Some Applications of the Singular Decomposition of a Matrix
Author(s): I. J. Good
Source: *Technometrics*, Vol. 11, No. 4 (Nov., 1969), pp. 823-831
Published by: American Statistical Association and American Society for Quality
Stable URL: http://www.jstor.org/stable/1266902
Accessed: 17/04/2009 14:25

Some Applications of the Singular Decomposition of a Matrix

I. J. GOOD

Virginia Polytechnic Institute, Blacksburg, Virginia

It is emphasized that the singular decomposition of a matrix has a variety of uses, especially in statistics, although it is seldom mentioned in books on either matrices or statistics. Some applications are surveyed and some new ones are given.

The singular decomposition of an arbitrary matrix is far more useful, both in statistics and in matrix algebra, than is commonly realized. That its importance has been much under-rated is clear since it is mentioned in very few books on either of these subjects. It was perhaps first discussed, in the more advanced form applicable to kernels of integral equations, by E. Schmidt (1907). In that context Smithies (1958) is recommended. The decomposition was applied in the matrix form, to "least squares principal component analysis" by Whittle (1952) and to contingency tables by Good (1965b). In the present paper, which is partly but not entirely expository, some of its applications will be described.

If $\mathbf{M}$ is an arbitrary real $m \times n$ matrix, of rank k, it can be expressed as the sum of k matrices of rank one in a variety of ways. Of these, perhaps the most useful is the "singular decomposition"

$$\mathbf{M} = \mu_1 \boldsymbol{\xi}_1 \mathbf{n}_1' + \mu_2 \boldsymbol{\xi}_2 \mathbf{n}_2' + \cdots + \mu_k \boldsymbol{\xi}_k \mathbf{n}_k' ,$$

where the column vectors $\boldsymbol{\xi}_1 , \boldsymbol{\xi}_2 , \cdots , \boldsymbol{\xi}_k$ are orthonormal (orthogonal and each of length 1) and each has m components, and the row vectors $\mathbf{n}_1' , \cdots , \mathbf{n}_k'$ are orthonormal and each has n components. (We shall throughout denote vectors by bold lower case letters; and matrices, known not to be vectors, by bold upper case letters.) The numbers $\mu_1 , \cdots , \mu_k$ are real and positive and are the square roots of the positive eigenvalues of the $m \times m$ matrix $\mathbf{MM}'$ or of the $n \times n$ matrix $\mathbf{M}'\mathbf{M}$, each of which is a symmetric square semipositive definite matrix. The vectors $\boldsymbol{\xi}_r$ and $\mathbf{n}_r$ are eigenvectors of $\mathbf{MM}'$ and $\mathbf{M}'\mathbf{M}$ respectively. The numbers $\mu_1 , \mu_2 , \cdots ,$ are the *singular values* of $\mathbf{M}$ and the vectors $\boldsymbol{\xi}_1 , \boldsymbol{\xi}_2 , \cdots , \mathbf{n}_1 , \mathbf{n}_2 , \cdots$ are the right and left *singular vectors*. When $\mathbf{M}$ is square and symmetric the singular decomposition reduces to the better known *spectral decomposition,* where the left and right singular vectors are identical and reduce to eigenvectors.

The orthonormal sets $\boldsymbol{\xi}_1 , \cdots , \boldsymbol{\xi}_k$, and $\mathbf{n}_1 , \cdots , \mathbf{n}_k$ can be completed to sets $\boldsymbol{\xi}_1 , \cdots , \boldsymbol{\xi}_m$, and $\mathbf{n}_1 , \cdots , \mathbf{n}_m$ (not uniquely if $k + 2 \leq \max (m, n)$). A *complete* singular decomposition of $\mathbf{M}$ is then $\sum_{i=1}^{m} \mu_i \boldsymbol{\xi}_i \mathbf{n}_i'$, where $\mu_{k+1} = \cdots = \mu_m = 0$

Received Jan. 1968

are the zero singular values. The singular decomposition is of course equal in numerical value to the *complete* singular decomposition.

For complex matrices and even for real or complex quaternion matrices there is a singular decomposition in which the orthogonality conditions are replaced by "unitarity" conditions $\boldsymbol{\xi}_r'\bar{\boldsymbol{\xi}}_s = \delta_r^s$, and the singular values are the square roots of eigenvalues of $\mathbf{M}\bar{\mathbf{M}}'$, but for simplicity only real matrices will be considered in this paper, with minor exceptions. The appropriate modifications for complex matrices will all be fairly obvious.

