J. Soc. Indust. Appl. Math.
Vol. 11, No. 2, June, 1963
Printed in U.S.A.

A NOTE ON PSEUDOINVERSES[1]

C. A. DESOER[2] AND B. H. WHALEN[3]

1. Introduction. When an operator, A, is invertible it is possible to solve the equation $Ax = y$ by operating on both sides with the inverse of A. Moore [1, 2] introduced the concept (for matrices) of the pseudoinverse in order to extend this technique to situations in which A has no inverse but $Ax = y$ has a solution. Moore's definition was based on the row and column spaces of A and was sufficient to make the pseudoinverse, $A^\dagger$, unique and to give it useful symmetry properties, notably $(A^\dagger)^\dagger = A$, $A^{\dagger *} = A^{*\dagger}$.

Penrose [3, 4] introduced independently, and much later, the identical operator, but in a different manner. Penrose's approach was axiomatic and algebraic. Kalman [7, 8] using Penrose's approach has applied the pseudoinverse to the theories of filtering and control. Greville [5, 6] has brought Moore's definition to the fore and has contributed (as did Penrose) to the computational aspects of pseudoinverses.

Because of the projection theorem for inner product spaces, (see [12, p. 51]) the pseudoinverse is very useful in least square curve-fitting and estimation problems. Penrose, Greville, Kalman and others have demonstrated and applied this fact.

The purpose of this paper is to introduce the pseudoinverse from a range-null-space point of view. The authors feel that this approach is beneficial in that *i.* the definition has a strong motivation, *ii.* the concepts are illuminated geometrically, *iii.* the proofs are quite simple, *iv.* the basis is eliminated, and *v.* the extension to bounded linear mappings with closed range between Hilbert spaces is immediate. In fact, this paper will deal exclusively with such mappings since all of the proofs follow easily the elementary properties of Hilbert spaces. Since every finite dimensional inner product space is a Hilbert space, the results will be valid for such spaces [9, p. 243].

2. Assumptions and basic facts. Let A be a bounded linear transformation of a Hilbert space $\mathfrak{X}$ into a Hilbert space $\mathfrak{Y}$. The Hilbert space adjoint of A, which will be denoted by A^* is defined by

$$\text{(a)} \qquad (Ax, y) = (x, A^*y) \text{ for all } x \text{ in } \mathfrak{X} \text{ and all } y \text{ in } \mathfrak{Y}$$

[1] Received by the editors April 20, 1962, and in final revised form November 30, 1962. This work has been supported by NSF Grant G-15965.

[2] Department of Electrical Engineering, University of California, Berkeley 4, California.

[3] RIAS, 7212 Bellona Avenue, Baltimore 12, Maryland. Formerly at University of California, Berkeley 4, California.

where $(\cdot,\cdot)$ denotes the inner product in both $\mathfrak{X}$ and $\mathfrak{Y}$. The fact that (a) defines a unique bounded linear transformation of $\mathfrak{Y}$ into $\mathfrak{X}$ is demonstrated in [11, p. 249].

Denote by $\mathfrak{N}(A)$ the set of x in $\mathfrak{X}$ such that

$$Ax = 0,$$

and by $\mathfrak{R}(A)$ the set of y in $\mathfrak{Y}$ such that

$$Ax = y \qquad \text{for some } x \text{ in } \mathfrak{X}.$$

Denote by $\mathfrak{N}(A)^{\perp}$ the set of x in $\mathfrak{X}$ such that

$$(x, x_0) = 0 \qquad \text{for every } x_0 \text{ in } \mathfrak{N}(A).$$

The following facts are easily deduced and may be found in [11, p. 250].

(1) $\mathfrak{Y} = \overline{\mathfrak{R}(A)} \oplus \overline{\mathfrak{R}(A)}^{\perp}$; $\quad \overline{\mathfrak{R}(A)}^{\perp} = \mathfrak{N}(A^*)$.

(2) $\mathfrak{X} = \mathfrak{N}(A) \oplus \mathfrak{N}(A)^{\perp}$; $\mathfrak{N}(A)^{\perp} = \overline{\mathfrak{R}(A^*)}$.

(3) $A^{**} = A$.

Combining (3) with Theorem VI, 6.4 of [9, p. 488], we get

(4) $\mathfrak{R}(A) = \overline{\mathfrak{R}(A)}$ if and only if $\mathfrak{R}(A^*) = \overline{\mathfrak{R}(A^*)}$.

