players of the team behind after 5 games may make a major effort in game 6 which cannot be sustained for game 7, while the players on the team ahead after 5 games individually husband resources in game 6. Perhaps readers will have other explanations.

References

1. H. Cramer, Mathematical Methods of Statistics, Princeton University Press, 1946, pp. 416–452.
2. Z. Hollander, Editor, The Modern Encyclopedia of Basketball, Four Winds Press, New York, 1973.
3. F. Mosteller, The World Series Competition, J. Amer. Statist. Assoc., 47 (1952) 355–380.
4. D. S. Neft, R. T. Johnson, R. M. Cohen, and J. A. Deutsch, The Sports Encyclopedia: Baseball, Grosset and Dunlap, New York, 1974.
5. R. A. Styer, The Encylopedia of Hockey, rev. ed., A. S. Barnes, New York, 1973.
6. The Official Associated Press Sports Almanac, Dell, New york, 1974.
7. M. Zelinka, How Many Games to Complete a World Series, in Statistics by Example: Weighing Chances, ed. by F. Mosteller, et al., Addison-Wesley, Reading, 1973.

CLOSEST UNITARY, ORTHOGONAL AND HERMITIAN OPERATORS TO A GIVEN OPERATOR

JOSEPH B. KELLER, Courant Institute of Mathematical Sciences, New York University

1. Introduction. It is often of interest to find an operator U_0, in some specified class of operators $\mathcal{U}$, which is closest to a given operator A. We shall describe various situations in which this kind of problem arises, and then show how to solve it for several choices of $\mathcal{U}$.

1. A simple example of this kind arises when one tries to determine a rotation matrix by measuring or by computing its entries. Because of inevitable experimental or computational errors, the resulting matrix A will generally not be orthogonal. Therefore we may wish to adjust its entries to make it orthogonal. A reasonable way to do so is to change it into that orthogonal matrix U_0 which is closest to A in some appropriate norm. Finding U_0 is a problem of the type described above with $\mathcal{U}$ being the class of orthogonal matrices. This problem was solved by K. Fan and A. J. Hoffman [1], who found the unitary and hermitian matrices closest to a given matrix. Their result is independent of the norm, provided the norm is unitarily invariant, i.e., provided that $\|A\| = \|UA\| = \|AU\|$ when U is unitary.

2. Another example arises in factor analysis of psychological test data and latent structure analysis of sociological data. In both of these cases an m by n data matrix A is supposed to be a product of an m by r matrix B times an r by n matrix C, with $r < m$ and $r < n$. The product BC has rank $\leqq r$, but the rank of A will generally exceed r, so there is no such factorization in general. Therefore it

is customary to seek the best approximate factorization. This corresponds to finding the matrix U_0 closest to A in the class $\mathcal{U}$ of m by n matrices of rank r, since every matrix in $\mathcal{U}$ can be factored in the desired manner, and every factorizable matrix is in $\mathcal{U}$. This problem was solved by C. Eckart and G. Young [2] and by J. B. Keller [3] using the euclidean norm. L. Mirsky [4] showed that the result is independent of the norm, provided the norm is unitarily invariant.

3. A modification of the above problem also arises in factor analysis, in which the matrix C is given. Then $\mathcal{U}$ is the class of matrices BC where B is an arbitrary m by r matrix and C is given. Alternatively, B may be given and C may be arbitrary. These problems were solved by B. Green [5], J. B. Keller [3] and P. Schönemann [6] using the euclidean norm.

4. A special case of problem 3 is that with $m = n = r$ and $A = I$, the identity matrix. Then the problem is to find a matrix B which makes BC closest to the identity. Such a matrix B we call a generalized left inverse of C. Similarly if B is given and if C makes BC closest to I, we call C a generalized right inverse of B. These problems have been solved by R. Penrose [7].

5. A Fredholm integral equation of the second kind is an equation of the form

$$\phi(x) + \int_a^b K(x, y)\phi(y)dy = f(x).$$

Here $\phi(x)$ is the unknown function, f is a given inhomogeneous term, and K is a given function called the kernel. The equation can be reduced to a system of n linear algebraic equations for n unknowns, and then it can be solved easily, if K is a degenerate kernel of rank n. This is a kernel of the form

$$K_n(x, y) = \sum_{j=1}^{n} g_j(x)h_j(y).$$

Therefore one way to obtain an approximate solution of the integral equation is to approximate its kernel K by a degenerate kernel of rank n. Naturally it is desirable to obtain the best such approximation, and this is again an instance of the general problem described above. The solution is given by R. Courant and D. Hilbert [8].