There is a distinct generalization of the spectral decomposition of a square symmetric matrix. It exists for any square $m \times m$ matrix that has m independent eigenvectors, that is, for matrices that are not *defective*. This is the well-known decomposition

$$\mathbf{M} = \lambda_1\mathbf{x}_1\mathbf{y}_1' + \lambda_2\mathbf{x}_2\mathbf{y}_2' + \cdots,$$

where $\lambda_1, \lambda_2, \cdots$ are the eigenvalues of $\mathbf{M}$, the right eigenvectors of $\mathbf{M}$ are $\mathbf{x}_1, \mathbf{x}_2, \cdots$, the left eigenvectors are $\mathbf{y}_1', \mathbf{y}_2', \cdots$, and $\mathbf{y}_s'\mathbf{x}_r = \delta_r^s$. Instead of being orthonormal the two sets of vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots$, and $\mathbf{y}_1, \mathbf{y}_2, \cdots$, are thus *biorthogonal*. We may refer to this decomposition as the *general spectral decomposition* of a square non-defective matrix. It is quite distinct from the singular decomposition except for symmetric square matrices. It exists only for square non-defective matrices, whereas the singular decomposition always exists.

The singular decomposition of $\mathbf{M}$ provides an immediate analysis of the effect of $\mathbf{M}$ regarded as a linear transformation acting on the vectors of Euclidean n-space, E_n. For any such vector is of the form

$$\mathbf{v} = \beta_1\mathbf{n}_1 + \beta_2\mathbf{n}_2 + \cdots + \beta_n\mathbf{n}_n,$$

where $\mathbf{n}_{k+1}, \cdots, \mathbf{n}_n$ completes the orthonormal basis of E_n, and

$$\mathbf{M}\mathbf{v} = \beta_1\mu_1\boldsymbol{\xi}_1 + \cdots + \beta_k\mu_k\boldsymbol{\xi}_k .$$

Thus $\mathbf{v}$ is first projected into the k-manifold spanned by $\mathbf{n}_1, \cdots, \mathbf{n}_k$, then the k coordinates are scaled by factors $\mu_1, \cdots, \mu_k$, and finally the resulting vector is pictured with the same coordinates in the k-manifold spanned by $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_k$. This last step, so to speak, is an orthogonal transformation. There are six interpretations of this kind corresponding to the six orders in which we can project, scale, and orthogonally transform.

It is trivial to prove that, if q is a positive integer,

$$(\mathbf{M}\mathbf{M}')^q = \mu_1^{2q}\boldsymbol{\xi}_1\boldsymbol{\xi}_1' + \cdots + \mu_k^{2q}\boldsymbol{\xi}_k\boldsymbol{\xi}_k'$$

$$(\mathbf{M}'\mathbf{M})^q = \mu_1^{2q}\mathbf{n}_1\mathbf{n}_1' + \cdots + \mu_k^{2q}\mathbf{n}_k\mathbf{n}_k'$$

$$(\mathbf{M}\mathbf{M}')^q\mathbf{M} = \mu_1^{2q+1}\boldsymbol{\xi}_1\mathbf{n}_1' + \cdots + \mu_k^{2q+1}\boldsymbol{\xi}_k\mathbf{n}_k'$$

$$(\mathbf{M}'\mathbf{M})^q\mathbf{M}' = \mu_1^{2q+1}\mathbf{n}_1\boldsymbol{\xi}_1' + \cdots + \mu_k^{2q+1}\mathbf{n}_k\boldsymbol{\xi}_k'$$

These equations are also true when $q = 0$ provided that we interpret $(\mathbf{M}\mathbf{M}')^0$ and $(\mathbf{M}'\mathbf{M})^0$ as matrices representing the projections of m- and n-space into the k-manifolds spanned by $(\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_k)$ and $(\mathbf{n}_1, \cdots, \mathbf{n}_k)$ respectively. Interpretations of the equations for negative integers q can be obtained by means of generalized matrix inverses (discussed next). More generally, $f((\mathbf{M}'\mathbf{M})^{\frac{1}{2}}) = \sum f(\mu_r)\mathbf{n}_r\mathbf{n}_r'$ for even Laurent series, f, and so on.

