Bell Laboratories

Cover Sheet for Technical Memorandum

The information contained herein is for the use of employees of Bell Laboratories and is not for publication (see GEI 13.9-3)

Title– Approximation of a ~~Real~~ Symmetric Matrix by a Positive Semidefinite Matrix of Prescribed Rank

Date–April 1, 1974

TM– 74-1229-10

Other Keywords –

Author(s)	Location and Room	Extension
J. DeLeeuw*	MH 7F-402	3948

Charging Case – 25952

Filing Case – 25952

ABSTRACT

In this paper we summarize some known results on the approximation of real square symmetric matrices by matrices of a lower rank. An application to multidimensional scaling is mentioned, but not analyzed in detail. This particular application requires a result which is slightly more general then the classical ones. The result is proved in detail.

*On leave of absence from the University of Leiden, Leiden, The Netherlands.

Pages Text 6 Other 5 Total 11

No. Figures -- No. Tables -- No. Refs. 9

Address Label

Bell Laboratories

subject: Approximation of a Real Symmetric Matrix by a Positive Semidefinite Matrix of Prescribed Rank-- Case 25952

date: April 1, 1974

from: J. DeLeeuw*

TM74-1229-10

MEMORANDUM FOR FILE

0. Introduction and Summary

Many computational procedures in data analysis (especially in the factor analysis and scaling area) need to solve the problem of approximating (in the least-squares sense) a given matrix by a matrix of rank not larger than a prescribed p. For rectangular or asymmetric square matrices, the solution, based on singular value or Eckart-Young decomposition, is well known. For symmetric positive semidefinite matri~~ces of order n,~~ the solution, based on spectral decomposition, is well known, too. The best approximation is simply a matrix with the same roots and vectors, except for the n-p smallest roots, which are set equal to zero. ~~The only~~ A problem which is not exactly classical is the approximation of a general symmetric matrix by a positive semidefinite symmetric matrix of rank ~~for~~ not greater than p. It is proved in this paper that the classical results still hold if the matrix has at least p positive eigen values. If it has $q < p$ positive eigen-values, the best positive semidefinite symmetric

*On leave of absence from the University of Leiden, Leiden, The Netherlands.

approximations of rank $q+1, q+2, \ldots, n$ are all the same. They are obtained by using the q roots and vectors corresponding to positive roots, while all other roots and vectors are set to zero. In terms of approximat~~ely~~ing a symmetric matrix by a product XX', with X an $n \times p$ matrix, this means that X has $p-q$ columns with zeroes and q columns with the appropriately scaled vectors corresponding to positive roots.

1. Notation

We use the notation S_{np} for the (closed) set of all $n \times n$ real symmetric matrices of rank $\rho \leq p$, the set S^+_{np} is the subset of all positive semidefinite (psd) matrices of rank $\rho \leq p$. We also use the notation $C \gtrsim 0$ to indicate that C is psd. For any $n \times n$ real symmetric matrix C we use $\lambda_1(C) \geq \ldots \geq \lambda_n(C)$ for the ordered eigenvalues, and $\gamma_1(C) \geq \ldots \geq \gamma_n(C)$ for the ordered absolute values of the eigenvalues. If no confusion is possible we leave out the argument C. We define for each real number x the decomposition $x = x^+ - x^-$ with $x^+ = \frac{1}{2}(|x| + x)$ and $x^- = \frac{1}{2}(|x| - x)$. Clearly ~~both~~ $x^+ \geq 0$ with equality iff $x \leq 0$, and $x^- \geq 0$ with equality iff $x \geq 0$. Moreover, $x^+x^- = 0$ and $x^2 = (x^+)^2 + (x^-)^2$. We apply this decomposition to the canonical form $C = K\Lambda K'$ with $K'K = KK' = I$ and Λ diagonal. This gives $C = C^+ - C^- = K\Lambda^+K' - K\Lambda^-K'$. The decomposition is unique, both C^+ and C^- are psd, and $C^+C^- = C^-C^+ = 0$. Finally we use R_n^m for the linear space of all real $n \times m$ matrices, S_n ~~is~~ the subspace of R_m^n which consists of the symmetric matrices.

