KYBERNETIKA — VOLUME 15 (1979), NUMBER 5

On the Pseudoinverse of a Sum of Symmetric Matrices with Applications to Estimation

PAVEL KOVANIC

A new formula for the pseudoinverse of a sum of symmetric matrices is presented, valid for arbitrary symmetric matrices without any restrictions relating to their column — or row — spaces. As an application of this formula a generalized version of the estimate minimizing the penalty is developed. This makes it possible to show that in a general case of estimation the problem is decomposed into two independent problems. One of them is related to data belonging to a subspace containing signal components but no noise components. This part of the problem can be easily solved, the result of estimation performed on this part of data being error-free.

INTRODUCTION

An important role in linear estimation theory is played by a symmetric matrix

$$K = XX^{\mathrm{T}} + V \tag{1}$$

which is the covariance matrix of an observation vector

$$y = y_{\mathrm{x}} + y_{\mathrm{e}}\,, \tag{2}$$

where the two vectors y_{x} and y_{e} represent random signal and noise, respectively. The noise covariance matrix V often contains the signal covariance matrix XX^{T} in its row-space (as well as in its column-space, they both are symmetrical). If V is nonsingular, this condition is fulfilled automatically. Then a "best" linear estimate of different functions of the y_{x} from observations y may be obtained by different generalizations of Gauss-Markov Theorem [1], [2], [3]. But also a more general situation deserves attention. For the model

$$y_{\mathrm{x}} = Xc\,, \tag{3}$$

where c is a nonrandom vector of parameters, Zyskind [4] showed that constraints on parameters lead to the problem of singular covariance matrix. As shown by Hallum, Lewis and Boullion [5], for such a model a minimum variance estimate of the c, with c restricted by linear restrictions and with the covariance matrix of y_e having an arbitrary rank, may be calculated directly without the need of a linear transformation used by other authors to obtain a linear model of smaller dimensions with a full rank covariance matrix. It is the purpose of this paper to show that mentioned result as well as a more general result can be obtained using an extension of a theorem on pseudo-inverse of a sum of symmetrical matrices.

PSEUDO-INVERSE OF A SUM OF SYMMETRICAL MATRICES

The "pseudo-inverse" or the "Moore-Penrose inverse" is a (unique) matrix A^+ satisfying four conditions

$$AA^+A = A\,, \tag{4}$$

$$A^+AA^+ = A^+\,, \tag{5}$$

$$(AA^+)^T = AA^+\,, \tag{6}$$

$$(A^+A)^T = A^+A\,. \tag{7}$$

The following theorem is presented in literature on generalized inverse of matrices [6]:

Theorem. If X is an $n \times q$ matrix contained in the column-space of an $n \times n$ symmetrical matrix V, then

$$(V + XX^T)^+ = V^+ - V^+X(I + X^TV^+X)^{-1}X^TV^+\,. \tag{8}$$

"To be in the column-space" means here the same as

$$VV^+X = X\,. \tag{9}$$

In a more general case the matrix $(I - VV^+)\,X$ does not equal to the zero matrix. For such a case the following extension holds:

Theorem. If V is an $n \times n$ symmetrical matrix and if X is an arbitrary $n \times q$ real matrix, then

$$(V + XX^T)^+ = V^+ - V^+X(I + X^TV^+X)^{-1}\,X^TV^+ + (X_\perp^+)^TX_\perp^+\,, \tag{10}$$