Generalized inverses have several applications to statistics: see, for example, Rao (1965), and also Rao (1967) with which the first half of the present paper has a fair amount of overlap. I believe that the simplest treatment of generalized inverses is by means of singular decompositions. Even the definition $\mathbf{M}^+ = \mu_1^{-1}\mathbf{n}_1\boldsymbol{\xi}_1' + \cdots + \mu_k^{-1}\mathbf{n}_k\boldsymbol{\xi}_k'$ suggests itself at once, and is simpler than the definition given by Moore (1935) and Penrose (1955). They defined the generalized inverse $\mathbf{N}$ by the four properties $\mathbf{MNM} = \mathbf{M}$, $\mathbf{NMN} = \mathbf{N}$, $(\mathbf{MN})' = \mathbf{MN}$, $(\mathbf{NM})' = \mathbf{NM}$ and proved that $\mathbf{N}$ exists and is unique. It must of course be of size $n \times m$. All four properties are at once seen to be satisfied by taking $\mathbf{N} = \mathbf{M}^+$, so $\mathbf{M}^+$ is the Moore–Penrose generalized inverse. The uniqueness is easy to prove by making use of the fact that any $n \times m$ matrix is uniquely expressible in the form

$$\sum_{s=1}^{n} \sum_{r=1}^{m} \gamma_{sr}\mathbf{n}_s\boldsymbol{\xi}_r' ,$$

where $(\mathbf{n}_s)$ and $(\boldsymbol{\xi}_r)$ are assigned orthonormal bases of E_n and E_m respectively. (The mn matrices $\mathbf{n}_s\boldsymbol{\xi}_r'$ form a basis for the vector space of $n \times m$ matrices.) Rao (1965) writes $\mathbf{M}^-$ for any "g-inverse" of $\mathbf{M}$, which is any matrix satisfying $\mathbf{MM}^-\mathbf{M} = \mathbf{M}$. It is easy to see that the general g-inverse is

$$\mathbf{M}^- = \mu_1^- \mathbf{n}_1 \boldsymbol{\xi}_1' + \cdots + \mu_p^- \mathbf{n}_p \boldsymbol{\xi}_p',$$

where $p = \min(m, n)$, $\mu_1, \cdots, \mu_p$ are *all* the singular values of $\mathbf{M}$ (non-negative square roots of all the eigenvalues of $\mathbf{MM}'$ if $m \leq n$, or of $\mathbf{M}'\mathbf{M}$ if $m \geq n$), and x^- means x^{-1} if $x \neq 0$ and is otherwise arbitrary.

Rao (1965, p. 25) shows that g-inverses of maximum rank, min (m, n), exist. This is obvious in terms of the singular decomposition, by taking all $\mu_r^- \neq 0$. It is also obvious from the singular decomposition that $\mathbf{M}^+$ *can be uniquely defined as the g-inverse of minimum rank,* since the rank of $\mathbf{M}^+$ is simply the number of μ_r's that do not vanish.

If $\mathbf{M}$ is square and non-defective it is natural to ask whether the general spectral decomposition can be used instead of the singular decomposition, for the representation of generalized inverses. It is easily seen that, in our earlier notation, $\lambda_1^-\mathbf{x}_1\mathbf{y}_1' + \cdots + \lambda_n^-\mathbf{x}_n\mathbf{y}_n'$ is a g-inverse, but $\lambda_1^{-1}\mathbf{x}_1\mathbf{y}_1' + \cdots + \lambda_k^{-1}\mathbf{x}_k\mathbf{y}_k'$ is *not* the Moore–Penrose generalized inverse, where k is the rank of $\mathbf{M}$ and $\lambda_1, \cdots, \lambda_k$ are the non-zero eigenvalues. Thus the singular decomposition is more appropriate in this context than the general spectral decomposition, and I suspect that it is nearly always more useful.

Solution of linear equations.

We can use the singular decomposition of $\mathbf{M}$ to solve the m simultaneous linear equations $\mathbf{Mx} = \mathbf{b}$, where $\mathbf{b}$ is given and $\mathbf{x}$ is to be determined. Write $\mathbf{b} = \beta_1\boldsymbol{\xi}_1 + \cdots + \beta_m\boldsymbol{\xi}_m$ and $\mathbf{x} = \gamma_1\mathbf{n}_1 + \cdots + \gamma_n\mathbf{n}_n$. Then we quickly obtain the consistency conditions $\beta_{k+1} \neq 0, \cdots, \beta_m \neq 0$ and, subject to these conditions, the solution $\gamma_1 = \mu_1^{-1}\beta_1, \cdots, \gamma_k = \mu_k^{-1}\beta_k$ with $\gamma_{k+1}, \cdots, \gamma_n$ arbitrary. Thus

$\mathbf{x} = \mathbf{M}^{-}\mathbf{b}$ is a solution if there is one and *the* general solution is $\mathbf{x} = \mathbf{M}^{-}\mathbf{b} + (\mathbf{M}^{-}\mathbf{M} - \mathbf{I})\mathbf{c}$ where $\mathbf{c}$ is arbitrary (Lanczos, 1958). Rao (1965, p. 24) gives this as a general solution.

