KYBERNETIKA — VOLUME 8 (1972), NUMBER 5

Minimum Penalty Estimate

PAVEL KOVANIC

A new generalized discrete linear estimate is introduced, called a ***minimum penalty estimate*** from which not only the well-known discrete versions of linear estimates such as the Zadeh-Ragazzini, Gauss-Markov and Semyonov estimates but also some new and more general estimators may be obtained. The usefulness of the generalized estimate consists especially in the possibility of using a priori information to lower the estimating error.

1. MAIN TYPES OF KNOWN LINEAR ESTIMATORS

1.1. Zadeh-Ragazzini estimator

In 1950 Zadeh and Ragazzini published their generalization of Wiener's theory of prediction. A few years later in papers by different authors this problem was formulated and solved for discrete variables (see e.g. [1]). A generalization [2] extended the field of applications. Numerous particular applications of unbiased minimum variance estimators such as the optimum interpolation, extrapolation, analysis, differentiation and integration as well as the generalized Neumann-David estimation of linear forms are specified in [2].

The data vectors are

$$\mathbf{y} = \mathbf{X}\mathbf{a} + \mathbf{x}_{\mathrm{I}} + \mathbf{x}_{\mathrm{II}} \tag{1}$$

and the required result of transformation is

$$z_0 = \mathbf{p}\mathbf{a} + \mathscr{L}_{\mathrm{I}}\{\mathbf{x}_{\mathrm{I}}\} + \mathscr{L}_{\mathrm{II}}\{\mathbf{x}_{\mathrm{II}}\}\,, \tag{2}$$

where the matrix $\mathbf{X}$ is an $n \times q$ nonrandom matrix with a rank which is not necessarily full, the vector $\mathbf{p}$ is a given nonrandom vector, $\mathbf{a}$ is an unknown nonrandom vector, $\mathbf{x}_{\mathrm{I}}$ is a random vector carrying information and $\mathbf{x}_{\mathrm{II}}$ represents a corrupting random component. Symbols $\mathscr{L}_{\mathrm{I}}$ and $\mathscr{L}_{\mathrm{II}}$ denote some linear operators.

The actual result of estimation is

$$z = \mathbf{w}\mathbf{y}\,, \tag{3}$$

where the $n \times 1$ vector $\mathbf{w}$ is the estimator. The generalized discrete Zadeh-Ragazzini estimator is the vector $\mathbf{w}$ minimizing the variance of the estimate z and satisfying the constraint of unbiasedness

$$\mathbf{w}\mathbf{X} = \mathbf{p}\,. \tag{4}$$

Such an estimator exists if and only if

$$\mathbf{p}(\mathbf{E} - \mathbf{X}^{+}\mathbf{X}) = \mathbf{0}\,. \tag{5}$$

The symbol $\mathbf{X}^{+}$ denotes the Penrose pseudo-inverse of the matrix $\mathbf{X}$.

1.2. Gauss-Markov estimate

The generalized Gauss-Markov estimator [3] published in 1966, may be considered to be a special case of the Zadeh-Ragazzini estimator. In this case, all components a_j ($j = 1, \ldots, m$) of the vector $\mathbf{a}$ (Eq. (1)) have to be estimated using vector estimators with different vectors $\mathbf{p}_j$ defined as

$$\begin{aligned} p_{ij} &= 1 \quad \text{for} \quad i = j\,, \\ &= 0 \quad \text{for} \quad i \neq j\,, \end{aligned} \tag{6}$$

where the number p_{ij} is the i-th component of the j-th vector $\mathbf{p}_j$.

Both operators, $\mathscr{L}_{\mathrm{I}}$ and $\mathscr{L}_{\mathrm{II}}$ in (2) are zero operators for this case. Some months later another generalization [4] was published making it possible to use the previous concept also for the case when the unbiased Gauss-Markov estimate does not exist. The unbiased constraint is replaced by the more general constraint that a quadratic error norm is minimized.

1.3. Semyonov estimate

For the continuous variable a linear estimation problem has been formulated and solved by Semyonov differing from the Zadeh-Ragazzini one. For discrete automatic control systems the analogy of this problem has been given in [5]. Data structure is as in (1) but the vector $\mathbf{a}$ is a random vector having a known expected value $\langle \mathbf{a} \rangle$ and a known covariance matrix. Instead of the conditional, an unconditional minimum of the variance of the estimate is sought. Systematic error of the Semyonov estimate is zeroed by addition of a proper term determined by the mean data vector which is supposed to be known.

2.1. Definitions

Among the data vectors $\mathbf{y}$, there exist vectors $\mathbf{y}_x$ representing a useful component containing information while the other represent noise, measuring errors and other undesirable disturbances.