2. The Basic Distance Functions

For each $C \in S_n$ we define

$$\mu_p(C) = \inf\{tr[(C-\hat{C})^2] \mid \hat{C} \in S_{np}\},$$

$$\mu_p^+(C) = \inf\{tr[(C-\hat{C})^2] \mid \hat{C} \in S_{np}^+\}$$

$$= \inf\{tr[(C-XX')^2] \mid X \in R_n^p\}.$$

For each $A \in R_n^m$ we define (assuming $p \leq m \leq n$)

$$\eta_p(A) = \inf\left\{tr[(A-XY')'(A-XY')] \mid X \in R_n^p,\ Y \in R_m^p\right\}.$$

The least squares version of the Eckart-Young Theorem tells us that

$$\eta_p(A) = \sum_{r=p+1}^{m} \lambda_r(A'A) = \sum_{r=p+1}^{n} \lambda_r(AA').$$

For symmetric matrices the Eckart-Young Theorem gives

$$\mu_p(C) = \sum_{r=p+1}^{n} \gamma_r^2(C).$$

If $C \gtrsim 0$ then obviously

$$\mu_p(C) = \mu_p^+(C) = \sum_{r=p+1}^{n} \lambda_r^2(C)$$

As a matter of fact this result remains true under the much weaker condition that $\lambda_p(C) \geq 0$. In this paper we are interested in the remaining case, in which we want to compute $\mu_p^+(C)$ for a matrix C with $\lambda_p(C) < 0$.

3. The Remaining Case

Clearly if $\hat{C} \gtrsim 0$ then

$$\begin{aligned} = tr[(C-\hat{C})^2] &= tr[(C^+-C^--\hat{C})^2] \\ &= tr[(C^-)^2] + 2tr[C^-\hat{C}] + tr[(C^+-\hat{C})^2] \\ &\geq tr[(C^-)^2] + tr[(C^+-\hat{C})^2] \end{aligned}$$

and ~~consequently~~

$$\mu_p^+(C) \geq \sum_{r=1}^{n} (\lambda_r^-)^2$$

$$= \sum_{r=1}^{p} (\lambda_r^-)^2 + \sum_{r=p+1}^{n} \lambda_r^2 .$$

Define $\hat{C}_p^+$ to be the best rank p approximation to C^+ if rank $(C^+) \geq p$ and define $\hat{C}_p^+ = C^+$ if rank $(C^+) < p$. Then

$$\mu_p^+(C) \leq \mathrm{tr}[(C-\hat{C}_p^+)^2] = \sum_{r=1}^{p} (\lambda_r^-)^2 + \sum_{r=p+1}^{n} \lambda_r^2.$$

This clearly implies

$$\mu_p^+(C) = \sum_{r=1}^{p} (\lambda_r^-)^2 + \sum_{r=p+1}^{n} \lambda_r^2 .$$

The infimum is attained for $\hat{C} = \hat{C}_p^+$. (If $C \lesssim 0$ then this implies that the infimum is attained for $\hat{C} = 0$.) In terms of

$$\mu_p^+(C) = \inf\left\{\mathrm{tr}[(C-XX')^2] \;\middle|\; X \in R_n^p\right\}$$

this means that the infimum is attained by letting $X = K_p\Lambda_p^{+\frac{1}{2}}$, where K_p and Λ_p^+ refer to the parts of K and Λ^+ corresponding with the p largest elements of Λ^+. If rank $(C^+) = \rho < p$ this obviously means that X has $p - \rho$ columns equal to zero. If $C \lesssim 0$ it even means that X = 0 gives the optimal solution.