G. H. Golub [9] has given the solutions of problems 1–4 together with a numerical procedure for computing them.

We shall describe two methods for analyzing such problems. The first is the direct method of showing that a particular U_0 is closer to A than any other U in $\mathcal{U}$ be comparing the norm $\|A - U_0\|$ of $A - U_0$ with $\|A - U\|$. The virtue of this method is its simplicity and generality. It applies to operators on infinite dimensional spaces as well as to matrices of finite dimension. However, it has the disadvantage that the solution U_0 must be known before it can be used. The second method is the usual indirect method which is used to find minima in calculus. It consists in representing U as a matrix, differentiating $\|A - U\|$ with respect to the matrix elements, and equating the derivatives to zero. In doing this, the condition that U is in $\mathcal{U}$ must be taken into account by the use of

Lagrange multipliers. This method leads to an equation which must be satisfied by U_0. Thus it has the virtue that U_0 need not be known in advance. Its disadvantage is that after U_0 has been found, it must still be shown that U_0 does yield the smallest value of $\|A - U\|$.

I wish to thank my colleague A. B. Novikoff for some stimulating discussions of these topics.

2. Closest hermitian and anti-hermitian operators. Let A be a linear operator and $\mathcal{U}$ a set of linear operators which map a unitary space into itself. The euclidean norm $\|U\|$ of an operator U is defined as the positive square root of the trace of U times its adjoint U^+, so that

$$\|U\|^2 = \operatorname{tr} UU^+. \tag{1}$$

This is an inner product norm, so the notion of orthogonal projection can be introduced in $\mathcal{U}$, but we shall not make use of it explicitly. In terms of the matrix elements u_{ij} of U with respect to an orthonormal basis of the space, $\|U\|^2$ is given by

$$\|U\|^2 = \sum_{i,j} |u_{ij}|^2. \tag{2}$$

We define $U_0 \in \mathcal{U}$ to be closest to A if U_0 minimizes the distance $\|A - U\|$ among all $U \in \mathcal{U}$. We shall now prove

THEOREM 1. *Among all hermitian operators, the unique operator closest to A is* $U_0 = \frac{1}{2}(A + A^+)$.

Proof. U is an hermitian operator if $U = U^+$, so obviously $U_0 = \frac{1}{2}(A + A^+)$ is hermitian and so is $V = U - U_0$. Therefore any hermitian operator U can be represented as a sum $U = U_0 + V$ where $V = V^+$. Now (1) yields

$$\begin{aligned} \|A - U\|^2 &= \|A - U_0 - V\|^2 \\ &= \|A - U_0\|^2 + \operatorname{tr}[(A - U_0)V^+ + V(A^+ - U_0^+)] + \|V\|^2. \end{aligned} \tag{3}$$

By using the facts that $U_0 = U_0^+$, $V = V^+$ and that the trace of a product is invariant under cyclic permutation of the factors, we find

$$\begin{aligned} \operatorname{tr}[(A - U_0)V^+ + V(A^+ - U_0^+)] &= \operatorname{tr}[(A - U_0)V + V(A^+ - U_0)] \\ &= \operatorname{tr}[(A + A^+ - 2U_0)V] = 0. \end{aligned} \tag{4}$$

The last equality in (4) follows from the definition of U_0. Upon using (4) in (3) we obtain

$$\|A - U\|^2 = \|A - U_0\|^2 + \|V\|^2 \geqq \|A - U_0\|^2. \tag{5}$$

This proves the theorem since $\|V\|^2 > 0$ unless $V = 0$.

An anti-hermitian operator is an operator U for which $U = -U^+$. By changing a few signs in the proof above, we can prove

THEOREM 2. *Among all anti-hermitian operators, the unique operator closest to A is* $U_0 = \frac{1}{2}(A - A^+)$.

3. Closest unitary and orthogonal operators.

THEOREM 3. *A closest unitary operator to A is any unitary operator U_0 which occurs in a polar decomposition $A=(AA^+)^{\frac{1}{2}}U_0$ of A. If A is invertible then $U_0=(AA^+)^{-\frac{1}{2}}A$ is the unique closest unitary operator to A.*

In the polar decomposition the positive square root is used. This theorem provides a characterization of the unitary factor U_0 in the polar decomposition.