where

$$X_\perp = (I - VV^+)\,X\,. \tag{11}$$

Proof. It can be easily verified by substitution into (4)–(7) that if

$$A^{\mathrm{T}}B = O \tag{12}$$

and

$$B^{\mathrm{T}}A = O\,, \tag{13}$$

then

$$(A + B)^+ = A^+ + B^+ \tag{14}$$

for any matrices A and B having appropriate dimensions. Using relations

$$VV^+XX^{\mathrm{T}}VV^+ = XX^{\mathrm{T}}VV^+ = VV^+XX^{\mathrm{T}} \tag{15}$$

resulting from the symmetry of the matrix XX^{T}, one may write

$$\begin{aligned} V + XX^{\mathrm{T}} &= V + VV^+XX^{\mathrm{T}} + (I - VV^+)XX^{\mathrm{T}} = \\ &= V + VV^+XX^{\mathrm{T}}VV^+ + VV^+XX^{\mathrm{T}}(I - VV^+) + \\ &+ (I - VV^+)XX^{\mathrm{T}}VV^+ + (I - VV^+)XX^{\mathrm{T}}(I - VV^+) = \\ &(V + VV^+XX^{\mathrm{T}}VV^+) + ((I - VV^+)XX^{\mathrm{T}}(I - VV^+))\,. \end{aligned} \tag{16}$$

The first bracketed term is orthogonal to the last one because of the property (6) of the pseudoinverse V^+ and because of the symmetry of the V. Therefore

$$(V + XX^{\mathrm{T}})^+ = (V + VV^+XX^{\mathrm{T}}VV^+)^+ + ((I - VV^+)XX^{\mathrm{T}}(I - VV^+))^+\,. \tag{17}$$

The matrix $VV^+XX^{\mathrm{T}}VV^+$ is in the column-space of the matrix V, therefore (8) may be applied:

$$\begin{aligned} &(V + VV^+XX^{\mathrm{T}}VV^+)^+ = \\ &= V^+ - V^+(VV^+X)(I + (VV^+X)^{\mathrm{T}}\,V^+(VV^+X))^{-1}\,(VV^+X)^{\mathrm{T}}\,V^+ = \\ &= V^+ - V^+X(I + X^{\mathrm{T}}V^+X)^{-1}\,X^{\mathrm{T}}V^+\,. \end{aligned} \tag{18}$$

The second right-hand term of (17) may be rewritten as

$$(X_\perp X_\perp^{\mathrm{T}})^+ = (X_\perp^+)^{\mathrm{T}}\,X_\perp^+ \tag{19}$$

using (11) and a known property of the pseudo-inverse. The proof is complete.

STATISTICAL APPLICATIONS

A. The Minimum Penalty Estimate

As shown in [2], [3], a generalized estimator called the minimum penalty estimator (MPE) exists from which a large class of different known linear estimators can be obtained as particular cases. This estimator has been developed under assumption "there are no observed signals not corrupted by noise". Using the formula (10) for the same generalized problem as in [3], we can come to an MPE not restricted by such an assumption:

Observed data are $n \times p$ random matrices

$$Y = Y_x + Y_e \,, \tag{20}$$

where Y_x represents random signals and Y_e is given by random error components and by noise. Requirements relating to results of estimation are characterized by a $t \times p$ matrix

$$Z_x = \tau_x\{Y_x\} \tag{21}$$

for a case when noise disappears, and by a matrix

$$Z_0 = \tau_0\{Y_x, Y_e\} \tag{22}$$

for the case with a non-zero noise. Symbols τ_x and τ_e denote some given operators. To solve the problem we need only correlations of required results of estimation with data.

The estimator will have a general linear form

$$Z = WY + C \,, \tag{23}$$

where W and C are some constant matrices having dimensions $t \times n$ and $t \times p$, respectively. We proceed in the same way as in [3]. To evaluate the quality of the estimate we introduce the norm $\|E\|$ of an error matrix E in the following manner:

$$\|E\| = \mathrm{tr}\{\langle EQE^{\mathrm{T}}\rangle\}^{1/2} \tag{24}$$

where Q is a given positive definite weighting matrix, the brackets $\langle\cdot\rangle$ denote the averaging, and $\mathrm{tr}\{\cdot\}$ states for the trace of a matrix. The error of the first type

$$E_x = WY_x + C - Z_x \tag{25}$$

relates to ideal situations with no noise while the error of the second type

$$E_0 = W(Y_x + Y_e) + C - Z_0 \tag{26}$$

is influenced by actual noise components. To take into account both errors one uses the penalty

