JOURNAL OF RESEARCH of the National Bureau of Standards — B. Mathematics and Mathematical Physics
Vol. 71B, Nos. 2 and 3, April–September 1967

Polar Factorization of a Matrix*

John Z. Hearon**

(April 24, 1967)

It is known that if A is a bounded linear operator with closed range on a Hilbert space then A can be factored as $A=UH$, with U a partial isometry and H nonnegative and self adjoint. For the finite-dimensional case a strictly matrix-theoretic derivation is given based on the concept of a generalized inverse. Certain properties of the factors are given as well as conditions under which H or both U and H are uniquely determined by A. A pivotal item in the derivation is the representation of a square partial isometry as the product of a unitary matrix and an orthogonal projection. This representation is new, of some interest in itself and greatly simplifies the derivations.

Key Words: Generalized inverse, matrix, partial isometry.

1. Introduction

It is well known but not well discussed in the matrix literature[1] that a square matrix, A, can be factored as $A=UH$ where U is a partial isometry and H is positive semidefinite. The primary purpose of this paper is to give a fairly direct demonstration of this factorization, and the main result is thus not new (see footnote 1). The demonstration, as well as the deduction of certain properties of the factors, is based on a characterization of partial isometries which is new and of some interest *per se*. While not strictly necessary it is possible and illuminating to cast part of the development in terms of generalized inverses of singular matrices.

2. Notation and Preliminaries

In what follows all matrices are considered to have complex entries. We denote by $\rho(A)$, $R(A)$, $N(A)$ and A^* rank, range, null space and conjugate transpose, respectively, of any given matrix. When A is nonsingular, A^{-1} denotes the inverse. For generalized inverses a special terminology is used. This terminology, previously introduced and related to others [5, 6][2] is as follows: For a given matrix A denote by $C_1(A)$ the set of all matrices B such that $ABA=A$. Then $C_2(A)$ is defined as the set of all matrices B such that $B\epsilon C_1(A)$ and $A\epsilon C_1(B)$; $C_3(A)$ is the set of all matrices B such that $B\epsilon C_2(A)$ and AB is hermitian; finally $C_4(A)$ is the set of all matrices B such that $B\epsilon C_3(A)$ and BA is hermitian. We note that the set $C_4(A)$ contains a single uniquely determined matrix which is the Moore-Penrose generalized inverse [7]. We call a matrix $B\epsilon C_i(A)$ a C_i-inverse of A. The relation between a C_3-inverse, as here defined, and the "weak generalized inverse" of Goldman and Zelen [3] has been noted elsewhere [5]. Repeated use will be made of the following fact: If $B\epsilon C_1(A)$ then $\rho(B)\geqslant\rho(A)=\rho(AB)=\rho(BA)$, with strict equality if and only if $B\epsilon C_2(A)$ [5, 9].

We call a matrix A a partial isometry if there exists a subspace, S, such that $x^*A^*Ax=x^*x$, when $x\epsilon S$, and $Ax=0$, when $x\epsilon S^\perp$, where $S^\perp$ is the orthogonal complement of S. This definition is equivalent to the requirement that A^*A be an orthogonal projection [2], [4, p. 150].

3. The Polar Factorization

We begin with the following two lemmas

LEMMA 1. *The square matrix,* A, *is a partial isometry if and only if* A=QE *where* Q *is an isometry and* E *is an orthogonal projection.*

PROOF. If $A=QE$, with $Q^*Q=I$ and $E=E^2=E^*$, we have $A^*A=E$ and A is a partial isometry. Let $A=QH$ be the usual polar factorization of A, where Q is unitary and H is positive semidefinite. If A is a partial isometry then $A^*A=H^2$ is hermitian and idempotent. If so then H, the positive semidefinite square root of H^2, is also hermitian idempotent.

REMARK. It is an obvious consequence of Lemma 1 that A is a partial isometry if and only if $A=FQ$ where F is an orthogonal projection and Q is unitary. For, from $A=QE$ we have $A=QEQ^*Q$, and we identify F with the orthogonal projection QEQ^*. Conversely $A=FQ=QQ^*FQ=QE$.

LEMMA 2. Let A be normal and $B\epsilon C_1(A)$. Then if $E=AB$ is normal, E is uniquely determined by A, and $EA=AE=A$.

*An invited paper.

**Present address: Mathematical Research Branch, NIAMD, National Institutes of health, Bethesda, Md. 20014.

[1] The factorization, with condition for both factors to be unique, cf. Theorem 2 to follow, is given as a problem in [4, p. 171]. The factorization is well known as a result for bounded operators with closed range on a Hilbert space [1, 8]. Desoer and Whalen [1] give the factorization where U has the property that U^* is the pseudo-inverse (which on a finite-dimensional inner-product space is the Moore-Penrose inverse) of U. This is equivalent to U being a partial isometry.