From (4), (1) and (2) it is apparent that

(5) If $\mathfrak{R}(A) = \overline{\mathfrak{R}(A)}$, then A is a one-to-one map of $\mathfrak{R}(A^*)$ onto $\mathfrak{R}(A)$ and A^* is a one-to-one map of $\mathfrak{R}(A)$ onto $\mathfrak{R}(A^*)$.

Since a closed subspace of a Hilbert space is a Hilbert space, one can consider the restriction of A to $\mathfrak{R}(A^*)$ as a bounded linear transformation onto a Hilbert space $\mathfrak{R}(A)$ if $\mathfrak{R}(A) = \overline{\mathfrak{R}(A)}$. Denote this transformation by B. It follows that

(6) If $\mathfrak{R}(A) = \overline{\mathfrak{R}(A)}$, the bounded linear transformation, B, defined above has a bounded linear inverse, B^{-1}. (See [9, p. 513].)

The following facts will be used in proofs of the theorems.

(7) A bounded linear operator E is an orthogonal projection of $\mathfrak{X}$ onto $\mathfrak{R}(E)$ if and only if $E^2 = E$ and $E = E^*$. (See [11, pp. 241–333].)

(8) Let x_1, x_2 belong to $\mathfrak{X}$. Then $x_1 = x_2$ if and only if $(x_1, x) = (x_2, x)$ for every x in $\mathfrak{X}$. (This result follows directly from the axioms of the inner product [11, p. 106].)

(9) Every nonnegative self adjoint bounded linear operator on a Hilbert space $\mathfrak{X}$, has a unique nonnegative self adjoint square root which is also a bounded linear operator on $\mathfrak{X}$. (See [10, p. 265].)

The following fact is presented only as an illustration of a class of operators to which the results of this paper apply.

(10) Let A and G be bounded linear transformations of $\mathfrak{X}$ into $\mathfrak{Y}$. Let the range of A be closed and the range of G finite dimensional. Then the range of $A + G$ is closed.

The proof is obtained by combining (4), Theorem 1A.II of [12, p. 53] and the fact that an operator with finite dimensional range can be written as a finite sum of dyads [12, pp. 26–27].

From (10) it is clear, since E_n (the Euclidean n-space) is complete and therefore a Hilbert space, that every matrix is a bounded linear transformation from a Hilbert space $\mathfrak{X}$ into a Hilbert space $\mathfrak{Y}$ and has a closed range. The pseudoinverse defined below is therefore an extension of the pseudoinverse of Moore and Penrose.

3. Definition of pseudoinverse. DEFINITION. *Let A be a bounded linear operator of a Hilbert space $\mathfrak{X}$ into a Hilbert space $\mathfrak{Y}$ such that $\mathcal{R}(A)$ is closed. $A^\dagger$ is said to be the pseudoinverse of A if*

i. $A^\dagger Ax = x$ *for all* $x \in \mathfrak{N}(A)^\perp = \mathcal{R}(A^*)$,

ii. $A^\dagger y = 0$ *for all* $y \in \mathcal{R}(A)^\perp = \mathfrak{N}(A^*)$,

iii. if $y_1 \in \mathcal{R}(A)$ *and* $y_2 \in \mathfrak{N}(A^*)$ *then* $A^\dagger(y_1 + y_2) = A^\dagger y_1 + A^\dagger y_2$.

The operator $A^\dagger A$ restricted to $\mathfrak{N}(A)^\perp$ is the identity map. Note that *i.* defines $A^\dagger$ on $\mathcal{R}(A)$ (this follows by (5)), and *ii.* defines $A^\dagger$ on $\mathcal{R}(A)^\perp$, hence by (1) and *iii.* $A^\dagger$ *is uniquely defined.* We assert that $A^\dagger$ *is a linear operator*: to see this decompose any $y \in \mathfrak{X}$ as $y = y_1 + y_2$ with $y_1 \in \mathcal{R}(A)$ and $y_2 \in \mathcal{R}(A)^\perp$; by (5) there is a unique vector $u \in \mathfrak{N}(A)^\perp$ such that $y_1 = Au$, then $y = Au + y_2$ where $u \in \mathfrak{N}(A)^\perp$ and $y_2 \in \mathcal{R}(A)$. With such decomposition the homogeneity and additivity of $A^\dagger$ is easily checked.

