KYBERNETIKA — VOLUME 10 (1974), NUMBER 4

Generalized Linear Estimate of Functions of Random Matrix Arguments

PAVEL KOVANIC

The optimum linear estimation problem in a generalized formulation is considered for a function of observational data matrix being a summ of random matrix signals and of a matrix of observation errors or noise. The criterium of optimality is the minimization of the penalty which is a linear combination of the squares of the norms of two kinds of errors: of the error matrix which would take place in estimation without noise and of the matrix of actual estimating errors. Results are applicable also in the case of deficient rank of covariance matrices and/or of signal model. Correlation of signals with the noise is allowed. A priori statistics of signals can be incorporated to improve the estimates. It is shown that different known generalizations of least squares estimates are special cases of the minimum penalty estimate.

INTRODUCTION

Importance of the linear estimation problem lies not only in its direct applications for various practical purposes but also in the necessity to include state and parameter estimations procedures into "higher" techniques such as, e.g., identification and optimal control. Also the modern state estimation methods themselves and the related theory of optimum linear filters are based on the classical method of least squares [1]. Making use of the minimum mean-square error estimate (MSE) and of its generalization (GMSE) is most popular. But this approach is not an unique alternative and — under certain conditions — it may be not the best alternative.

There are two kinds of errors arising in estimation: The *first* one is the error in treatment of data or signals *not corrupted by noise* and another statistical errors. The *second* one arises in the treatment of data or signals *in the presence of noise.* The MSE technique requires the minimization of the error of the second kind under condition that the error of the first kind is zero. Solution of this problem does not exist always but the GMSE-approach [2] can be used always according to which the error of the second kind is minimized under condition that the error of the first kind reaches its minimum. The error of the second kind is in both cases the variance and we may speak of conditional minimum variance estimate (CMVE). We shall however consider a more general case when data are random variables. In this formulation not only the actual but also the required result of estimation is a random quantity and instead of the variance of the estimate, its mean square error is to be taken into account as the error of the second kind. Instead of CMVE we thus deal the more general case, the conditional minimum square error estimate (CMSE).

Confusion might take place relating to the actual sense of the error of the first kind. Zero value of this error is sufficient for the unbiasedness of the estimate but it is not necessary. Unbiased linear estimates not satisfying this requirement can be found minimizing the error of the second kind unconditionally. Such approach is based on an idea of Semyonov (1954). As mentioned in [3], it was an alternative and more general formulation of the filtering problem to that one given by Wiener and generalized by Zadeh and Ragazzini (1950). Therefore, the estimate minimizing the error of the second kind unconditionally, the unconditional minimum mean-square error estimate (UMSE) may be called also the Semyonov estimate. In the case of the UMSE, the signal is considered to be not only unknown but a random variable having a known mean and a given variance. Hence the error of the second kind is not identical with the variance of the estimate in this case. The mean square value of the estimating error of the UMSE can be less than that of the CMSE whereas opposite is true for the error of the first kind.

It has been shown in [4] that minimizing unconditionally a linear combination of mean square errors of both mentioned kinds (a quantity called the penalty) one obtains a generalized estimate, the minimum penalty estimate (MPE). The UMSE is obtained as an extremal special case of the MPE when taking zero weight of the error of the first kind, whereas the CMSE can be considered to be an opposite extremal case of unlimited weight of this error. Between these extremal cases one has a set of compromising estimates.

It is the purpose of this paper to extend the minimum penalty estimate concept to the case when data or signals are random matrices.

PRELIMINARIES

The model is given by $n \times p$ random matrices

$$Y = Y_x + Y_e\,, \tag{1}$$

where Y_x represents signals, the matrix Y_e noises and another random errors, the Y being composed of the observed data. Mean values and covariances of all components are supposed to be known.