A subspace $\mathscr{S}_1$ of an n-dimensional vector space $\mathscr{S}$ can be defined in the following way

$$\langle \mathbf{y}_x \mathbf{y}_x^\mathrm{T} \rangle = \mathbf{S}_1 \mathbf{D}_1 \mathbf{S}_1^\mathrm{T}, \tag{7}$$

where the $n \times n$ matrix $\langle \mathbf{y}_x \mathbf{y}_x^\mathrm{T} \rangle$ is the mathematical expectation of the random matrix $\mathbf{y}_x \mathbf{y}_x^\mathrm{T}$. The columns of the $n \times m$ matrix $\mathbf{S}_1$, satisfying the orthonormal condition

$$\mathbf{S}_1^\mathrm{T} \mathbf{S}_1 = \mathbf{E}_{m \times m} \tag{8}$$

and having the same rank m as the matrix $\langle \mathbf{y}_x \mathbf{y}_x^\mathrm{T} \rangle$ form an orthonormal base of the subspace $\mathscr{S}_1$. The $m \times m$ diagonal matrix $\mathbf{D}_1$ has positive diagonal terms d_i,

$$d_1 \geqq d_2 \geqq \ldots \geqq d_m > 0\,. \tag{9}$$

Such decomposition is a special (symmetrical) case of the singular value decomposition [6] which exists for an arbitrary real nonzero matrix. It is unique.

Disturbing components of data vectors belonging to the same subspace $\mathbf{S}_1$ will be denoted y_1 and the factorization

$$\langle (\mathbf{y}_x + \mathbf{y}_1)(\mathbf{y}_x + \mathbf{y}_1)^\mathrm{T} \rangle = \mathbf{S}_1(\mathbf{M}_1 + \mathbf{D}_1)\, \mathbf{S}_1^\mathrm{T} \tag{10}$$

with a symmetrical matrix $\mathbf{M}_1$ will be used.

Considering the set of all possible data vectors $\mathbf{y}$ and its characteristics $\langle \mathbf{y}\mathbf{y}^\mathrm{T} \rangle$ having a rank s we may find that only a subspace $\mathscr{S}$ of the n-dimensional vector space $\mathscr{N}$ contain some data vectors,

$$s \leqq n\,. \tag{11}$$

For the subspace $\mathscr{S}_1$ a complementary subspace $\mathscr{S}_2$ can be defined in $\mathscr{S}$. Data vector component belonging to the subspace $\mathscr{S}_2$ will be denoted $\mathbf{y}_2$,

$$\mathbf{y}_2 \in \mathscr{S}_2\,. \tag{12}$$

Such a vector is orthogonal to the both components $\mathbf{y}_1$ and $\mathbf{y}_x$. A data vector of general type is therefore represented as

$$\mathbf{y} = \mathbf{y}_x + \mathbf{y}_1 + \mathbf{y}_2\,. \tag{13}$$

(14) $$\langle \mathbf{y}_2 \mathbf{y}_2^{\mathrm{T}} \rangle = \mathbf{S}_2 \mathbf{D}_2 \mathbf{S}_2^{\mathrm{T}}$$

defines an $(s - m) \times (s - m)$ diagonal matrix $\mathbf{D}_2$ and a semiorthonormal $n \times$ $\times (s - m)$ matrix $\mathbf{S}_2$ analogously with (7)–(9).

It follows from the definitions that

(15) $$\mathbf{S}_1^{\mathrm{T}} \mathbf{S}_2 = \mathbf{0}_{m \times (s-m)}$$

and

(16) $$\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \rangle = \mathbf{S}\mathbf{M}\mathbf{S}^{\mathrm{T}},$$

using the block matrix notation

(17) $$\mathbf{S}_{n \times s} = \left\| \mathbf{S}_1 \mid \mathbf{S}_2 \right\|$$

and

(18) $$\mathbf{M} = \left\| \begin{array}{c|c} \mathbf{M}_1 + \mathbf{D}_1 & \mathbf{M}_0 \\ \mathbf{M}_0^{\mathrm{T}} & \mathbf{D}_2 \end{array} \right\|,$$

where

(19) $$m \leqq s \leqq n,$$

$\mathbf{M}_0$ is an $m \times (s - m)$ matrix and the rank of the $s \times s$ matrix $\mathbf{M}$ is full.