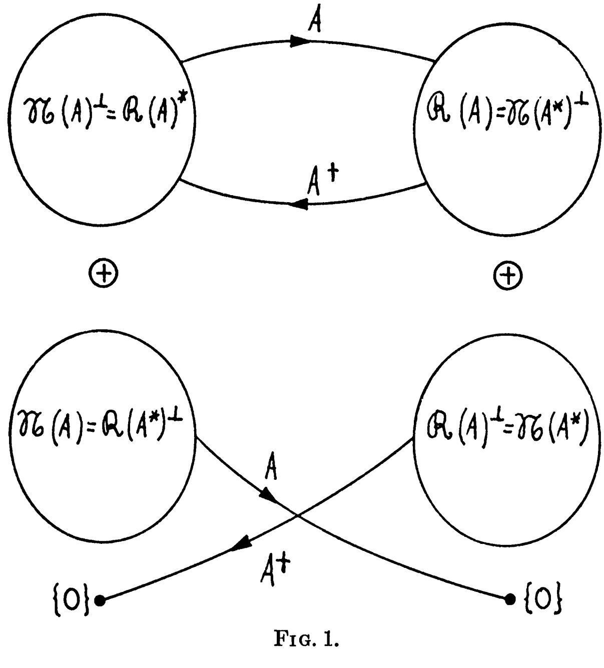

FIG. 1.

The following corollary follows immediately from the definition.

COROLLARY 1. $\mathcal{R}(A^\dagger) = \mathcal{R}(A^*) = \mathfrak{N}(A)^\perp; \mathfrak{N}(A^\dagger) = \mathfrak{N}(A^*) = \mathcal{R}(A)^\perp$.

The relationship implied by the definition and Corollary 1 may be visualized by Fig. 1. The first two sets on the left of the figure represent the direct sum (2) and the first two on the right represent the direct sum (1). By definition, A maps $\mathfrak{N}(A)$ into 0 and $A^\dagger$ maps $\mathcal{R}(A)^\perp$ into 0. Also, $A^\dagger$ restricted to $\mathcal{R}(A)$ is the inverse map of the one to one linear mapping, B, which we defined as A restricted to $\mathcal{R}(A^*)$. From (6), we know that B has a bounded linear inverse, therefore $A^\dagger$ is bounded and has a closed range by Corollary 1. Therefore $A^{\dagger\dagger}$ is defined.

COROLLARY 2. *$A^\dagger A$ is the orthogonal projection of $\mathfrak{X}$ onto $\mathfrak{N}(A) = \mathcal{R}(A^*)^\perp$.*

Proof. For any $x \in \mathfrak{X}$, consider the orthogonal decomposition $x = x_1 + x_2$, with $x_1 \in \mathfrak{N}(A)^\perp$ and $x_2 \in \mathfrak{N}(A)$. Now $A^\dagger Ax = A^\dagger Ax_1 = x_1$ from which the conclusion follows.

COROLLARY 3. $(A^\dagger)^\dagger = A$.

This fact is intuitively obvious from the symmetry of Fig. 1. It is easily checked in detail by referring to the definition and Corollary 1.

COROLLARY 4. $A^\dagger AA^\dagger = A^\dagger; (A^\dagger A)^* = A^\dagger A$.

Proof. The second equality is immediate by (4) and the fact that $A^\dagger A$ is an orthogonal projection. For any $y \in \mathfrak{Y}$, consider the orthogonal decomposition $y = y_1 + y_2$ with $y_1 \in \mathcal{R}(A)$, $y_2 \in \mathcal{R}(A)^\perp$. By definition of $A^\dagger$, $A^\dagger y = A^\dagger y_1 \in \mathfrak{N}(A)^\perp$. Multiply this equality by $A^\dagger A$, then by Corollary 2, $A^\dagger AA^\dagger y = A^\dagger AA^\dagger y_1 = A^\dagger y_1 = A^\dagger y$, and the first equality is established.

COROLLARY 5. $AA^\dagger A = A; (AA^\dagger)^* = AA^\dagger$.

Proof. Let $A^\dagger = B$, then by Corollary 3, $B^\dagger = A$. With this substitution the two equalities to be proven are identical to those of Corollary 4. The same idea applied to Corollary 2 leads, with the help of Corollary 1, to

COROLLARY 6. *$AA^\dagger$ is the orthogonal projection of $\mathfrak{Y}$ onto $\mathcal{R}(A)$.*

NOTE. For the finite dimensional case, Penrose [3] has shown that if the four equations of Corollaries 4 and 5 are considered as equations for the unknown matrix $A^\dagger$, then these equations have a unique solution which he defines to be the pseudoinverse.