The required result of estimation under condition of zero noise should be a $t \times p$

matrix

(2) $$Z_x = \mathscr{T}_x\{Y_x\}$$

and in the presence of noise a matrix

(3) $$Z_0 = \mathscr{T}_0\{Y_x, Y_e\}$$

of the same dimensions, where $\mathscr{T}_x$ and $\mathscr{T}_0$ are some given operators. The estimate will be of the general linear form

(4) $$Z_x = WY + C\,,$$

where W and C are some constant matrices having dimensions $t \times n$ and $t \times p$, respectively.

As a measure of a random matrix E we introduce a scalar quantity

(5) $$\|E\| = [\operatorname{tr}\{\langle EQE^T\rangle\}]^{1/2}\,,$$

the square root of the trace of the expected value of a quadratic form EQE^T for a given weighting matrix Q which is positive definite. It can be shown, that the ordinary conditions

(6) $$\|E\| = 0 \quad \text{for} \quad E = 0 \quad \text{and} \quad \|E\| > 0 \quad \text{for} \quad E \neq 0\,,$$

(7) $$\|E_1 + E_2\| \leqq \|E_1\| + \|E_2\|\,,$$

and

(8) $$\|aE\| = |a|\,\|E\|$$

for a real scalar a are fulfiled.

The error of the first kind is defined as

(9) $$\|E_X\| = \|WY_X + C - Z_X\|\,,$$

the error of the second kind being

(10) $$\|E_0\| = \|W(Y_x + Y_e) + C - Z_0\|\,.$$

The penalty

(11) $$p = p_0\|E_0\|^2 + p_x\|E_x\|^2$$

can be defined using the weights p_0 and p_x,

(12) $$p_0 > 0\,,$$

(13) $$p_0 + p_x > 0\,.$$

Formulation of the problem

Given data of the form (1) and the matrices (2) and (3) characterizing a function of the data. Values of this function are to be estimated. Given first and second statistical moments of all elements of the matrices (1), (2) and (3). Determine matrices W and C for which the linear estimate (4) minimizes unconditionally the penalty (11).

Centralization of variables

Substituting (9) and (10) into (11) one obtains after formal transformations the penalty

$$p = \mathrm{tr}\,\{p_0\langle \hat{Z}_0 Q \hat{Z}_0^{\mathrm{T}}\rangle + p_x\langle \hat{Z}_x Q \hat{Z}_x^{\mathrm{T}}\rangle + W(p_0\langle \hat{Y} Q \hat{Y}^{\mathrm{T}}\rangle + p_x\langle \hat{Y}_x Q \hat{Y}_x^{\mathrm{T}}\rangle)\, W^{\mathrm{T}} - (p_0\langle \hat{Z}_0 Q \hat{Y}^{\mathrm{T}}\rangle + p_x\langle \hat{Z}_x Q \hat{Y}_x^{\mathrm{T}}\rangle)\, W^{\mathrm{T}} - W(p_0\langle \hat{Y} Q \hat{Z}_0^{\mathrm{T}}\rangle + p_x\langle \hat{Y}_x Q \hat{Z}_x^{\mathrm{T}}\rangle) + (p_0 + p_x)(C - C_b)\, Q(C^{\mathrm{T}} - C_b^{\mathrm{T}})\}\,, \tag{14}$$

where

$$\hat{Z}_0 = Z_0 - (p_0\langle Z_0\rangle + p_x\langle Z_x\rangle)/(p_0 + p_x)\,, \tag{15}$$

$$\hat{Z}_x = Z_x - (p_0\langle Z_0\rangle + p_x\langle Z_x\rangle)/(p_0 + p_x)\,, \tag{16}$$

$$\hat{Y}_x = Y - (p_0\langle Y\rangle + p_x\langle Y_x\rangle)/(p_0 + p_x)\,, \tag{17}$$

$$\hat{Y}_x = Y_x - (p_0\langle Y\rangle + p_x\langle Y_x\rangle)/(p_0 + p_x)\,, \tag{18}$$

$$C_b = (p_0\langle Z_0\rangle + p_x\langle Z_x\rangle - W(p_0\langle Y\rangle + p_x\langle Y_x\rangle))/(p_0 + p_x)\,. \tag{19}$$