$$p = p_0\|E_0\|^2 + p_x\|E_x\|^2 \tag{27}$$

with a positive weight p_0 and with a weight p_x satisfying a condition

$$p_0 + p_x > 0 \,. \tag{28}$$

It can be shown like in [3] that the constant matrix C in (23) vanishes after an appropriate centralization of variables. We assume below such a centralization as having been performed according to the formulae given in [3]. Then the penalty (27) is minimized if the equation

$$WM = p_0\langle Z_0 Q Y^T \rangle + p_x \langle Z_x Q Y_x^T \rangle \tag{29}$$

holds. Here the M states for a weighted sum of covariance matrices

$$M = p_0 \langle Y Q Y^T \rangle + p_x \langle Y_x Q Y_x^T \rangle . \tag{30}$$

It follows from this definition that there are no observed vectors outside the range-space of the M, therefore one may take

$$WMM^+ = W . \tag{31}$$

Then

$$W = \left(p_0 \langle Z_0 Q Y^T \rangle + p_x \langle Z_x Q Y_x^T \rangle\right) M^+ \tag{32}$$

is the minimum penalty estimator. In [3] a pseudo-inverse of M has been obtained using (8) but we are able now to discuss a more general situation by means of the extended formula (10).

Denoting

$$B_0 = \langle Y_e Q Y_e^T \rangle , \tag{33}$$

$$B = B_0 + \langle Y_x Q Y_e^T \rangle + \langle Y_e Q Y_x^T \rangle , \tag{34}$$

and using a square-root matrix X determined by a decomposition

$$XX^T = \langle Y_x Q Y_x^T \rangle , \tag{35}$$

one has

$$M = p_0 B + (p_0 + p_x) XX^T . \tag{36}$$

Denote $\mathscr{R}(A)$ the range space and $\mathscr{N}(A)$ the zero space of an $n \times m$ matrix A,

$$\mathscr{R}(A) = \{y \in \mathscr{R}^n : y = Ax, x \in \mathscr{R}^m\} , \tag{37}$$

$$\mathscr{N}(A) = \{x \in \mathscr{R}^m : 0 = Ax\} . \tag{38}$$

The range space $\mathscr{R}(XX^T)$ is a "signal space" while the range space $\mathscr{R}(B_0)$ is a "noise space". Consider four subspaces of the n-dimensional vector space $\mathscr{R}^n$:

$\mathscr{S}_1 = \mathscr{R}(XX^T) \cap \mathscr{R}(B_0)$ (containing both signal and noise),

$\mathscr{S}_2 = \mathscr{R}(XX^T) \cap \mathscr{N}(B_0)$ (containing signals but no noise),

$\mathscr{S}_3 = \mathscr{N}(XX^T) \cap \mathscr{R}(B_0)$ (containing noise but no signal),

$\mathscr{S}_4 = \mathscr{N}(XX^T) \cap \mathscr{N}(B_0)$ (containing no signal and no noise).

In a general case, only the subspace $\mathscr{S}_4$ is empty (for a nonregular matrix M). Formula (10) may be applied to obtain

$$(39)\qquad M^+ = (1/p_0)\, B^+(I - X_1(s_1 I + X_1^T B^+ X_1)^{-1}\, X_1^T B^+) = 1/(p_0 + p_x)\, X_2^{+T} X_2^+ \,,$$

where

$$(40)\qquad s_1 = p_0/(p_0 + p_x)\,,$$

$$(41)\qquad X_1 = BB^+X\,,$$

and

$$(42)\qquad X_2 = (I - BB^+)\,X\,.$$

As above in (18) there is no necessity to distinguish the X_1 from the X when substituted into the first term of (39).

After substitution of M^+ into (29) one has the minimum penalty estimate valid for the general case under consideration.