Proof. Since U_0 is unitary, $U_0U_0^+=I$. Let U be any unitary operator and let $V=U-U_0$. From the unitarity of U and U_0 it follows that

$$U_0V^+ + VU_0^+ + VV^+ = 0. \tag{6}$$

Thus any unitary U can be written as $U=U_0+V$ where V satisfies (6). Now we can write

$$\begin{aligned}\|A-U\|^2 = \|A-U_0-V\|^2 &= \|A-U_0\|^2 - \operatorname{tr}[(A-U_0)V^+ + V(A^+-U_0^+) - VV^+] \\ &= \|A-U_0\|^2 - \operatorname{tr}[AV^+ + VA^+].\end{aligned} \tag{7}$$

The last equality follows from (6). Next we use the polar decomposition of A in the last term in (7), then use the invariance of the trace of a product under cyclic permutation of the factors and then use (6) to obtain

$$\begin{aligned}\operatorname{tr}[AV^+ + VA^+] &= \operatorname{tr}[(AA^+)^{\frac{1}{2}}U_0V^+ + VU_0^+(AA^+)^{\frac{1}{2}}] \\ &= \operatorname{tr}[(AA^+)^{\frac{1}{2}}(U_0V^+ + VU_0^+)] = -\operatorname{tr}[(AA^+)^{\frac{1}{2}}VV^+].\end{aligned} \tag{8}$$

Since $(AA^+)^{\frac{1}{2}}$ and VV^+ are both hermitian and nonnegative, the trace of their product is nonnegative.

Upon using (8) in (7) and noting that the last trace in (8) is nonnegative, we obtain

$$\|A-U\|^2 = \|A-U_0\|^2 + \operatorname{tr}[(AA^+)^{\frac{1}{2}}VV^+] \geqq \|A-U_0\|^2. \tag{9}$$

This proves the first part of the theorem. If A is invertible then AA^+ is invertible, so the trace in (9) can vanish if and only if $V=0$. Therefore inequality holds in (9) unless $U=U_0$. This proves the second part of the theorem.

If A is real then $A^+=A^T$ where A^T, the transpose of A, is also real. Therefore U_0 is real and orthogonal. As a consequence Theorem 3 yields

THEOREM 4. *A closest orthogonal operator to a real operator A is any orthogonal operator U_0 which occurs in a polar decomposition $A=(AA^T)^{\frac{1}{2}}U_0$ of A. If A is invertible then $U_0=(AA^T)^{-\frac{1}{2}}A$ is the unique closest orthogonal operator to A.*

If A is not necessarily real then by slightly modifying the proof of Theorem 3 we can prove

THEOREM 5. *A closest orthogonal operator to A is any orthogonal operator U_0 which occurs in a polar decomposition* $\operatorname{Re}A=(\operatorname{Re}A\operatorname{Re}A^T)^{\frac{1}{2}}U_0$ *of* $\operatorname{Re}A$. *If* $\operatorname{Re}A$ *is invertible then* $U_0=(\operatorname{Re}A\operatorname{Re}A^T)^{-\frac{1}{2}}\operatorname{Re}A$.

4. Generalized left and right inverses. In the introduction we defined a generalized left inverse of A as an operator U_0 which minimizes $\| UA - I \|$. Now we shall prove

THEOREM 6. *U_0 is a generalized left inverse of A if it is a solution of the equation $U_0AA^+ = A^+$. If A is invertible then $U_0 = A^{-1}$.*

Proof. Any operator U can be written as $U = U_0 + V$ where $V = U - U_0$. Then

$$\begin{aligned} \| UA - I \|^2 &= \|(U_0 + V)A - I \|^2 \\ &= \| U_0A - I \|^2 + \| VA \|^2 + \operatorname{tr}[(U_0A - I)A^+V^+ - VA(A^+U_0^+ - I)]. \end{aligned} \tag{10}$$

The trace in (10) vanishes in view of the equation satisfied by U_0 and (10) yields, since $\| VA \|^2 \geqq 0$,

$$\| UA - I \|^2 = \| U_0A - I \|^2 + \| VA \|^2 \geqq \| U_0A - I \|^2. \tag{11}$$

This proves the theorem. Similarly we can prove

THEOREM 7. *U_0 is a generalized right inverse of A if it is a solution of the equation $A^+AU_0 = A^+$. If A is invertible then $U_0 = A^{-1}$.*

Penrose [7] has shown for matrices, i.e., operators on a finite dimensional space, that there exists a unique U_0, called a generalized inverse of A, which satisfies both $U_0AA^+ = A^+$ and $A^+AU_0 = A^+$. From the theorems above, this U_0 is both a generalized left and genralized right inverse of A. Furthermore he has shown that U_0B is the unique best approximate solution of the equation $AX = B$. The best approximate solution is defined to be that X which minimizes $\|AX - B\|$, and if there is more than one minimizer, then it has the least value of $\|X\|$.