Required results of estimation are

(20) $$z_0 = \mathscr{T}_0\{\mathbf{y}\}$$

and

(21) $$z_x = \mathscr{T}_x\{\mathbf{y}_x\},$$

where $\mathscr{T}_0$ and $\mathscr{T}_x$ are some given operators. For the estimator of the type (3), the mean square of errors of estimation are

(22) $$\langle e_0^2 \rangle = \langle (z_0 - \mathbf{w}\mathbf{y})^2 \rangle = \mathbf{w}\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \rangle \mathbf{w}^{\mathrm{T}} - \mathbf{w}\langle \mathbf{y} z_0 \rangle - \langle z_0 \mathbf{y}^{\mathrm{T}} \rangle \mathbf{w}^{\mathrm{T}} + \langle z_0^2 \rangle$$

and

(23) $$\langle e_x^2 \rangle = \langle (z_x - \mathbf{w}\mathbf{y}_x)^2 \rangle = \mathbf{w}\langle \mathbf{y}_x \mathbf{y}_x^{\mathrm{T}} \rangle \mathbf{w}^{\mathrm{T}} - \mathbf{w}\langle \mathbf{y}_x z_x \rangle - \langle z_x \mathbf{y}_x^{\mathrm{T}} \rangle \mathbf{w}^{\mathrm{T}} + \langle z_x^2 \rangle .$$

Both errors are functions of the sought estimator $\mathbf{w}$. By its variation one gets the gradients

(24) $$\mathbf{g}_0 = \mathbf{w}\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \rangle - \langle z_0 \mathbf{y}^{\mathrm{T}} \rangle$$

and

(25) $$\mathbf{g}_x = \mathbf{w}\langle \mathbf{y}_x \mathbf{y}_x^{\mathrm{T}} \rangle - \langle z_x \mathbf{y}_x^{\mathrm{T}} \rangle .$$

For a given gradient $\mathbf{g}_x$, the last equation can be considered as a constraint.

To evaluate the quality of the results of an estimating procedure we introduce some nonnegative weights c_0 and c_x for the individual errors $\langle e_0^2 \rangle$ and $\langle e_x^2 \rangle$, respectively. The quantity

$$c = c_0 \langle e_0^2 \rangle + c_x \langle e_x^2 \rangle \tag{26}$$

will be called the *penalty*. The ratio

$$r = \frac{c_x}{c_0} \tag{27}$$

characterizes the relative weight of both penalty components.

It can be reasonable to represent the data vector component $\mathbf{y}_x$ in the form

$$\mathbf{y}_x = \mathbf{X}\mathbf{a}\,, \tag{28}$$

where $\mathbf{X}$ is an $n \times q$ nonrandom matrix having a rank m,

$$m \leqq n, q\,. \tag{29}$$

Both cases $n > q$ and $q \leqq n$ are allowed. The $q \times 1$ vector $\mathbf{a}$ is a random vector normalized so that

$$\langle \mathbf{a}\mathbf{a}^{\mathsf{T}} \rangle = \mathbf{E}_{q \times q}\,. \tag{30}$$

This assumption represents no less of generality, such a normalization may be achieved choosing properly the matrix $\mathbf{X}$ for which the singular value decomposition is

$$\mathbf{X} = \mathbf{S}_1 \mathbf{D}_x \mathbf{Q}_1\,. \tag{31}$$

The $m \times q$ matrix $\mathbf{Q}_1$ also satisfies the semiorthonormal conditions

$$\mathbf{Q}_1 \mathbf{Q}_1^{\mathsf{T}} = \mathbf{E}_{m \times m} \tag{32}$$

and the $m \times m$ diagonal matrix $\mathbf{D}_x$ has positive elements for which we obtain from (7)

$$\mathbf{D}_1 = \mathbf{D}_x^2\,. \tag{33}$$

For an operator $\mathscr{T}_x$ (21) of the linear type we have

$$z_x = \mathbf{p}\mathbf{a} \tag{34}$$

with an $1 \times q$ nonrandom vector $\mathbf{p}$, and

$$\langle z_x \mathbf{y}_x^{\mathsf{T}} \rangle = \mathbf{p}\mathbf{X}^{\mathsf{T}}\,. \tag{35}$$

Using matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ we may write

$$\langle z_0 \mathbf{y}^{\mathsf{T}} \rangle = \mathbf{u}_1 \mathbf{S}_1^{\mathsf{T}} + \mathbf{u}_2 \mathbf{S}_2^{\mathsf{T}}\,. \tag{36}$$

2.2. Main results

In the Appendices I – IX the proofs of following theorems are given:

Theorem I. *Among all linear estimates having the form* (3), *the best one minimizing the penalty* (26) *is*

$$\tilde{z}_r = \mathbf{w}_r \mathbf{y}\,, \tag{37}$$

where the best unconditional estimator $\mathbf{w}_r$ *equals*

$$\mathbf{w}_r = (\langle z_0 \mathbf{y}^{\mathsf{T}} \rangle + r\langle z_x \mathbf{y}_x^{\mathsf{T}} \rangle)(\langle \mathbf{y}\mathbf{y}^{\mathsf{T}} \rangle + r\langle \mathbf{y}_x \mathbf{y}_x^{\mathsf{T}} \rangle)^{+}\,. \tag{38}$$