B. Noise-Free Estimation

An interesting aspect is worth to be analyzed in more details. Let

$$(43)\qquad \langle Z_0 Q Y^T \rangle = \langle Z_x Q Y_x^T \rangle = L_1 X_1^T + L_2 X_2^T$$

with the same X_1 and X_2 as above.

Then

$$(44)\qquad W = W_1 + W_2\,,$$

where

$$(45)\qquad W_1 = L_1(s_1 I + X_1^T B^+ X_1)^{-1}\, X_1^T B^+$$

and

$$(46)\qquad W_2 = L_2 X_2^+\,.$$

Both components of the estimator W have the same structure. To see this write $(\cdot)^+$ instead of $(\cdot)^{-1}$ and represent W_2 in the form

$$(47)\qquad W_2 = L_2(s_2 I + X_2^T B_2^+ X_2)^+\, X_2^T B_2^+$$

with $s_2 = 0$ and $B_2 = I$, a trivial covariance matrix. We have two estimators having two mutually orthogonal domains, $\mathscr{R}(B)$ and

$$(48)\qquad \mathscr{R}(X_2) = \mathscr{R}((I - BB^+)\,X) = \mathscr{N}(B)\,.$$

An arbitrary observed data matrix may be written as

$$Y = BB^{+}Y + (I - BB^{+})\,Y = Y_1 + Y_2\,. \tag{49}$$

The result of estimation is therefore

$$Z = W_1 Y_1 + W_2 Y_2 = Z_1 + Z_2\,. \tag{50}$$

This means that our estimating problem is decomposed into two independent problems. Moreover, if noise is uncorrelated with random components of signals,

$$\langle Y_e Q Y_x^{\mathrm{T}} \rangle = 0\,, \tag{51}$$

then

$$(I - B_0 B_0^{+})\,Y_e = 0\,. \tag{52}$$

The projection Y_2 of observed data matrix onto the subspace $\mathcal{N}(B_0)$ is therefore

$$(I - B_0 B_0^{+})\,Y_x = (I - B_0 B_0^{+})\,X A_2 = X_2 A_2\,, \tag{53}$$

where A_2 is a certain matrix. We have thus

$$Z_2 = L_2 X_2^{+} X_2 A_2\,. \tag{54}$$

For a matrix X_2 having a full rank

$$X_2^{+} X_2 = I\,, \tag{55}$$

and we obtain

$$Z_2 = L_2 A_2\,. \tag{56}$$

But this is exactly what we wanted to obtain as a result of estimation performed on data occupying the subspace $\mathcal{N}(B_0)$. Actually,

$$\langle Z_2 Q Y_2^{\mathrm{T}} \rangle = L_2 \langle A_2 A_2^{\mathrm{T}} \rangle X_2^{\mathrm{T}} = L_2 X_2^{\mathrm{T}} \tag{57}$$

because of the definition (35) of the matrix X.

We are comming to the conclusion that the estimation performed on data belonging to a noise-free subspace $\mathcal{N}(B_0)$ is completely error-free as has been stated intuitively in [3].

(Received December 8, 1978.)

REFERENCES

[1] T. O. Lewis, P. L. Odell: A generalization of the Gauss-Markov theorem. Amer. Stat. Ass. Journal (Dec. 1966), 1063—1066.

[2] P. Kovanic: Minimum penalty estimate. Kybernetika *8* (1972), 5, 367—383.

[3] P. Kovanic: Generalized linear estimate of functions of random matrix arguments. Kybernetika *10* (1974), 4. 303—316.

[4] G. Zysking: On canonical forms, non-negative covariance matiices and best and simple least squares linear estimators in linear models. Ann. Math. Stat. *38* (1966), 1092—1108.

[5] C. R. Hallum, T. O. Lewis, T. L. Boullion: Estimation in the general linear model with a positive semidefinite covariance matrix. Communications in statistics *1* (*2*), (1973) 157—166.

[6] R. M. Pringle, A. A. Rayner: Generalized inverse matrices with applications to statistics. Griffin, London 1971.

Ing. Pavel Kovanic, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.