Least squares theory.

The ordinary theory of least squares can be expressed in terms of singular decompositions. I here give some of the details of this approach. It seems to me more intuitive than the usual matrix approach in which numerous matrices are multiplied together. Let $\mathbf{z}$ be an observed m-vector whose expected value is $\mathbf{Mc}$, where $\mathbf{M}$ is a known $m \times n$ matrix of rank k ($m \geq n \geq k$) and $\mathbf{c}$ is an n-vector of unknown parameters. Let $S = (\mathbf{z} - \mathbf{Mb})'(\mathbf{z} - \mathbf{Mb})$. We wish to minimize S by an appropriate selection of $\mathbf{b}$ and would like $\mathbf{b}$ to be an unbiased estimator of $\mathbf{c}$. Let the singular decomposition of M be

$$\mathbf{M} = \sum_1^k \mu_j \boldsymbol{\xi}_j \mathbf{n}_j' ,$$

and let the corresponding decompositions of $\mathbf{z}$, $\mathbf{b}$, and $\mathbf{c}$ be

$$\mathbf{z} = \sum_1^m \zeta_j \boldsymbol{\xi}_j \text{ , } \mathbf{b} = \sum_1^n \beta_j \mathbf{n}_j \text{ , } \mathbf{c} = \sum_1^n \nu_j \mathbf{n}_j \text{ ,}$$

where $\boldsymbol{\xi}_{k+1}, \cdots, \boldsymbol{\xi}_m$ and $\mathbf{n}_{k+1}, \cdots, \mathbf{n}_n$ are any unit vectors that complete the orthonormal bases in m-space and n-space respectively. Then, easily,

$$S = \sum_1^k (\zeta_j - \beta_j\mu_j)^2 + \sum_{k+1}^m \zeta_j^2 .$$

Thus S is minimized by taking

$$\mathbf{b} = \sum_1^k \zeta_j \mu_j^{-1} \mathbf{n}_j + \sum_{k+1}^n \beta_j \mathbf{n}_j ,$$

the minimum value of S is $\sum_{k+1}^m \zeta_j^2$, and β_j is an unbiased estimate of γ_j if $j \leq k$. If $\gamma_{k+1}, \cdots, \gamma_n$ are to be estimable there must be $n - k$ linear constraints on the γ's and $\gamma_{k+1}, \cdots, \gamma_n$ must be expressible as known linear functions of $\gamma_1, \cdots, \gamma_k$. Then we can choose $\beta_{k+1}, \cdots, \beta_n$ as the same linear functions of $\beta_1, \cdots, \beta_k$, and thus get an unbiased estimator of $\mathbf{c}$. If $k = n$, then of course this last stage is unnecessary. When $k < n$ we are said to be in the "singular case" of least squares. For other treatments of this case see, for example, Kendall and Stuart, pp. 84–86; and Rao (1965), pp. 181–182.

Solution of matrix equations.

To solve $\mathbf{XX'X} = \mathbf{M}$ where $\mathbf{X}$ clearly must be $m \times n$, we write $\mathbf{X}$ in the form $\sum_{r,s} \gamma_{r,s} \boldsymbol{\xi}_r \mathbf{n}_s'$ and readily deduce that $\gamma_{r,s} = 0$ if $r \neq s$, and that $\gamma_r^3 = \mu_r$. There are therefore precisely 3^k solutions obtained by replacing the non-zero singular values of M by their cube roots in all possible ways. Similarly we obtain $|2q + 1|^k$ solutions for $(\mathbf{XX'})^q\mathbf{X} = \mathbf{M}$ for any integer q, including negative integers if we use generalized inverses. The Moore–Penrose generalized inverse, $\mathbf{M}^+$, is the *unique* solution of $(\mathbf{XX'})^{-}\mathbf{X} = \mathbf{M}$, whatever g-inverse is taken for $(\mathbf{XX'})^{-}$.

If $\mathbf{M}$ is square and not defective the general spectral decomposition can be used in the obvious manner to find the k^p solutions of $\mathbf{X}^p = \mathbf{M}$. For defective square matrices, see Gantmacher (1959), I, p. 231.