In (14), only the last term depends on the constant matrix C. This term is a positive semidefinite matrix, its trace is a non-negative number. The optimum choice of the constant term of the estimator (4) for an arbitrary W is thus

$$C = C_b\,. \tag{20}$$

It can be mentioned that using substitutions (17) and (19), one gets the estimating formula (4) in a simplified form

$$\hat{Z} = W\hat{Y}\,, \tag{21}$$

where again

$$\hat{Z} = Z - (p_0\langle Z_0\rangle + p_x\langle Z_x\rangle)/(p_0 + p_x)\,. \tag{22}$$

A case characterized by the relations

$$Z_0 \equiv Z_x \tag{23}$$

and

$$\langle Y_e \rangle = 0 \tag{24}$$

is met usually in practice. Identity (23) expresses the requirement to minimize the influence of the noise upon the results of estimation. The condition (24) does not represent a loss of generality. The eventual non-zero mean of the noise may be included into the data matrix Y_x. In the case of validity of the conditions (23) and (24) the transformation leading to the estimator (21) is simply the centralization of variables.

Geometrical aspects

Let us denote

$$M = p_0\langle \hat{Y} Q \hat{Y}^T \rangle + p_0\langle \hat{Y}_x Q \hat{Y}_x^T \rangle \tag{25}$$

and

$$B = \langle \hat{Y}_e Q \hat{Y}_e^T \rangle + \langle \hat{Y}_e Q \hat{Y}_x^T \rangle + \langle \hat{Y}_x Q Y_e^T \rangle . \tag{26}$$

Both members of the right-hand side of (25) are positive semidefinite, they are summs of covariance matrices. One can therefore obtain a decomposition

$$\langle \hat{Y}_x Q \hat{Y}_x^T \rangle = X X^T \tag{27}$$

which is not unique, one has certain freedom in choosing a $n \times q$ matrix X having the same rank m as the matrix $\langle \hat{Y}_x Q \hat{Y}_x^T \rangle$, whereby

$$m \leqq n . \tag{28}$$

The rank s of the positive semidefinite matrix B is also not necessarily full,

$$s \leqq n . \tag{29}$$

The column-space of the matrices X and B will be denoted $\mathscr{S}_x$ and $\mathscr{S}_e$, respectively. A natural assumption is

$$\mathscr{S}_x \subseteq \mathscr{S}_e \tag{30}$$

representing no loss of generality. If some signals would be outside of the "error space" then such signals can be determined exactly and it is not necessary to include them into the model being considered. Denoting $\mathscr{S}_n$ the n-dimensional space, we conclude from (29)

$$\mathscr{S}_e \subseteq \mathscr{S}_n . \tag{31}$$

The case of no data vectors existing within a part of the space $\mathscr{S}_n$ is thus taken into account, too.

To handle the rectangular matrices as well as possibly singular quadratic matrices we shall use the most general version of a generalized inverse matrice [5]: The one-condition generalized inverse A^{g1} of an arbitrary matrix A is a matrix satisfying the condition

$$AA^{g1}A = A\,. \tag{32}$$

The inverse A^{g1} is not unique, in determining this matrix there remains a certain choice making it possible to add further conditions when the matrix A^{-1} does not exist.

It follows from (30) that the columns-space of the matrix M (25) is in a general case not identical with $\mathscr{S}_n$. Each data matrix may be represented as

$$Y = MA\,. \tag{33}$$

The equation

$$(I - MM^{g1})\,Y = 0 \tag{34}$$

resulting from (32) and (33) can be interpreted as "there exist no data outside the subspace $\mathscr{S}_e$ of the n-dimensional space $\mathscr{S}_n$". (The symbol I denotes the unity matrix.)

Taking into account (30) we may apply the equation (34) also to the signal matrix Y_x.