[2] Figures in brackets indicate the literature references at the end of this paper.

PROOF. From $ABA=A=EA$ it follows that $Ax=\lambda x$, implies $Ex=x$, provided $\lambda\neq 0$. Let $\rho(A)=r$. Then there are linearly independent x_i such that $Ex_i=x_i$, $1\leq i\leq r$, and since $\rho(E)=\rho(A)$ we have $R(E)=R(A)$. Since E is a normal projection, it is an orthogonal projection and thus uniquely determined by its range and hence by A. Further, since E and A are normal, $EA=A$ shows that $N(E)=N(A)$. Hence E and A have a complete set of eigenvectors in common and must commute.

THEOREM 1. *Let* A *be any* n-*square matrix. Then there exists a partial isometry* U *and a positive semidefinite matrix* H, *such that*

(i) $A=UH$,
(ii) $U^*A=H$,
(iii) $N(U)=N(H)=N(A)$,
(iv) U *maps all of* n-*space onto* R(A)
(v) $UH=HU$ *if and only if* A *is normal, and in this case* U *is normal.*

PROOF. Let $A=QH$ be the usual polar factorization of A, where Q is unitary and H is positive semidefinite. Let P be any C_1-inverse of H such that $E=HP$ is an orthogonal projection. Then $HPH=EH=H$ and we have $A=QH=QEH=UH$, where $U=QE$ is, by Lemma 1, a partial isometry. Thus (i) is proved. Now $U^*U=E$ and hence, from (i), $U^*A=EH=H$ which is (ii). It is clear that $N(U)=N(E)$ and that $N(A)=N(H)$: and since $\rho(E)=\rho(H)$, $EH=H$ shows that $N(E)=N(H)$. Thus $N(U)=N(E)=N(H)=N(A)$, which gives (iii). Given A, the projection $E=HP$ is, by Lemma 2, uniquely determined. If P is chosen to be nonsingular, as is plainly possible (see after (2) below), then $A=UH=UEP^{-1}=UP^{-1}$, and (iv) is evident. Suppose A to be normal. Then $A=QH=HQ$ and from this and $EH=HE=H$, which we have from Lemma 2 (but which in this case is obvious from $EH=H$ since E and H are hermitian), it follows that $EA=AE=A$. But then $AE=A=HQE=HU=UH$. Conversely, suppose $UH=HU$. We have at once that $AE=EA=A$, which shows that $N(A^*)=N(E)=N(A)$. Given this, and $N(H)=N(E)$, we have from $A^*=HQ^*$ that $Qy\epsilon N(E)$ whenever $y\epsilon N(E)$. We can now assert that $HQy=QHy=0$, when $y\epsilon N(E)$. Finally, $HU=HQE=UH=QH$ implies that $HQx=QHx$, when $x\epsilon R(E)$. We have proved that $HQ=QH$ and hence that A is normal. Given this, from $A=QH=QEH=HQ=EHQ=EQH$, we have $QEx=EQx$ when $x\epsilon R(H)$, and we have seen that $Qy\epsilon N(E)=N(H)$ when $y\epsilon N(E)$. Thus $QE=EQ$ and U is normal.

THEOREM 2. *Let* $A=UH$, *where* U *is a partial isometry and* H *is positive semidefinite. Consider the conditions:* (i) $U^*A=H$, (ii) $\rho(U)=\rho(H)$, (iii) $N(U)=N(H)$. *Then, if* (i) *holds,* H *is uniquely determined;* (iii) *holds if and only if* (i) *and* (ii) *hold, and in that case both* U *and* H *are uniquely determined.*

PROOF. By Lemma 1, we may replace U by QE with Q unitary and E an orthogonal projection. Then, if (i) holds, $U^*A=EH=H$. This being so, we have $A=UH=QEH=QH$, and $H^2=A^*A$. Thus H is the unique positive semidefinite square root of A^*A. We next show that (iii) is equivalent to (i) and (ii) together. Let (i) and (ii) hold. Then, with $U=QE$, (i) gives $U^*A=EH=H$, which with (ii) implies (iii). Let (iii) hold. We obviously then have (ii). Further, with $U=QE$, (iii) states $N(E)=N(H)$. Let $x_1, x_2, \ldots, x_r$ be any orthonormal basis of $N(E)=N(H)$. Then from $E=I-\Sigma x_i x_i^*$, we have $EH=HE=H$. This being the case, $U^*A=EH=H$ which is (i). Now let $A=U_1H_1=U_2H_2$ be any two factorizations of A and assume (iii). Since (iii) implies (i), $H_1=H_2=H$ and we have $U_1H=U_2H$ which implies $U_1x=U_2x$, when $x\epsilon R(H)$. But (iii) now also requires $N(U_1)=N(U_2)=N(H)$ and hence $U_1y=U_2y$ for $y\epsilon N(H)$. Thus $U_1=U_2$.