It is curious that the singular decomposition is mentioned in very few treatises on matrices; but see Gantmacher (1959) I, 276–278, who applies it to the polar decomposition of a linear transformation. (Even he is not quite explicit.) The polar decomposition of matrices is treated in Vol. II. For a complex matrix it is a generalization of the expression of a complex number in polar coordinates. For a real matrix the polar decomposition is $\mathbf{M} = \mathbf{SO}$, where $\mathbf{S}$ is symmetric and non-negative definite and $\mathbf{O}$ is orthogonal if $\mathbf{M}$ is of rank m, and in any case $\mathbf{O}$ represents an "orthogonal transformation of the $\boldsymbol{\eta}$ space to the $\boldsymbol{\xi}$ space," that is $\mathbf{OO}' = \sum_{r=1}^{k} \boldsymbol{\xi}_r\boldsymbol{\xi}_r'$. Writing $\mathbf{O} = \sum_{r,s=1}^{m} \alpha_{rs}\boldsymbol{\xi}_r\boldsymbol{\eta}_s'$, we readily infer that $\mathbf{O} = \sum_{r=1}^{k} \pm \boldsymbol{\xi}_r\boldsymbol{\eta}_r'$ and then that $S = \sum_{r=1}^{k} \mu_r\boldsymbol{\xi}_r\boldsymbol{\xi}_r'$ and $\mathbf{O} = \sum_{r=1}^{k} \boldsymbol{\xi}_r\boldsymbol{\eta}_r'$.

Calculation of the singular decomposition.

The direct use of the definitions of $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\eta}_k$ as eigenvectors of $\mathbf{MM}'$ and $\mathbf{M}'\mathbf{M}$ is not necessarily the best method for calculating the singular decomposition of $\mathbf{M}$, because the formation of the matrix products involves nm^2 or mn^2 multiplications. Instead we can compute μ_1, ξ_1, and η_1 simultaneously (if μ_1 is $\max_r \mu_r$) by means of the obvious generalization of the well known iterative calculation of the leading eigenvector of a symmetric matrix. That is to say, we take an arbitrary n-vector, $\mathbf{y}_0$, of unit length, form $\mathbf{x}_0 = \mathbf{My}_0$, "normalize" $\mathbf{x}_0$ (scale it to length 1), form $\mathbf{y}_1 = \mathbf{M}'\mathbf{x}_0$, normalize $\mathbf{y}_1$, form $\mathbf{x}_1 = \mathbf{My}_1$ and so on. We see easily, by means of the algebra of the singular decomposition, that we converge at exponential speed to $\boldsymbol{\xi}_1$ and $\boldsymbol{\eta}_1$ unless $\mathbf{y}_0$ happens by bad luck to be orthogonal to $\boldsymbol{\eta}_1$, which is most unlikely if many decimal places are used. Next we form $\mathbf{M} - \mu_1\boldsymbol{\xi}_1\boldsymbol{\eta}_1'$ and repeat the process to get μ_2, $\boldsymbol{\xi}_2$, and $\boldsymbol{\eta}_2$; and so on.

This iterative procedure is extremely easy to program for a computer and will usually be numerically stable when only the first few terms of the singular decomposition are required. (This would cover the case of *principal* components, —see the next paragraph.) If all the terms are wanted then analogues of many of the comments and techniques in Wilkinson (1965), chapter 9, would be relevant. For example, the iterative method is especially pertinent for large sparse matrices.

The least squares property of the singular decomposition, and component analysis.

It is not difficult to prove that the sum of the first j terms ($j = 1, 2, \cdots, k$) of the singular decomposition of $\mathbf{M}$ gives the matrix $\mathbf{N}$ of rank j that best approximates $\mathbf{M}$ in the sense of least squares, i.e. for which we minimize

$$\sum_{r,s} (m_{rs} - n_{rs})^2 = \text{trace } [(\mathbf{M} - \mathbf{N})(\mathbf{M}' - \mathbf{N}')]$$

Moreover, the minimum is $\sum \mu_r^2$ summed for $r > j$. We can apply this least-squares property to a matrix of observations in which each element is diminished by the mean of the elements in its row, and we can then express the above property in the jargon of the statistician thus: We can account for a fraction $(\mu_1^2 + \cdots + \mu_j^2)/(\mu_1^2 + \cdots + \mu_k^2)$ of the total sample variance by means of a matrix of rank j. This presentation of (principal) component analysis was given by Whittle (1952). Previous writers had first formed the sample covariance matrix

MM′ and had worked with its spectral decomposition. The singular decomposition of **M** has the advantages that

(i) it involves less calculation at any rate for the *principal* components (see the section above dealing with the calculation of the singular decomposition);

(ii) if the vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \cdots$, can be given a physical meaning, then one would expect to find associated physical meanings for $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \cdots$, respectively, so that it is natural to refer to $\boldsymbol{\xi}_j$ and $\boldsymbol{\eta}_j$ as a "conjugate" pair of vectors. Compare the suggested "botryological" technique in Good (1965a, pp. 52–53) for putting documents and index terms into conjugate clumps, a technique that would also apply, for example, to patients and symptoms (Good, 1965c). A clump of symptoms would often correspond to a specific complaint or to an advisable treatment and the conjugate clump of patients to those who suffer from this complaint or who would merit this treatment. ("Botryo" is an English prefix meaning a cluster.)