Theorem II. *Among all linear estimates having the form* (3), *the best one minimizing the penalty* (26) *and satisfying the constraint* (25) *is*

$$\tilde{z}_c = \mathbf{w}_c \mathbf{y}\,, \tag{39}$$

where the best conditional estimator $\mathbf{w}_c$ *equals to*

$$\mathbf{w}_c = (\langle z_x \mathbf{y}_x^{\mathsf{T}} \rangle + \mathbf{g}_x)(\langle \mathbf{y}_x \mathbf{y}_x^{\mathsf{T}} \rangle)^{+} + \\ + [\langle z_0 \mathbf{y}^{\mathsf{T}} \rangle - (\langle z_x \mathbf{y}_x^{\mathsf{T}} \rangle + \mathbf{g}_x)(\langle \mathbf{y}_x \mathbf{y}_x^{\mathsf{T}} \rangle)^{+} \langle \mathbf{y}\mathbf{y}^{\mathsf{T}} \rangle](\langle \mathbf{y}_2 \mathbf{y}_2^{\mathsf{T}} \rangle)^{+}\,. \tag{40}$$

Theorem III. *The best choice of the constraint* (25) *minimizing the error square* $\langle e_x^2 \rangle$ *is*

$$\mathbf{g}_x = \mathbf{w}_c \langle \mathbf{y}_x \mathbf{y}_x^{\mathsf{T}} \rangle - \langle z_x \mathbf{y}_x^{\mathsf{T}} \rangle = \mathbf{0} \tag{41}$$

or its equivalent

$$\mathbf{w}_c \mathbf{X} = \mathbf{p} \mathbf{X}^{+} \mathbf{X}\,. \tag{42}$$

Theorem IV. *The estimator having the form*

$$\mathbf{w}_r = \mathbf{v}_r \mathbf{K}_r \mathbf{S}^{\mathsf{T}}\,, \tag{43}$$

where

$$\mathbf{v}_r = \| \mathbf{p} \mathbf{Q}_1^{\mathsf{T}} \mathbf{D}_x (r \mathbf{M}_r^{-1}) + \mathbf{u}_1 \mathbf{M}_r^{-1} \mid \mathbf{u}_2 \|\,, \tag{44}$$

$$\mathbf{K}_r = \left\| \begin{array}{c|c} \mathbf{E} & -\mathbf{M}_0 \mathbf{D}_2^{-1} \\ -\mathbf{D}_2^{-1} \mathbf{M}_0^{\mathsf{T}} \mathbf{M}_r^{-1} & \mathbf{D}_2^{-1} + \mathbf{D}_2^{-1} \mathbf{M}_0^{\mathsf{T}} \mathbf{M}_r^{-1} \mathbf{M}_0 \mathbf{D}_2^{-1} \end{array} \right\|, \tag{45}$$

and

$$\mathbf{M}_r = \mathbf{M}_1 + (1 + r)\,\mathbf{D}_1 - \mathbf{M}_0 \mathbf{D}_2^{-1} \mathbf{M}_0^{\mathsf{T}}\,, \tag{46}$$

is identical with the best unconditional estimator (38) *and it includes the best conditional estimator* (40) *as a special case,*

(47) $$w_c = \lim_{r \to \infty} \mathbf{w}_r = \boldsymbol{p}\mathbf{Q}_1^{\mathrm{T}}\mathbf{D}_x^{-1}(\boldsymbol{S}_1^{\mathrm{T}} - \boldsymbol{M}_0\boldsymbol{D}_2^{-1}\boldsymbol{S}_2^{\mathrm{T}}) + \boldsymbol{u}_2\boldsymbol{D}_2^{-1}\boldsymbol{S}_2^{\mathrm{T}}.$$

Theorem V. *The mean error of transformation arising in application of the best unconditional estimator* (43) *to a vector of the type* $\mathbf{y}_x$ (28) *equals*