The column-space of the matrix X is identical with the row-spaces of both matrices $\langle \hat{Z}_0 Q \hat{Y}_x^T \rangle$ and $\langle \hat{Z}_x Q \hat{Y}_x^T \rangle$. One has therefore

$$\langle \hat{Z}_x Q \hat{Y}_x^T \rangle = \langle \hat{Z}_x Q \hat{Y}_x^T \rangle \,(X^T)^{g1}\, X^T \tag{35}$$

as well as

$$\langle \hat{Z}_0 Q \hat{Y}_x^T \rangle = \langle \hat{Z}_0 Q \hat{Y}_x^T \rangle \,(X^T)^{g1}\, X^T\,. \tag{36}$$

A GENERAL CASE OF THE BEST LINEAR ESTIMATOR

The expression (14) for the penalty can be rewritten substituting $C = C_b$ and using (32):

$$\begin{aligned} p = \operatorname{tr} \{ & p_0 \langle \hat{Z}_0 Q \hat{Z}_0^T \rangle + p_x \langle \hat{Z}_x Q \hat{Z}_x^T \rangle - (p_0 \langle \hat{Z}_0 Q \hat{Y}^T \rangle + \\ & + p_x \langle \hat{Z}_x Q \hat{Y}_x^T \rangle)\, M^{g1} (p_0 \langle \hat{Y} Q \hat{Z}_0^T \rangle + p_x \langle \hat{Y}_x Q \hat{Z}_x^T \rangle) + \\ & + [WM - (p_0 \langle \hat{Z}_0 Q \hat{Y}^T \rangle + \\ & + p_x \langle \hat{Z}_x Q \hat{Y}_x^T \rangle)]\, M^{g1} [M^T W^T - (p_0 \langle \hat{Y} Q \hat{Z}_0^T \rangle + \\ & + p_x \langle \hat{Y}_x Q \hat{Z}_x^T \rangle)]\}\,. \end{aligned} \tag{37}$$

The penalty is thus minimized for

$$W_b M = p_0 \langle \hat{Z}_0 Q \hat{Y}^T \rangle + p_x \langle \hat{Z}_x Q \hat{Y}_x^T \rangle\,. \tag{38}$$

Multiplying (38) by M^{g1} we obtain $W_b M M^{g1}$. The component $W_b(I - MM^{g1})$ may be chosen freely, as this part of the estimator will not add anything to the result of estimation because of (34). A simple choice is

$$W_b(I - MM^{g1}) = 0 . \tag{39}$$

We have therefore

$$W_b = W_b M M^{g1} = (p_0 \langle \hat{Z}_0 Q \hat{Y}^T \rangle + p_x \langle \hat{Z}_x Q \hat{Y}_x^T \rangle) M^{g1} . \tag{40}$$

This result can be presented in a more explicit form. It is convenient to start with the case of a matrix X_0 having full rank m. The result will be generalized afterwards. We have seen already that a matrix X satisfying (27) is contained in the column-space of the symmetrical matrix B (26). Therefore, a matrix lemma [5] may be applied to determine a generalized inversion of the summ of two matrices:

$$M^{g1} = (1/p_0) B^{g1}[I - X_0(p_0/(p_0 + p_x) I + X_0^T B^{g1} X_0)^{-1} X_0^T B^{g1}] . \tag{41}$$

Using (1), (35), (36), (40) and (41) we come to the result

$$\begin{aligned} W_b = {} & 1/(1 + r) (\langle \hat{Z}_0 Q \hat{Y}_x^T \rangle + \\ & + r \langle \hat{Z}_x Q \hat{Y}_x^T \rangle) (X_0^T)^{g1} (1/(1 + r) I + X_0^T B^{g1} X_0)^{-1} X_0^T B^{g1} + \\ & + \langle \hat{Z}_0 Q \hat{Y}_x^T \rangle B^{g1}[I - X_0(1/(1 + r) I + X_0^T B^{g1} X_0)^{-1} X_0^T B^{g1}] , \end{aligned} \tag{42}$$

where

$$r = p_x / p_0 \tag{43}$$

is the relative weight of the penalty components.