Analysis of contingency tables.

Statistical methodology is poor in *formal* methods for suggesting hypotheses: the statistician often selects his hypotheses (a) in terms of the scientific background of his problem, (b) by the selection of a conventional model, (c) by the selection of a mathematically simple or physically simple model. These three techniques are of course not independent of each other, and often a hypothesis comes under all three headings. This is often true, for example, when the hypothesis of independence of rows and columns is postulated for a contingency table. Note that this hypothesis can be expressed by saying that the contingency table is well approximated by a matrix of rank 1. If this hypothesis is rejected, then the next simplest hypothesis is apparently that it is well approximated by a matrix of rank 2, and so on. Thus the singular decomposition of a contingency table gives rise to a succession of null hypotheses, as pointed out by Good (1965b, p. 64) where χ^2 significance tests for these hypotheses are also proposed. When a hypothesis is correctly accepted we have "improved" the observations. (Cf. Good, 1958.) We can express the "hypothesis of rank j" by saying that the contingency table arose by sampling from a *mixture of j different populations for each of which the rows and columns were statistically independent.* Thus the model is both mathematically and physically simple. This model for the structure of a contingency table seems likely to be relatively more often physically reasonable than the corresponding model for principal component analysis in general. [In error, Good (1965b) implied that the method was analogous to factor analysis instead of to component analysis. Various properties of singular values are given there such as a relationship with eigenvalues due to Weyl (1949), and a minimax property for bilinear forms analogous to a property of quadratic forms due to Courant and Hilbert.] For multidimensional contingency tables see Good (1963). Three-dimensional factor analysis is discussed by Tucker (1964).

The eigenvalues of **AC** *and the determinant* $|\mathbf{I} - \mathbf{AC}|$.

Let **A** and **C** be $m \times m$ matrices where **A** has rank $k < m$. Then, by expressing **A** as the sum of k matrices of rank 1, for example by the singular decomposition, we can reduce the calculation of the eigenvalues of $\mathbf{AC}$ to that of those of a $k \times k$ matrix, in virtue of the following lemma. Similarly the calculation of $|\mathbf{I} - \mathbf{AC}|$ can be reduced to that of a $k \times k$ determinant. This could sometimes make practicable a calculation otherwise impracticable, even on a high-speed computer. It could even happen that the process could be practicable when the order of $\mathbf{A}$ and $\mathbf{C}$ is several thousand. The result has application to statistical problems connected with the Wishart distribution (Good and Jensen, 1968). It should also be useful in the theory of linear transformations in Hilbert space. It is not mentioned by Smithies (1958) so it might be novel in spite of its simplicity.

Lemma (i) *Let* $\mathbf{M}$ *and* $\mathbf{N}$ *be respectively* $m \times k$ *and* $k \times m$ *real or complex matrices. Then* $\mathbf{MN}$ *and* $\mathbf{NM}$ *have the same set of non-zero eigenvectors.* (This is well known when $\mathbf{M}$ and $\mathbf{N}$ are real and $\mathbf{N} = \mathbf{M}'$, and presumably also in general, but I do not know a reference.) *Also, if* $m \geq k$,

$$|\lambda \mathbf{I}_m - \mathbf{NM}| = |\lambda I_k - \mathbf{MN}|\, \lambda^{m-k} \tag{1}$$

where $\mathbf{I}_m$ *and* $\mathbf{I}_k$ *are the* $m \times m$ *and* $k \times k$ *identity matrices.*

(ii) *Let* $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_k, \mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_k$ *be arbitrary m-component column vectors, where* $k < m$. *Then the non-zero eigenvalues of* $(\mathbf{a}_1\mathbf{b}_1' + \cdots + \mathbf{a}_k\mathbf{b}_k')\mathbf{C}$, *where* $\mathbf{C}$ *is* $m \times m$, *coincide with those of the* $k \times k$ *matrix* $\mathbf{D} = \{\mathbf{b}_r'\mathbf{C}\mathbf{a}_s\}$ $(r, s = 1, 2, \cdots, k)$. *Moreover*

$$|\lambda \mathbf{I}_m - (\mathbf{a}_1\mathbf{b}_1' + \cdots + \mathbf{a}_k\mathbf{b}_k')\mathbf{C}| = |\mathbf{I}_k - \mathbf{D}|\, \lambda^{m-k} \tag{2}$$

(iii) If $(\mathbf{A}_1\mathbf{B}_1' + \cdots + \mathbf{A}_k\mathbf{B}_k')\mathbf{C}$ is square, then it has the same set of non-zero eigenvalues as the block matrix $\{\mathbf{B}_r'\mathbf{C}\mathbf{A}_s\}$.