5. Lagrange multiplier method. We shall now illustrate the indirect method by using it to find a closest unitary matrix U to the matrix A of dimension n. We first introduce the matrix Λ of Lagrange multipliers λ_{ij} and consider the function $\varepsilon(U)$ defined by

$$\varepsilon(U) = \|A - U\|^2 + \frac{1}{2}\sum_{i,j,k} [\lambda_{ik}(u_{ij}u^*_{kj} - \delta_{ik}) + \lambda^*_{ik}(u^*_{ij}u_{kj} - \delta_{ik})]. \tag{12}$$

We now differentiate (12) with respect to u_{ij} and set $\partial\varepsilon/\partial u_{ij} = 0$. This yields

$$-a^*_{ij} + \frac{1}{2}\sum_k (\lambda_{ik}u^*_{kj} + \lambda^*_{ki}u^*_{kj}) = 0. \tag{13}$$

In differentiating (12) we used the fact that $\|U\|^2 = n$ to avoid differentiating $\|U\|^2$. Differentiation of (12) with respect to u^*_{ij} yields the complex conjugate of (13). If $\Lambda + \Lambda^+$ is nonsingular, the solution of (13) is

$$U^* = 2(\Lambda + \Lambda^+)^{-1}A^*. \tag{14}$$

To determine Λ we use (14) in $UU^+ = I$ and obtain

$$4(\Lambda^* + \Lambda^T)^{-1}AA^+(\Lambda^T + \Lambda^*)^{-1} = I. \tag{15}$$

Multiplication of (15) on the left and on the right by $(\Lambda^T + \Lambda^*)$ yields

$$4AA^+ = (\Lambda^T + \Lambda^*)^2. \tag{16}$$

The solution of (16) is

$$\Lambda^T + \Lambda^* = 2(AA^+)^{\frac{1}{2}}. \tag{17}$$

Then (14) and (17) yield the solution, which we denote by U_0,

$$U_0 = (AA^+)^{-\frac{1}{2}}A. \tag{18}$$

This result holds only if $(AA^+)^{\frac{1}{2}}$ is nonsingular, which is the case if and only if A is nonsingular.

The result (18) for the closest unitary matrix to A is ambiguous because the choice of square root has not been determined. To determine it we use (18) to evaluate $\|A - U_0\|^2$ and find

$$\|A - U_0\|^2 = \|A\|^2 + n - \mathrm{Tr}(AU_0^+ + U_0A^+) = \|A\|^2 + n - 2\,\mathrm{Tr}(AA^+)^{\frac{1}{2}}. \tag{19}$$

We see from (19) that to minimize $\|A - U_0\|$ we must choose the square root which maximizes $\mathrm{Tr}(AA^+)^{\frac{1}{2}}$. This is just the positive square root. The proof that U_0 given by (18) is actually closest to A is given in Section 3.

In case A is real and U is orthogonal, we may seek U in either the class with determinant plus one or that with determinant minus one. Then we must restrict the square root in (18) so that det $U_0 = +$ or -1 and maximize $\mathrm{Tr}(AA^+)^{\frac{1}{2}}$ subject to this restriction.

Research supported by the National Science Foundation under Grant No. NSF GP-32996X.

References

1. K. Fan and A. J. Hoffman, Some metric inequalities in the space of matrices, Proc. Amer. Math. Soc., 6 (1955) 111–116.

2. C. Eckart and G. Young, The approximation of one matrix by another of lower rank, Psychometrika, 1 (1936) 211–218.

3. J. B. Keller, Factorization of matrices by least squares, Biometrika, 49 (1962) 239–242.

4. L. Mirsky, Symmetric gauge functions and unitarily invariant norms, Quart. J. Math., Oxford, (2), 11 (1960) 50–59.

5. B. Green, The orthogonal approximation of an oblique structure in factor analysis, Psychometrika, 17 (1952) 429–440.

6. P. Schoenemann, A generalized solution of the orthogonal procrustes problem, Psychometrika, 31 (1966) 1–10.

7. R. Penrose, A generalized inverse for matrices, Proc. Cambridge Philos. Soc., 51 (1955) 406–413; 52 (1956) 17–19.

8. R. Courant and D. Hilbert, Methods of Mathematical Physics, (1), Interscience, New York, 1953, p. 161.

9. G. H. Golub, Least squares, singular values and matrix approximations, Apl. Mat., 13 (1968) 44–51.