The matrix $(1/(1 + r) I + X_0^T B^{g1} X_0)^{-1}$ exists always because of full rank of the matrix X_0, positive semidefiniteness of the nonzero matrix B and because of non-negativeness of the parameter $1 + r$. Nonsingularity of the matrix $X_0^T B^{g1} X_0$ will be shown below.

Now it is possible to proceed to the case of the $n \times q$ matrix X with no assumptions on its rank. Such matrix may always be represented as

$$X = X_0 F \tag{44}$$

using a full-rank matrix X_0 and an $m \times q$ matrix F satisfying the semiorthonormal condition

$$F F^T = I_{m \times m} . \tag{45}$$

It can be verified by the substitution into (32) that

$$X^{g1} = F^T X_0^{g1} \tag{46}$$

and

$$(47)\qquad F^{T}(1/(1+r)I + X_0^{T}B^{g1}X_0)^{-1} F = [1/(1+r)X^{g1}X + X^{T}B^{g1}X]^{g1},$$

where the substitution

$$(48)\qquad F^{T}F = X^{g1}X$$

is admissible because of the full rank of the matrix X_0.

One has therefore the best linear estimator in the form

$$(49)\qquad W_b = (1+r)^{-1}(\langle \hat{Z}_0 Q \hat{Y}_x^{T}\rangle + r\langle \hat{Z}_x Q \hat{Y}_x^{T}\rangle)(X^{T})^{g1}((1+r)^{-1}X^{g1}X + X^{T}B^{g1}X)^{g1}X^{T}B^{g1} + \langle \hat{Z}_0 Q \hat{Y}_e^{T}\rangle B^{g1}[I - X((1+r)^{-1}X^{g1}X + X^{T}B^{g1}X)^{g1}X^{T}B^{g1}]$$

applicable without any restrictions of generality.

ERRORS

Error of the first kind

To simplify writing of formulae one may make use of the possibility to introduce additional conditions defining the generalized inverse more specifically. We need here the decomposition of symmetrical positive-semidefinite matrices

$$(50)\qquad B = S_e D_e^2 S_e^{T}$$

and

$$(51)\qquad \langle \hat{Y}_x Q \hat{Y}_x^{T}\rangle = S_x D_x^2 S_x^{T},$$

where the elements of diagonal matrices D_e^2 and D_x^2 are positive latent roots of the matrices. Matrices S_e and S_x having dimensions $n \times s$ and $n \times m$ are column-orthonormal,

$$(52)\qquad S_e^{T}S_e = I_{s\times s},$$

$$(53)\qquad S_x^{T}S_x = I_{m\times m}.$$

Clearly

$$(54)\qquad X_0 = S_x D_x$$

and by (44) also

$$(55)\qquad X = S_x D_x F$$

are allowed.

The Moore-Penrose generalized inverse [5] will be used below for matrices B and X, defined as

(56) $$B^g = S_e D_e^{-2} S_e^T$$

and

(57) $$X^g = F^T D_x^{-1} S_x^T$$

whereby

(58) $$D_x^{-1} S_x^T = (X_0^g)^T = X_0^{gT} .$$

These inverses are unique. They satisfy (32) as well as three additional conditions [5].

Denoting

(59) $$V_0 = \langle \hat{Z}_0 Q \hat{Y}_x^T \rangle X^{gT} ,$$

(60) $$V_x = \langle \hat{Z}_x Q \hat{Y}_x^T \rangle X^{gT} ,$$

(61) $$V_e = \langle \hat{Z}_0 Q \hat{Y}_e^T \rangle S_e D_e^{-1} ,$$

(62) $$G = D_e^{-1} S_e^T X ,$$

(63) $$H = [(1 + r)^{-1} X^g X + X^T B^g X]^g ,$$

(64) $$P = (1 + r)^{-1} (V_0 + r V_x) - V_e G$$

one obtains from (49)