Proof of (i). Let $\mathbf{x}$ be a right eigenvector of M corresponding to a non-zero eigenvalue of $\mathbf{MN}$. Then $\mathbf{MNx} = \lambda\mathbf{x}$, where $\mathbf{Nx} \neq \mathbf{0}$. Therefore $\mathbf{NM}.\mathbf{Nx} = \lambda\mathbf{Nx}$, so λ is an eigenvalue of $\mathbf{NM}$ (and $\mathbf{Nx}$ is a corresponding right eigenvector). Thus every non-zero eigenvalue of $\mathbf{MN}$ is also an eigenvalue of $\mathbf{NM}$, and similarly every non-zero eigenvalue of $\mathbf{NM}$ is also an eigenvalue of $\mathbf{MN}$. The last sentence of Part (i) is now obvious.

Proof of (ii) In (i), let $\mathbf{M} = \mathbf{A}$, where $\mathbf{A}$ consists of the column vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_k$ placed side by side; and let $\mathbf{N} = \mathbf{B}'\mathbf{C}$, where $\mathbf{B}$ consists of the column vectors $\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_k$ placed side by side. Then

$$\mathbf{MN} = \mathbf{AB}'\mathbf{C} = (\mathbf{a}_1\mathbf{b}_1' + \cdots + \mathbf{a}_k\mathbf{b}_k')\mathbf{C}$$

whereas $\mathbf{NM} = \mathbf{B}'\mathbf{CA} = \{\mathbf{b}_r'\mathbf{C}\mathbf{a}_s\} = \mathbf{D}$, and Part (ii) is established.

Note that the case $k = 1$ of Part (ii) of the lemma follows also from the easily proved special case of (1):

$$|\mathbf{I} - \mathbf{AN}| = |\mathbf{I} - \mathbf{NA}|, \quad \text{or} \quad |\lambda\mathbf{I} - \mathbf{AN}| = |\lambda\mathbf{I} - \mathbf{NA}|, \tag{3}$$

where $\mathbf{A}$ and $\mathbf{N}$ are square, by putting $\mathbf{A} = [\mathbf{a}_1, \mathbf{0}]$, $\mathbf{N} = [\mathbf{b}_1, \mathbf{0}]'\mathbf{C}$, that is, by making $\mathbf{a}_1$ and $\mathbf{b}_1$ square by "filling up with zeros". (The identity (3) is immediate if $\mathbf{A}$ is non-singular, and follows generally by a continuity argument.) The case $k = 1$, $\lambda = 1$, of Part (ii) occurs also in Anderson (1958), p. 108. Also Scheffé (1959), p. 417, mentions that $1 + \mathbf{z}'\mathbf{A}^{-1}\mathbf{z} = |\mathbf{A} + \mathbf{zz}'|/|\mathbf{A}|$, and refers to

Wilks (1932), pp. 487–488, for a proof. This again is the case $k = 1$, $\lambda = 1$. More generally, from Equation (2), $|\lambda\mathbf{I} + \mathbf{Z}'\mathbf{A}^{-1}\mathbf{Z}| = |\lambda\mathbf{A} + \mathbf{ZZ}'|/|\mathbf{A}|$.

Part (ii) can be usefully combined with the singular decomposition of $\mathbf{M}$ in order to find the eigenvalues of $\mathbf{MC}$.

Part (iii) follows from Part (ii) by noticing that a matrix product $\mathbf{AB}'$, where $\mathbf{A}$ and $\mathbf{B}$ are $p \times m$, is the sum of m terms of the form $\mathbf{ab}'$.