(48) $$\langle e_{xr} \rangle = \langle z_x - \mathbf{w}_r \boldsymbol{X}\boldsymbol{a} \rangle = \boldsymbol{p}_r \langle \boldsymbol{a} \rangle$$

and the mean square value of this error is

(49) $$\langle e_{xr}^2 \rangle = \boldsymbol{p}_r \boldsymbol{p}_r^{\mathrm{T}},$$

where

(50) $$\boldsymbol{p}_r = \boldsymbol{p}(\mathbf{E} - \mathbf{Q}_1^{\mathrm{T}}\boldsymbol{D}_x(r\boldsymbol{M}^{-1})\,\boldsymbol{D}_x\mathbf{Q}_1) - (\boldsymbol{u}_1 - \boldsymbol{u}_2\boldsymbol{D}_2^{-1}\boldsymbol{M}_0^{\mathrm{T}})\,\boldsymbol{M}_r^{-1}\boldsymbol{D}_x\mathbf{Q}_1,$$

the limit values for an increasing r (case (47)*) being*

(51) $$\lim_{r \to \infty} \langle e_{xr} \rangle = \boldsymbol{p}[\mathbf{E} - \mathbf{X}^{+}\mathbf{X}]\,\langle \boldsymbol{a} \rangle$$

and

(52) $$\lim_{r \to \infty} \langle e_{xr}^2 \rangle = \boldsymbol{p}[\mathbf{E} - \mathbf{X}^{+}\mathbf{X}]\,\boldsymbol{p}^{\mathrm{T}},$$

whereby the last value represents the minimum of the norm $\langle e_{xr}^2 \rangle$,

(53) $$\lim_{r \to \infty} \langle e_{xr}^2 \rangle \leqq \langle e_{xr}^2 \rangle$$

for all nonnegative r.

Theorem VI. *The mean square of estimating error arising in application of the best unconditional estimator* (43) *to a vector of the general type* $\mathbf{y} \in \mathscr{S}$ *equals*

(54) $$\langle e_0^2 \rangle_r = \langle (z_0 - \mathbf{w}_r \mathbf{y})^2 \rangle = \langle z_0^2 \rangle - \boldsymbol{v}_r\boldsymbol{K}_r\boldsymbol{u}^{\mathrm{T}} - \boldsymbol{u}\boldsymbol{K}_r^{\mathrm{T}}\boldsymbol{v}_r + \boldsymbol{v}_r\boldsymbol{K}_r\boldsymbol{M}\boldsymbol{K}_r^{\mathrm{T}}\boldsymbol{v}_r^{\mathrm{T}},$$

where

(55) $$\boldsymbol{u} = \|\boldsymbol{u}_1 \mid \boldsymbol{u}_2\|,$$

the special case for $r = 0$ *being given as*

(56) $$\langle e_0^2 \rangle_{r=0} = \langle z_0^2 \rangle - \boldsymbol{u}\boldsymbol{M}^{-1}\boldsymbol{u}^{\mathrm{T}},$$

whereby the last value represents the minimum,

(57) $$\langle e_0^2 \rangle_{r=0} \leqq \langle e_0^2 \rangle_r$$

for all nonnegative r.

Theorem VII. *The conventional form of the best conditional estimator* (40) *respecting the constraint* (41) *is*

$$\mathbf{w}_c = [p(T^{-1}\mathbf{S}^{\mathsf{T}}\mathbf{X})^+ + u(T^{\mathsf{T}})^{-1}(\mathbf{E} - (T^{-1}\mathbf{S}^{\mathsf{T}}\mathbf{X})(T^{-1}\mathbf{S}^{\mathsf{T}}\mathbf{X})^+)]\, T^{-1}\mathbf{S}^{\mathsf{T}} . \tag{58}$$

Theorem VIII. *Among all linear estimates, the best one minimizing the penalty*

$$\tilde{c}/c_0 = \langle (z_0 - \tilde{z}_0)^2 \rangle + r\langle (z_x - \tilde{z}_x)^2 \rangle \tag{59}$$

is

$$\tilde{z}_0 = \mathbf{w}_0\mathbf{y} + b \tag{60}$$

or – when applied to the vector $\mathbf{y}_x$ *–*

$$\tilde{z}_x = \mathbf{w}_0\mathbf{y}_x + b \tag{61}$$

with the constant term

$$b = [\langle z_0 \rangle + r\langle z_x \rangle - \mathbf{w}_0(\langle \mathbf{y} \rangle + r\langle \mathbf{y}_x \rangle)]/[1 + r] \tag{62}$$

and with the vector

$$\mathbf{w}_0 = (\langle \hat{z}_0 \hat{\mathbf{y}}^{\mathsf{T}} \rangle + r\langle \hat{z}_x \hat{\mathbf{y}}_x^{\mathsf{T}} \rangle)(\langle \hat{\mathbf{y}}\hat{\mathbf{y}}^{\mathsf{T}} \rangle + r\langle \hat{\mathbf{y}}_x \hat{\mathbf{y}}_x^{\mathsf{T}} \rangle)^+ , \tag{63}$$