(65) $$W_b = (PHG^T + V_e) D_e^{-1} S_e^T$$

which substituted into (9) written as

(66) $$\|E_x\| = \|W_b \hat{Y}_x - \hat{Z}_x\|$$

gives the error of the first kind in the form

(67) $$\|E_x\| = [\mathrm{tr}\, \{\langle \hat{Z}_x Q \hat{Z}_x^T \rangle - V_x V_x^T + [PHG^T G - (V_x - V_e G)] [PHG^T G - (V_x - V_e G)]^T\}]^{1/2} .$$

Error of the second kind and the penalty

This error is defined by (10) and (15)–(19) as

(68) $$\|E_0\| = \|W_b \hat{Y} - \hat{Z}_0\| .$$

 Using the notation introduced above one yields

$$\|E_0\|^2 = \mathrm{tr}\,\{\langle \hat{Z}_0 Q \hat{Z}_0^T \rangle - V_0 V_0^T - V_e V_e^T + + (V_0 - V_e G)(I + G^T G)^{-1} (V_0 - V_e G)^T + + [PH - (V_0 - V_e G)(I + G^T G)^{-1}]\, G^T (I + G G^T)\, G [PH - - (V_0 - V_e G)(I + G^T G)^{-1}]^T\} . \tag{69}$$

Substitution of (67) and (69) into (11) gives the penalty resulting from the use of the best estimator (42):

Extremal values of the errors

$$p = p_0\, \mathrm{tr}\,\{\langle \hat{Z}_0 Q \hat{Z}_0^T \rangle + r \langle \hat{Z}_x Q \hat{Z}_x^T \rangle - - [(V_0 + r V_x)\, G^T + V_e]\, [I - G H G^T]\, [(V_0 + r V_x)\, G^T + V_e]^T\} . \tag{70}$$

It follows from (67) that the minimum of the error $\|E_x\|$ is reached for

$$[PH]_1\, G^T G = V_x - V_e G . \tag{71}$$

In this equation, the term $[PH]_1$ depends on the parameter r. To solve this equation for the value r_1, minimizing the error $\|E_x\|$, we use the relations given above and take into account that

$$S_e S_e^T S_x = S_x \tag{72}$$

as follows from (31). Therefore

$$S_x^T S_e (S_x^T S_e)^T = I_{m \times m} , \tag{73}$$

the matrix $S_x^T S_e$ is a full-rank matrix as well as the matrix

$$G_0 = D_e^{-1} S_e^T S_x D_x . \tag{74}$$

The matrix $G_0^T G_0$ is thus non-singular and the minimality condition is

$$(1 + r)^{-1}\, [V_0 - V_x (I + (G_0^T G_0)^{-1}) + V_e G_0 (G_0^T G_0)^{-1}] = 0 . \tag{75}$$

Excluding the case of zero matrix value of the bracketed term we may conclude that the single minimum of the error takes place when

$$r = r_1 \to \infty . \tag{76}$$

The minimum value of the error $\|E_x\|$ is thus reached for

$$\|E_x\|^2_{\min} = \mathrm{tr}\,\{\langle \hat{Z}_x Q \hat{Z}_x^T \rangle - \langle \hat{Z}_x Q \hat{Y}_x^T \rangle\, X^{gT} X^g \langle \hat{Y}_x Q \hat{Z}_x^T \rangle\} \tag{77}$$

being non-zero in a general case.