4. Application

In metric multidimensional scaling the model is usually transformed (by squaring and doubly centering) into an equation of the form

$$C = XX' + E$$

where the "errors" in E are supposed to be small, and the number of columns in X is unspecified. We then compute the eigenvalues and eigenvectors of C, select an appropriate dimensionality on the basis of these results, and compute X by the rule $K\Lambda^{\frac{1}{2}}$. Negative eigenvalues are always ignored as being "error" in some sense, large negative eigenvalues are taken as evidence that the model does not fit the data very well. In this procedure it is clear that no precisely defined least squares problem is solved, because a precise definition would mean a prior specification of the dimensionality p. Due to the fortunate mathematical properties of eigenvectors this is not very serious (a successive solution gives the same eigenvectors as a simultaneous solution for fixed p). Moreover in practical situations we will almost always have $\lambda_p(C) \geq 0$. There is, however, nothing in the procedure of forming C which guarantees that this is true. This becomes more important in applications of the basic approximation method to nonmetric scaling where it is necessary to specify p in advance. For a rigorous solution of these problems we have to consider the posssibility that $\lambda_p(C) < 0$, and in this case our "optimal" configuration must have a number of zero columns (it would be interesting to know how this possibility is handled in TORSCA and KYST, which use the approximation procedure in the early stages). The solution outlined in this paper is even more important for some of the new scaling methods currently being developed which use the approximation in each iteration until convergence, and not only to compute an initial approximation. It was with this last problem in mind that this paper was written.

MH-1229-JD-pb J. DeLeeuw

Att.
References

REFERENCES

The paper by Schönemann, Bock, and Tucker on the Eckart-Young Theorem has an excellent list of references to various applications and to some of the mathematical literature (Schönemann, P. H., Bock, R. D., and Tucker, L. R. Some notes on a Theorem by Eckart and Young. Psychometric Laboratory, University of North Carolina, Research Memo 25, July 1965). We give some *additional* references to the mathematical literature dealing with the general problem of approximation in matrix space, and with the approximation by a matrix of fixed, predescribed rank in particular. All our results can essentially be found (with different proofs, and less explicit) in a nice paper by J. B. Keller (Factorization of matrices by least squares, *Biometrika*, 1962, *49*, 239-242). Several interesting contributions in this field have been collected in the proceedings of "Programmation en Mathematiques Numériques," Colloque Internationale du Centre National de la Recherche Scientifique, No. 165, Besancon, 7-14 Septembre 1966, Editions du CNRS, Paris, 1968. In the following list references to these proceedings are indicated by the letters PMN, followed by the page numbers. We also use the abbreviation CRAS for the Comptes Rendus de l'Académie des Sciences, Paris. There is some additional related literature on the distance of a matrix to the set of singular matrices which we mention only briefly (P. Franck, CRAS 253 and CRAS 256; J. L. Rigal, CRAS 252).

1. Metric Problems in Matrix Space (general)

A. S. Householder: Some applications of the theory of Norms PMN, p. 27-36.

M. Fiedler: Metric problems in the space of matrices PMN, p. 93-103.

L. Mirsky: Symmetric gauge functions and unitary invariant norms
Quart. J. Math, 1960, 11, 50-59.

2. Approximation by a Matrix of Predescribed Rank

M. Fiedler and V. Ptak: Sur la meilleure approximation des transformations linéaires par des transformations de rang préscrit.
CRAS, 1962, 254, 3805-3807.

J. Gaches, J. L. Rigal and X. Rousset de Pina: Distance Euclidienne d'une application linéaire au lieu des applications de rang γ.
CRAS, 1965, 260, 5672-5674.

J. F. Maitre: Approximation de rang donné dans un espace de matrices.
PMN, p. 105-110.

J. F. Maitre and Nguyen Huu Vinh: Valeurs singulieres généralisées et meilleure approximation de rang r d'une opérateur linéaire.
CRAS, 1966, 262, 502-504.

J. F. Maitre and Nguyen Huu Vinh: Evaluation de la distance
d'une matrice à l'ensemble des
matrices de rang r.
CRAS, 1966, 262, 910-912.

J. F. Maitre and Nguyen Huu Vinh: Distance d'une matrice a
l'ensemble des matrices de rang r
au sense de norme $S_{\psi\varphi}$.
Actes 5^{e} congres
AFIRO (Lille 1966), p. 477-480, Editions
AFIRO, Paris, 1966.