where

$$\hat{\mathbf{y}} = \mathbf{y} - (\langle \mathbf{y} \rangle + r\langle \mathbf{y}_x \rangle)/(1 + r) , \tag{64}$$

$$\hat{\mathbf{y}}_x = \mathbf{y}_x - (\langle \mathbf{y} \rangle + r\langle \mathbf{y}_x \rangle)/(1 + r) ,$$

$$\hat{z}_0 = z_0 - (\langle z_0 \rangle + r\langle z_x \rangle)/(1 + r) , \tag{66}$$

$$\hat{z}_x = z_x - (\langle z_0 \rangle + r\langle z_x \rangle)/(1 + r) , \tag{67}$$

which may be rewritten in the form identical with (43) *if in all definitions instead of the variables* $\mathbf{y}$, $\mathbf{y}_x$, z_0 *and* z_x, *the variables* $\hat{\mathbf{y}}$, $\hat{\mathbf{y}}_x$, $\hat{z}_0$ *and* $\hat{z}_x$ *are substituted.*

The estimate corresponding to Theorem VIII will be called the *minimum penalty estimate.*

Theorem IX. *Mean errors of the minimum penalty estimates* (60) *and* (61) *are*

$$\langle z_0 - \tilde{z}_0 \rangle = [\mathbf{w}_0(\langle \mathbf{y}_x \rangle - \langle \mathbf{y} \rangle) + (\langle z_0 \rangle - \langle z_x \rangle)]\, r/(1 + r) \tag{68}$$

and

$$\langle z_x - \tilde{z}_x \rangle = [\mathbf{w}_0(\langle \mathbf{y} \rangle - \langle \mathbf{y}_x \rangle) + (\langle z_x \rangle - \langle z_0 \rangle)]/(1 + r) , \tag{69}$$

whereby the penalty (59) *is*

$$\tilde{c}/c_0 = \langle \hat{z}_0^2 \rangle + r\langle \hat{z}_x^2 \rangle - \mathbf{w}_0[\langle \hat{\mathbf{y}}\hat{z}_0 \rangle + r\langle \hat{\mathbf{y}}_x \hat{z}_x \rangle] . \tag{70}$$

Theorems I – VII are related to estimates having the form (3) without a constant term. The conditional estimate according to Theorems II and VII represents a generalization of Zadeh-Ragazzini's estimate (or minimum variance unbiased estimate) including the case of a rank deficiency and also the case when the condition (5) is not satisfied. In the last case the unbiased constraint (4) is replaced in accordance with Theorem III by a more general constraint (42) which results from the requirement to minimize a quadratic error norm $\langle e_x^2 \rangle$. The generalized Gauss-Markov estimate [4] may be therefore also obtained as a special case of the conditional estimate. However, the conditional estimate can be obtained according to Theorem IV as a limit case of the unconditional estimate introduced in Theorem I. Hence, the unconditional estimate is a most general one among estimates having the form (3). Although it is the best one minimizing the penalty, it is not unbiased in some cases, as shown by Theorem V. Comparing Theorem V and Theorem VI we may see, that for an infinite penalty ratio r, the error norm $\langle e_x^2 \rangle$ reaches its minimum, while for zero r a minimum of the norm $\langle e_0^2 \rangle$ is achieved. Thus, the choice of penalty ratio r determines the relative importance of both mentioned error norms. The requirement of unbiasedness or its generalized formulation (42) is equivalent to the statement that the norm $\langle e_x^2 \rangle$ is of prime importance. The penalty can be higher in this case than in the case of the unconditional estimate.

A general type of linear estimate should include a constant term as in (60) and (61). It is shown by Theorem VIII how to find the best estimator of this type. Its penalty can be lower than that of an unconditional estimate (37) for the same value of the ratio r. As follows from Theorem IX, the estimate $\tilde{z}_0$ (60) is asymptotically unbiased in the case $r = 0$, while the estimate $\tilde{z}_x$ (61) for $r \to \infty$. Moreover, both estimate (60) and (61) are asymptotically unbiased for all values of the penalty ratio r if following identities take place:

$$\langle \mathbf{y} \rangle = \langle \mathbf{y}_x \rangle , \tag{71}$$

$$\langle z_x \rangle = \langle z_0 \rangle . \tag{72}$$

As seen from (13), when the means $\langle \mathbf{y}_1 \rangle$ and $\langle \mathbf{y}_2 \rangle$ are zero or when they are included into the component $\mathbf{y}_x$, the identity (71) holds. In the case that the operators (20) and (21) are linear and

$$\mathscr{T}_0\{\mathbf{y}_1\} \equiv \mathscr{T}_0\{\mathbf{y}_2\} \equiv 0 , \tag{73}$$

we obtain from (13) also (72). Thus, if all parameters appearing in (62) – (67) are known, the best solution of the linear estimation problem is given by Theorem VIII.