As to the error of the second kind $\|E_0\|$, its minimum value

(78)
$$\|E_0\|_{\min} = [\operatorname{tr}\{\langle \hat{Z}_0 Q \hat{Z}_0^T \rangle - \langle \hat{Z}_0 Q \hat{Y}_x^T \rangle X^{gT} X^g \langle \hat{Y}_x Q \hat{Z}_0^T \rangle - \langle \hat{Z}_0 Q \hat{Y}_e^T \rangle B^g \langle \hat{Y}_e Q \hat{Z}_0^T \rangle + (\langle \hat{Z}_0 Q \hat{Y}_x^T \rangle X^{gT} - \langle \hat{Z}_0 Q \hat{Y}_e^T \rangle B^g X)(I + X^T B^g X)^{-1} (X^g \langle \hat{Y}_x Q \hat{Z}_0^T \rangle - X^T B^g \langle \hat{Y}_e Q \hat{Z}_0^T \rangle)\}]^{1/2}$$

is obtained from the definition

(79)
$$\|E_0\| = \|W_b \hat{Y} - Z_0\|$$

written as

(80)
$$\|E_0\| = [\operatorname{tr}\{\langle \hat{Z}_0 Q \hat{Z}_0^T \rangle - V_0 V_0^T - V_e V_e^T + (V_0 - V_e G)(I + G^T G)^{-1} (V_0 - V_e G)^T + (PH - (V_0 - V_e G)(I + G^T G)^{-1}) G^T (I + G G^T) G (PH - (V_0 - V_e G)(I + G^T G)^{-1})^T\}]^{1/2},$$

the minimum being reached for

(81)
$$[PH]_2 G = (V_0 - V_e G)(I + G^T G)^{-1} G .$$

Solution of this equation for the value r_2 of the parameter r gives the condition of minimality of the error $\|E_0\|$ in the following form:

(82)
$$r_2[V_x - (V_0 - V_e G)(F F^T + G^T G)^g G^T G] = 0 .$$

If the bracketed matrix term is not zero then the unique solution is

(83)
$$r_2 = 0 .$$

Hence, it may be summarized that the estimating errors of both kinds reach their minimum values at two opposite ends of the range of the parameter r.

SPECIAL CASES: UNCONDITIONAL AND CONDITIONAL MINIMUM SQUARE ERROR ESTIMATES

It is apparent that the latter of both extremal cases considered above (the case $r = 0$) represents a generalization of a discrete version of the Semyonov estimate (UMSE-unconditional minimum mean-square error estimate). The error $\|E_0\|$ is identical with the penalty in this case and it is minimized unconditionally via minimization of the penalty.

It has to be shown that the former of the extremal cases (the case $r \to \infty$) is the case of generalized conditional minimum square error estimate (CMSE). Let us consider the problem of this estimate independently:

Denoting

$$WS_e D_e = V \tag{84}$$

one can write the error $\|E_x\|$ in the form

$$\|E_x\| = [\mathrm{tr}\,\{\langle \hat{Z}_x Q \hat{Z}_x^T \rangle - V_x V_x^T + (VG_0 - V_x)(VG_0 - V_x)^T\}]^{1/2} \tag{85}$$

minimized unconditionally by

$$VG_0 = V_x \,. \tag{86}$$

This equation has always a solution as V_x is in the row-space of G_0 (see definitions (60) and (74).

We prefer here to use the full-rank matrix G_0 (74) because of simpler calculations and to generalize the formulae for a more general G (62) afterwards. Also the matrices V_x and V_e are taken here for a full-rank X.

General solution of the equation (86) for the unknown matrix V is [5]

$$V = V_x G_0^g + L[I - G_0 G_0^g] \,, \tag{87}$$

where

$$G_0^g = (G_0^T G_0)^{-1} G_0^T \tag{88}$$

and the matrix L is an arbitrary matrix of proper dimensions. Each V satisfying (86) must be thus of the form (87) and only the matrix L may be subject to any additional conditions. Using estimator (87) one obtains the error of the second kind

$$\begin{aligned} \|E_0\| = [\mathrm{tr}\,\{\langle \hat{Z}_0 Q \hat{Z}_0^T \rangle - V_0 V_0^T - V_e V_e^T + \\ + (V_x G_0^g - V_e)(V_x G_0^g - V_e)^T + (V_x - V_0)(V_x - V_0)^T + \\ + (L - V_e)(I - GG^g)(L - V_e)^T\}]^{1/2} \end{aligned} \tag{89}$$

minimized by

$$L = V_e \,. \tag{90}$$

But this substitution makes (87) identical with the special case of (49) for $r \to \infty$:

$$V = V_x G_0^g + V_e[I - G_0 G_0^g] \,. \tag{91}$$

It may be concluded that the minimum penalty estimate in the case $r \to \infty$ leads to minimization of the error of the second kind $\|E_0\|$ constrained by the requirement to minimize the error of the first kind $\|E_x\|$ unconditionally.