4. Properties of the pseudoinverse.

THEOREM 1. $A^\dagger = (A^*A)^\dagger A^*$.

Thus the computation of $A^\dagger$ requires only that of the pseudoinverse of a self adjoint operator. It is easy to check that if B is self adjoint and if it has a spectral representation of the form

$$A = \sum \lambda_i E_i$$

(in which the summation is carried over all non-zero eigenvalues of B), then

$$B^\dagger = \sum \lambda_i^{-1} E_i .$$

Proof. If $y \in \mathfrak{N}(A^*) = \mathfrak{R}(A)^\perp$, the theorem is true since for such y, $(A^*A)^\dagger A^*y = 0$ and, by definition, $A^\dagger y = 0$. Thus the conclusion of the theorem remains to be checked for $y \in \mathfrak{R}(A)$. To each $y \in \mathfrak{R}(A)$ there is a unique x in $\mathfrak{N}(A)^\perp = \mathfrak{R}(A^*)$ such that $y = Ax$. On the other hand, from (5), A^*A is a one-to-one map of $\mathfrak{R}(A^*)$ onto itself and by the definition of the pseudoinverse, for any $x \in \mathfrak{R}(A^*)$, $(A^*A)^\dagger A^*Ax = x$ or $(A^*A)^\dagger A^*y = x$. Therefore, for all $y \in \mathfrak{R}(A)$, $(A^*A)^\dagger A^*y = A^\dagger y$.

THEOREM 2. $(A^*)^\dagger = (A^\dagger)^*$.

Proof. By Corollary 5, $AA^\dagger = (AA^\dagger)^* = A^{\dagger *}A^*$. By Corollaries 6 and 2, $AA^\dagger$ is the orthogonal projection of $\mathfrak{Y}$ onto $\mathfrak{R}(A)$ and $A^{*\dagger}A^*$ is the orthogonal projection of $\mathfrak{Y}$ on $\mathfrak{R}(A^{**}) = \mathfrak{R}(A)$. Hence $A^{*\dagger}A^* = A^{\dagger *}A^*$, i.e., $A^{\dagger *}x = A^{*\dagger}x$ for all $x \in \mathfrak{R}(A^*) = \mathfrak{N}(A)^\perp$. But $\mathfrak{N}(A^{\dagger *}) = \mathfrak{N}(A^{*\dagger}) = \mathfrak{N}(A)$ because, by Corollary 1, $\mathfrak{N}(A^{*\dagger}) = \mathfrak{N}(A^{**}) = \mathfrak{N}(A)$ and $\mathfrak{N}(A^{\dagger *}) = \mathfrak{N}(A^{\dagger\dagger}) = \mathfrak{N}(A)$. Hence $A^{\dagger *}x = A^{*\dagger}x = 0$ for all $x \in \mathfrak{N}(A)$, and the theorem follows from (2).

THEOREM 3. (*Equivalent definition*). *Let* $x, y \in \mathfrak{X}$ *and* $x = x_1 + x_2$, $y = y_1 + y_2$ *where* $x_1, y_1 \in \mathfrak{N}(A)^\perp$ *and* $x_2, y_2 \in \mathfrak{N}(A)$. *Then* $B = A^\dagger$ *if and only if i.* $(BAx, y) = (x_1, y_1)$, *for all* x, y *in* $\mathfrak{X}$ *and ii.* $\mathfrak{N}(B) = \mathfrak{N}(A^*) = \mathfrak{R}(A)^\perp$.

Proof. Necessity. By Corollary 2, for x, y in $\mathfrak{X}$, $(A^\dagger Ax, y) = (x_1, y_1 + y_2) = (x_1, y_1)$; *ii.* follows from the definition of $A^\dagger$.

Sufficiency. From *i.*, $(BAx_1, y) = (x_1, y_1) = (x_1, y)$ for all $x_1 \in \mathfrak{N}(A)$ and all $y \in \mathfrak{X}$; hence from (8), $BAx_1 = x_1$ for all $x_1 \in \mathfrak{N}(A)$. This fact together with *ii.* is the definition of $A^\dagger$.

THEOREM 4. *Let* $y \in \mathfrak{Y}$ *and* $x_1 = A^\dagger y$. *Then*

$$\|Ax_1 - y\| \leqq \|Ax - y\| \qquad \text{for all } x \text{ in } \mathfrak{X}, \tag{b}$$

and $\|x_1\| \leqq \|x_0\|$ *for all* x_0 *satisfying inequality* (b).