A generalized discrete version of Semyonov estimate can be obtained from supposing zero penalty ratio $r = 0$. In this case the error norm $\langle e_x^2 \rangle$ is fully ignored.

The square error $\langle e_0^2 \rangle$ reaches the least possible value, but this advantage is paid by information on statistical distribution of data vectors components within the subspace $\mathscr{S}_1$, as characterized by the covariance matrice $\mathbf{M}_1$ needed for (43)–(46). As shown by (47), this matrice disappears in the opposite case $r \to \infty$, such information is then not necessary but the penalty is larger. We are referring to the formulae of Theorem IV as they can be used also for the general linear estimate after substitutions corresponding to Theorem VIII.

APPENDICES

A. I. Existence and uniqueness of the best unconditional estimator

It follows from the regularity of the symmetrical matrix $\mathbf{M}$ (18) that the matrix $\mathbf{M}_1 + \mathbf{D}_1$ is also regular. The matrix

$$\mathbf{M}_x = \mathbf{D}_x^{-1}(\mathbf{M}_1 + \mathbf{D}_1)\,\mathbf{D}_x^{-1} \tag{I.1}$$

is therefore also symmetrical and regular and its characteristic equation

$$\operatorname{Det}(\mathbf{M}_x - (-r)\,\mathbf{E}) = 0 \tag{I.2}$$

has only positive roots $(-r)$.

Thus, the matrix $\mathbf{M}_x + r\mathbf{E}$ as well as the matrix $\mathbf{M}_1 + (1 + r)\,\mathbf{D}_1$ is not singular for a positive value of the parameter r. Regularity for zero r follows from the definition.

Symmetrical matrix

$$\mathbf{M}_s = \left\| \begin{array}{c|c} \mathbf{M}_1 + (1 + r)\,\mathbf{D}_1 & \mathbf{M}_0 \\ \mathbf{M}_0^{\mathrm{T}} & \mathbf{D}_2 \end{array} \right\| \tag{I.3}$$

is regular because of regularity of matrices $\mathbf{M}_1 + (1 + r)\,\mathbf{D}_1$ and $\mathbf{D}_2$. One has therefore

$$(\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \rangle + r\langle \mathbf{y}_x \mathbf{y}_x^{\mathrm{T}} \rangle)^+ = (\mathbf{S}\mathbf{M}_s\mathbf{S}^{\mathrm{T}})^+ = \mathbf{S}\mathbf{M}_s^{-1}\mathbf{S}^{\mathrm{T}}, \tag{I.4}$$

where the $\mathbf{A}^+$ denotes the Penrose pseudo-inverse of the matrix $\mathbf{A}$.

We have supposed that no data vectors $\mathbf{y}$ exist outside the subspace $\mathscr{S}$,

$$\mathbf{y}\mathbf{S}\mathbf{S}^{\mathrm{T}} = \mathbf{y} \tag{I.5}$$

for an arbitrary data vector $\mathbf{y}$. If an estimator $\mathbf{w}$ had a component lying outside the subspace $\mathscr{S}$, the product of such a component with a data vector would be zero. Therefore, we are free to choose this component arbitrarily. We choose zero, assuming

$$\mathbf{w}\mathbf{S}\mathbf{S}^{\mathrm{T}} = \mathbf{w}\,. \tag{I.6}$$

Multiplying $\mathbf{w}_r$ (38) by $\mathbf{S}\mathbf{M}_s\mathbf{S}^T$ and taking into account (15), (24), (25), (27), (I.4) and (I.5) and orthonormal properties of the matrices $\mathbf{S}_1$ and $\mathbf{S}_2$, we obtain

(I.7) $$c_0\mathbf{g}_0 + c_x\mathbf{g}_x = \mathbf{0}_{1\times n}\,.$$

Thus, the estimator $\mathbf{w}_r$ (38) actually satisfies the necessary condition for the minimization of the penalty (26).

The solution of (I.7) for the unknown vector $\mathbf{w}_r\mathbf{S}$ is unique because of (I.4). The estimator $\mathbf{w}_r$ can be determined from the expression $\mathbf{w}_r\mathbf{S}$ also uniquely because of the assumption (I.6).