If H is hermitian, then $H=T\operatorname{diag}(\Lambda, 0)T^*$, where T is unitary, Λ is real, diagonal and nonsingular. In the following discussion let this unitary similarity via T be denoted by $H\sim\operatorname{diag}(\Lambda, 0)$. Then for arbitrary K, L, and D of appropriate sizes and shapes any P such that

$$P\sim\begin{bmatrix}\Lambda^{-1} & K\\ L & D\end{bmatrix} \tag{1}$$

is a C_1-inverse of H. For, from

$$E=HP\sim\begin{bmatrix}I & \Lambda K\\ 0 & 0\end{bmatrix} \tag{2}$$

we have that HP is idempotent and has the rank of H and this is known [5] to be necessary and sufficient for $P\epsilon C_1(H)$. Now E, in (2), is hermitian if and only if $K=0$. Thus given $K=0$, any P as in (1) will serve in the proof of (i), (ii), (iii) and (v) of Theorem 1, and any P as in (1) with D nonsingular will serve in the proof of (iv) of Theorem 1. Now we could, in the proof of Theorem 1 except for (iv), forthwith have taken $P\epsilon C_4(H)$ or $P\epsilon C_3(H)$, for in both cases $E=HP$ is hermitian. For the proof of (iv), we could have then noted that for $P\epsilon C_4(H)$, $P+E_0$ is nonsingular when E_0 is the principal idempotent matrix of H (and of P) associated with the zero root,[3] and $H(P+E_0)=HP=E$. Of course the Theorem 1 could be proved, without reference to generalized inverses, by simply producing P as in (1) with $K=0$, noting that E as in (2) is then hermitian idempotent, and that, subject to $K=0$, E is invariant under choices of P. The pivotal idea of the proof is the observation that given $A=QH$, we have (i) of Theorem 1 at once, in view of Lemma 1, if we can produce an orthogonal projection, E, such that $EH=H$. This possibility is suggested by considering generalized inverses and that it is indeed possible is perceived at once by considering the Moore-Penrose generalized inverse, but as we have seen, other "inverses" will serve as well.

In the proof of (iv) of Theorem 1 and in the above discussion we have encountered an observation which may be set out as a corollary.

COROLLARY. *If* A *is any square matrix, there exist matrices* P *such that* AP *is a partial isometry. Further there exist such matrices* P *which are normal, in particular positive definite.*

[3] E_0 is the orthogonal projection upon $N(H)=N(P)$.

PROOF. As we have seen any P as in (1) with $K=0$ has the required property. Any P as in (1) with $K=0$, $L=0$, D normal and nonsingular is normal and nonsingular and has the required property. In particular, if $K=0$, $L=0$, and D is positive definite, we have a positive definite P from (1).

From Theorem 1, the corollary and the usual polar factorization $A=QH$, we have the following statement: If A is any square matrix, there exists an isometry Q and a partial isometry U such that $Q^*A=U^*A=H$, where H is positive semidefinite. If A is nonsingular there exists a positive definite matrix, C, such that $AC=Q$ is an isometry, but there always exists a positive definite P such that $AP=U$ is a partial isometry.

4. References

[1] C. A. Desoer and B. H. Whalen, A note on pseudoinverses, J. Soc. Indust. Appl. Math. **11,** 442–447 (1963).

[2] I. Erdelyi, On partial isometries in finite-dimensional euclidean spaces, J. SIAM Appl. Math. **14,** 453–467 (1966).

[3] A. J. Goldman and M. Zelen, Weak generalized inverses and minimum variance linear unbiased estimation, J. Res. NBS, **68B** (Math. and Math. Phys.) No. 4, 151–172 (1964).

[4] P. Halmos, Finite-dimensional vector spaces (D. Van Nostrand, New York, 1958).

[5] J. Z. Hearon, Construction of EPr generalized inverses by inversion of nonsingular matrices, J. Res. NBS **71B** (Math. and Math. Phys.) Nos. 2 & 3, (1967).

[6] J. Z. Hearon, A generalized matrix version of Rennie's inequality, J. Res. NBS **71B** (Math. and Math. Phys.) No. 2, (1967).

[7] R. Penrose, A generalized inverse for matrices, Proc. Cambridge Philos. Soc. **51,** 406–418 (1955).

[8] F. Riesz and B. Nagy, Functional Analysis, Frederick Ungar, New York, 1955.

[9] C. A. Rohde, Some results on generalized inverses, SIAM Rev. **8,** 201–205 (1966).

(Paper 71B2 & 3–200)