Estimability

We have seen that the minimum penalty estimate always exists. In this sense the problem of estimability is dropped. But the ordinary concept of estimability relates to the existence of such CMSE for which

(92) $$E_x = W\hat{Y}_x - \hat{Z}_x \equiv 0 .$$

It is clear from (6), that it means also that the error of the first kind $\|E_x\|$ is zero.

For linear operators the required result $\hat{Z}_x$ of the estimation in the case of data represented as

(93) $$\hat{Y}_x = XA ,$$

where the matrix A is a random matrix for which

(94) $$\langle AQA^T \rangle = I ,$$

should equal

(95) $$Z_x = P_a A$$

with a given operator P_a. The matrix X is non-random. As shown above, the estimator of the CMSE type is obtained from (49) by $r \to \infty$

(96) $$W_{b\infty} = \langle \hat{Z}_x Q \hat{Y}_x^T \rangle X^{gT}(X^T B^g X)^g X^T B^g + \\ + \langle \hat{Z}_0 Q \hat{Y}_e^T \rangle B^g[I - X(X^T B^g X)^g X^T B^g] .$$

It is easily seen that for the Moore-Penrose inversion

(97) $$X[(X^T B^g X)^g X^T B^g X] = X$$

and

(98) $$X^{gT}[(X^T B^g X)^g X^T B^g X] = X^{gT} .$$

The condition of the estimability thus follows by substitution of (92), (94), (96) and (97) into (91)

(99) $$P_a X^T X^{gT} = P_a .$$

One may conclude that every linear function of data is "minimum-penalty-estimable", but for that ones for which (99) holds, an estimate exists minimizing the error $\|E_0\|$ under constraint that the error matrix E_x as well as the error $\|E_x\|$ are zero.

Least squares estimates as special cases of the minimum penalty estimate

A generalization of the discrete Zadeh-Ragazzini estimate for the case of estimable linear functions and for vector data has been presented in [6]. This result can be obtained from (96) by replacing the matrix Q by the scalar 1 and by using (99).

In order to obtain from (96) the Lewis-Odell's generalization of the Gauss-Markov estimate applicable also in estimating non-estimable linear functions of data in the sense of CMSE, following simplifications have to be supposed:

1. There are no correlations of data with noise
2. Noise variance matrix B (26) is nonsingular
3. Data are vectors, $Q = 1$
4. $\langle \hat{Z}_0 Q \hat{Y}_e^T \rangle = 0$
5. $P_a = I$.

It has been shown in [1] that all the results obtained via the linear filter development are special cases of the results obtainable from the method of least squares viewpoint. In the terms of the present paper, the main result of [1] presenting the optimum linear estimates of a random process having a priori statistics, is a special case of (40) for $r = 0$, and it may be understood as the Semyonov estimate.

(Received December 3, 1973.)

REFERENCES

[1] P. Swerling: Modern state estimation methods from the viewpoint of the method of least squares. IEEE Trans. on automatic control *AC-16* (1971), 6, 707—719.

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

[3] I. D. Krutko: Statističeskaja dinamika impulsnych sistem (Statistical dynamics of impulse systems). Sovetskoe radio, Moskva 1963.

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

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

[6] P. Kovanic: Optimum digital operators. In: Information Processing 68. North-Holland Publishing Co., Amsterdam 1969, 249—255.

Ing. Pavel Kovanic, CSc.; Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 180 76 Praha 8. Czechoslovakia.