Note that if $x_2 \in \mathfrak{N}(A)$, $A(x_1 + x_2) = Ax_1$. Hence Theorem 4 shows that if $Ax = y$ has a solution, all solutions are of the form $A^\dagger y + x_2$ with $x_2 \in \mathfrak{N}(A)$ and any vector of this form is a solution.

Proof. Let $y = y_1 + y_2$ with $y_1 \in \mathfrak{R}(A)$ and $y_2 \in \mathfrak{R}(A)^\perp$. Then

$$\|Ax_1 - y\| = \|AA^\dagger y - y\| = \|y_1 - y\| = \|y_2\|.$$

On the other hand, for any x in $\mathfrak{X}$, let $Ax = y_3 \in \mathfrak{R}(A)$,

$$\|Ax - y\|^2 = \|y_3 - y_1 - y_2\|^2 = \|y_3 - y_1\|^2 + \|y_2\|^2$$

and (b) follows. Any vector x_0 satisfying (b) is of the form $x_1 + \hat{x}$ where $(x_1, \hat{x}) = 0$. Therefore $\|x_1\| \leqq \|x_0\|$.

Finally, we state a theorem which is a generalization of the polar decomposition of a matrix [7].

THEOREM 5. *Let* A *be a bounded linear map of one Hilbert space* $\mathfrak{X}$ *into* *another,* $\mathfrak{Y}$. *If the range of A is closed, then A can be factored as*

$$A = UH$$

where $U^* = U^\dagger$, *H is a nonnegative self adjoint operator in* $\mathfrak{X}$ *and U maps* $\mathfrak{X}$ *onto* $\mathcal{R}(A)$.

Proof. Since A^*A is nonnegative and self adjoint, it has a unique bounded nonnegative self adjoint square root (see [10, p. 265].) Call this square root H. Since $\mathcal{R}(H) = \mathcal{R}(H^*H) = \mathcal{R}(A^*A) = \mathcal{R}(A^*)$, $\mathcal{R}(H)$ is closed. Thus, H has a pseudoinverse, $H^\dagger$. Next define the operator U by $U = AH^\dagger$. Now, since H commutes with $H^\dagger$,

$$U^*U = (AH^\dagger)^*(AH^\dagger) = H^\dagger A^*AH^\dagger = H^\dagger HH^\dagger H = H^\dagger H.$$

By Corollary 2, U^*U is an orthogonal projection of $\mathfrak{X}$ onto $\mathcal{R}(H)$, i.e., $U^*Ux = x$ for all $x \in \mathcal{R}(H)$. However, $\mathcal{R}(H) = \mathcal{R}(A^*) = \mathcal{R}(U^*)$, thus $U^*Ux = x$ for all $x \in \mathcal{R}(U^*)$. Finally, for all $y \in \mathcal{R}(U)^\perp = \mathfrak{N}(U^*)$, $U^*y = 0$. Hence by definition of $U^\dagger$, $U^* = U^\dagger$.

REFERENCES

[1] E. H. Moore, *General Analysis*, Part 1, Memoirs, Amer. Philos. Soc. 1 (1935), esp. pp. 147–209.

[2] E. H. Moore, Bulletin Amer. Math. Soc., 26 (1920), pp. 394–395.

[3] R. Penrose, *On a generalized inverse for matrices*, Proc. Cambridge Philos. Soc., 51, Part 3 (1955), pp. 406–413.

[4] R. Penrose, *On best approximate solutions of linear matric equations*, Ibid., 52, Part 1 (1956), pp. 17–19.

[5] T. N. E. Greville, *The pseudoinverse of a rectangular or singular matrix and its application to the solution of systems of linear equations*, SIAM REV., 1 (1959), pp. 38–43.

[6] T. N. E. Greville, *Some applications of the pseudoinverse of a matrix*, Ibid., 2, 1 (1960), pp. 15–22.

[7] R. E. Kalman, Y. C. Ho and K. S. Narenda, *Controllability of linear dynamical systems*, Contributions to Differential Equations, vol. I, Interscience Publishers, Inc., New York, 1962.

[8] R. E. Kalman, New methods and results in linear prediction and filtering theory, RIAS Tech. Report 61–1.

[9] N. Dunford, and J. Schwartz, *Linear Operators*, Part I, Interscience, New York, 1958.

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

[11] A. E. Taylor, *Introduction to Functional Analysis*, John Wiley, New York, 1958.

[12] B. Friedman, *Principles and Techniques of Applied Mathematics*, John Wiley, New York, 1956.