The continuous analogue of Part (ii) of the lemma is that the non-zero eigenvalues of $\int b(x, u)c(u, y)\, du$ are the same as those of $\int c(x, u)b(u, y)\, du$, and the non-zero eigenvalues of the kernel $\int a(x, u)\, c(u, y)\, du$, where $a(x, y)$ is degenerate, $a(x, y) = \sum_{r=1}^{k} a_r(x)b_r(y)$, are the same as those of the matrix

$$\{\int b_r(x)c(x, y)a_s(y)\, dx\, dy\}.$$

As a neat algebraic example of Part (i) it is perhaps worth showing that the equation

$$\mathbf{AB} - \mathbf{BA} = c\mathbf{I} \text{ (} c \text{ real or complex, } c \neq 0) \tag{4}$$

cannot be true for (square) matrices of finite order. This is at once clear when $\mathbf{A}$ and $\mathbf{B}$ are symmetric since then $\mathbf{AB} - \mathbf{BA}$ is skew-symmetric. In quantum mechanics $\mathbf{A}$ and $\mathbf{B}$ are usually assumed to be Hermitian and then $\mathbf{AB} - \mathbf{BA}$ is skew-Hermitian, but this observation does not rule out its being $c\mathbf{I}$ if c is purely imaginary. But without restrictions on $\mathbf{A}$ and $\mathbf{B}$, Equation (4) implies that, for each real or complex λ,

$$|\mathbf{AB} - \lambda\mathbf{I}| = |\mathbf{BA} - (\lambda - c)\mathbf{I}|$$

so that the set of eigenvalues of $\mathbf{BA}$ is the same as that of $\mathbf{AB}$ but slid c to the left. But these two sets are the same, apart from zeros, which is impossible if they are both finite and contain more than one point. The case $\mathbf{AB} = \mathbf{0}$ requires special but trivial treatment: it implies $\mathbf{BA} = -c\mathbf{I}$ and hence $\mathbf{AB} = -c\mathbf{I}$, a contradiction.

Acknowledgment

I am indebted to Dr. Donald R. Jensen for some useful comments.

References

[1] Anderson, T. W. (1958), *Introduction to Multivariate Statistical Analysis* (New York: Wiley; London: Chapman and Hall).

[2] Gantmacher, F. R. (1934), "On the algebraic analysis of Krylov's method of transforming the secular equation," *Trans. Second Math. Congress, II*, 45–48.

[3] Gantmacher, F. R. (1949), *The Theory of Matrices*, two volumes (New York: Chelsea).

[4] Good, I. J. (1958), "The interaction algorithm and practical Fourier analysis," *J. Roy. Statist. Ser. B, 20*, 361–372; *22* (1960), 372–375.

[5] Good, I. J. (1963), "Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables," *Ann. Math. Statist. 34*, 911–934.

[6] Good, I. J. (1965a), "Speculations concerning the first ultraintelligent machine," *Advances in Computers, 6*, 31–88.

[7] Good, I. J. (1965b), *The Estimation of Probabilities: an essay on modern Bayesian methods* (M. I. T. Press).

[8] Good, I. J. (1965c), "Categorization of classification," in *Mathematics and Computer Science in Biology and Medicine* (London: H. M. S. O.), 115-128.
[9] Good, I. J. and Jensen, D. R. (1968), "Some distributions related to that of Wishart."
[10] Kendall, M. G. and Stuart, A. (1961), *The Advanced Theory of Statistics* (London: Griffin).
[11] Lanczos, C. (1958), "Linear systems in self-adjoint form," *Amer. Math. Monthly, 65*, 665–679.
[12] Moore, E. H. (1935), *General Analysis* (Philadelphia: American Philos. Soc.).
[13] Penrose, R. (1955), "A generalized inverse for matrices," *Proc. Cambridge Philos. Soc. 51*, 406–413.
[14] Rao, C. R. (1965), *Linear Statistical Inference and its Applications* (New York: Wiley).
[15] Rao, C. R. (1967), "Calculus of generalized inverses of matrices, Part I—general theory," *Sankhya, Ser. A, 29*, 317–342.
[16] Scheffé, H. (1959), *The Analysis of Variance* (New York: Wiley).
[17] Schmidt, E. (1907), "Zur theorie der linearen und nichtlinearen Integralgleichungen. Erster Tei", *Math Annalen, 63*, 433–476.
[18] Smithies, F. (1958), *Integral Equations* (Cambridge: University Press).
[19] Tucker, L. R. (1964), "The extension of factor analysis to three-dimensional matrices," in *Contributions to Mathematical Psychology* (ed. N. Frederiksen and H. Gulliksen; New York: Holt, Rinehart and Winston), 110–127.
[20] Weyl, H. (1949), "Inequalities between the two kinds of eigenvalues of a linear transformation," *Proc. Nat. Acad. Sci. 35*, 408–411.
[21] Whittle, Peter (1952), "On principal components and least square methods of factor analysis." *Skand. Aktuar 35*, 223–239.
[22] Wilkinson, J. H. (1965), *The Algebraic Eigenvalue Problem* (Oxford: Clarendon Press).
[23] Wilks, S. S. (1932), "Certain generalizations in the analysis of variance," *Biometrika 24*, 471–494.