such that $\text{Re } N$ is positive semidefinite. This can be achieved as follows: we use the Cholesky decomposition for symmetric positive semidefinite matrices [3], that is $P^T H P = \bar{L} \bar{L}^T$, where P is a permutation matrix and \bar{L} is a lower triangular matrix with non-negative elements on the diagonal. Then we proceed to the fourth step to compute \bar{H}_2 , the symmetric polar factor of \bar{L}^T . We set $H_2 = P \bar{H}_2 P^T$, hence $H_2^2 = P \bar{H}_2^2 P^T = P \bar{L} \bar{L}^T P^T = H$. We illustrate this by the following example. Let

$$A = \begin{bmatrix} 1.5 & 0.5 & -0.5 & -0.5 \\ -0.5 & 0.5 & -0.5 & -0.5 \\ 0.5 & -0.5 & 0.5 & 0.5 \\ -0.5 & 0.5 & -0.5 & 1.5 \end{bmatrix}$$

$\sigma(A) = \{0, 1 \pm i, 2\}$. A is a real normal matrix with no real negative eigenvalues,

however A is singular. Applying the modified algorithm we have,

$$N = \begin{bmatrix} 1.25645 & 0.22754 & -0.22754 & -0.15776 \\ -0.22754 & 0.54934 & -0.54934 & -0.22754 \\ 0.22754 & -0.54934 & 0.54934 & 0.22754 \\ -0.15776 & 0.22754 & -0.22754 & 1.25645 \end{bmatrix}$$

Also $\sigma(N) = \{0, \sqrt[4]{2}e^{\pm i\pi/8}, \sqrt{2}\}$. Thus the desired principal square root of A is obtained. The residual $E = A - N^2$ after 10 sweeps is not greater than 10^{-6} , for any element e_{ij} .

REMARK 4.2. If $A \in C^{n \times n}$ is normal with no real negative eigenvalues then we can compute the principal square root of A by recasting the complex problem to a real problem. Also if A has real negative eigenvalues then we can rotate A by small angle to give $A_1 = Ae^{i\Delta\theta}$, where $\Delta\theta$ is small positive real angle so that $A^{1/2} = A_1^{1/2}e^{i\Delta\theta/2}$, $A_1^{1/2}$ is computed as outlined above.

REFERENCES

- [1] M.D. Asic and D.D. Adamovic, 'Limit points of sequences on metric spaces', *Amer. Math. Monthly* **77** (1970), 613–616.
- [2] E.D. De Nman, 'Roots of real matrices', *Linear Algebra Appl.* **36** (1981), 133–139.
- [3] J.J. Dongarra, J.R. Bunch, C.B. Moler and G. Stewart, *LINPACK user's guide* (SIAM, Philadelphia, PA, 1979).
- [4] P.J. Eberlein, 'On one-sided Jacobi methods for parallel computation', *Siam J. Algebraic Discrete Methods* **8** (1987), 790–796.
- [5] I.C. Gohberg and M.G. Krein, *Introduction to the theory of linear nonself adjoint operators*, Transl. Math. Monographs **18** (Amer. Math. Soc., Providence, R.I., 1969).
- [6] G.H. Golub and C.F. Van Loan, *Matrix computations*, Johns Hopkins (University Press, Baltimore, MD, 1983).
- [7] R. Grone, C.R. Jhonson, E.M. Sa and H. Wolkowicz, 'Normal matrices', *Linear Algebra Appl.* **87** (1987), 213–225.
- [8] P.R. Halmos, 'Positive approximants of operators', *Indiana Univ. Math. J.* **21** (1972), 951–960.
- [9] N.J. Higham, 'Computing the polar decomposition with applications', *Siam J. Sci. Statist. Comput.* **7** (1986), 1160–1174.
- [10] N.J. Higham, 'Newton's method for the matrix square root', *Math. of Computation* **46** (1986), 537–549.
- [11] N.J. Higham, 'Computing real square roots of a real matrix', *Linear Algebra Appl.* **88/89** (1987), 405–430.
- [12] Ky Fan, 'Maximum properties and inequalities for the eigenvalues of completely continuous operators', *Proc. Nat. Acad. Sci. U.S.A.* **37** (1951), 760–766.

- [13] Ky Fan and A.J. Hoffman, 'Some metric inequalities in the space of matrices', *Proc. Amer. Math. Soc.* **6** (1955), 111–116.
- [14] P. Lancaster, *Theory of matrices* (Academic Press, New York, 1969).
- [15] B. Philippe, 'An algorithm to improve nearly orthonormal sets of vectors on a vector processor', *SIAM J. Algebraic Discrete Methods* **8** (1987), 396–403.
- [16] C.R. Putnam, 'On square roots of normal operators', *Proc. Amer. Math. Soc.* **8** (1957), 768–769.
- [17] R.F. Rinehart, 'The equivalence of definitions of a matrix function', *Amer. Math. Monthly* **62** (1955), 395–414.
- [18] P.H. Schonemann, 'A generalized solution of the orthogonal Procrustes problem', *Psychometrika* **31** (1966), 1–10.
- [19] G.W. Stewart, *Introduction to matrix computation* (Academic Press, New York, 1973).
- [20] Y.T. Tsay, L.S. Sheigh and J.S.H. Tsai, 'A fact method for computing the principal n th roots of complex matrices', *Linear Algebra Appl.* **76** (1986), 205–221.
- [21] J.H. Wilkinson, *The algebraic eigenvalue problem* (Oxford University Press, New York, 1977).

Department of Mathematics
Faculty of Science
University of Qatar
Doha
Qatar