A. II. Existence and uniqueness of the best conditional estimator

Using notation

(II.1) $$\mathbf{f} = \langle z_x\mathbf{y}_x^T\rangle + \mathbf{g}_x\,,$$

(II.2) $$\mathbf{F} = \langle \mathbf{y}_x\mathbf{y}_x^T\rangle\,,$$

(II.3) $$\mathbf{G} = \langle \mathbf{y}_2\mathbf{y}_2^T\rangle\,,$$

(II.4) $$\mathbf{Y} = \langle \mathbf{y}\mathbf{y}^T\rangle\,,$$

(II.5) $$\mathbf{z} = \langle z_0\mathbf{y}^T\rangle + r\langle z_x\mathbf{y}_x^T\rangle\,,$$

one can write the constraint (25) as

(II.6) $$\mathbf{w}_c\mathbf{F} = \mathbf{f}\,.$$

Denoting the vector of Lagrangian multipliers $\mathbf{k}$, one gets the equation of minimization of the penalty (26) respecting the condition (II.6)

(II.7) $$\mathbf{w}_c(\mathbf{Y} + r\mathbf{F}) - \mathbf{z} - \mathbf{k}\mathbf{F} = \mathbf{0}_{1\times n}\,.$$

It follows from (7) and (II.2) that

(II.8) $$\mathbf{S}_1\mathbf{S}_1^T\mathbf{F} = \mathbf{F}$$

and

(II.9) $$\mathbf{f}\mathbf{F}^+\mathbf{F} = \mathbf{f}\,,$$

where the pseudo-inverse $\mathbf{F}^+$ is

(II.10) $$\mathbf{F}^+ = \mathbf{S}_1\mathbf{D}_1^{-1}\mathbf{S}_1^T\,.$$

Therefore, the projection of the estimator $\mathbf{w}_c$ on the subspace $\mathscr{S}_1$ can be obtained in the form

(II.11) $$\mathbf{w}_c\mathbf{S}_1\mathbf{S}_1^T = \mathbf{f}\mathbf{F}^+$$

satisfying (II.6).

Taking into account (I.6), (15) and (17) and substituting (II.11) into the projection of (II.7) on the subspace $\mathscr{S}_2$ one gets

(II.12) $$\mathbf{w}_c \mathbf{S}_2 \mathbf{S}_2^{\mathrm{T}} = \mathbf{z}\mathbf{G}^+ - \mathbf{f}\mathbf{F}^+\mathbf{Y}\mathbf{G}^+ ,$$

where

(II.13) $$\mathbf{G}^+ = \mathbf{S}_2 \mathbf{D}_2^{-1} \mathbf{S}_2^{\mathrm{T}} .$$

Sum of (II.11) and (II.12) together with (I.6) gives (40). The necessary condition for a conditional extremum is satisfied.

The penalty (26) corresponding to the estimator $\mathbf{w}_c$ (40) equals to

(II.14) $$\tilde{c}/c_0 = r\mathbf{f}\mathbf{F}^+\mathbf{f}^{\mathrm{T}} + \mathbf{f}\mathbf{F}^+[\mathbf{E} - \mathbf{Y}\mathbf{G}^+]\,\mathbf{Y}[\mathbf{E} - \mathbf{G}^+\mathbf{Y}]\,\mathbf{F}^+\mathbf{f}^{\mathrm{T}} - \mathbf{z}\mathbf{G}^+\mathbf{z}^{\mathrm{T}} - - \mathbf{f}\mathbf{F}^+[\mathbf{E} - \mathbf{Y}\mathbf{G}^+]\,\mathbf{z}^{\mathrm{T}} - \mathbf{z}[\mathbf{E} - \mathbf{G}^+\mathbf{Y}]\,\mathbf{F}^+\mathbf{f}^{\mathrm{T}} + \langle z_0^2 \rangle + r\langle z_x^2 \rangle .$$

Any other linear estimator $\hat{\mathbf{w}}$ (a vector of the same dimension as $\mathbf{w}_c$) satisfying the same constraint

(II.15) $$\hat{\mathbf{w}}\mathbf{F} = \mathbf{f}$$

would have the form

(II.16) $$\hat{\mathbf{w}} = \mathbf{f}\mathbf{F}^+ + \mathbf{q}\mathbf{G}^+$$

with a certain vector $\mathbf{q}$. Denoting $\hat{c}$ the penalty which corresponds to the estimator $\hat{\mathbf{w}}$ we may get from (II.14) and (II.15) using the identity

(II.17) $$\mathbf{G}^+\mathbf{Y}\mathbf{G}^+ = \mathbf{G}^+$$

and defining

(II.18) $$\mathbf{d} = [\mathbf{q} + \mathbf{f}\mathbf{F}^+\mathbf{Y} - \mathbf{z}]\,\mathbf{S}_u ,$$

where

(II.19) $$\mathbf{S}_u = \mathbf{S}_2 \mathbf{D}_2^{-1/2} ,$$

the relation

(II.20) $$\hat{c}/c_0 - \tilde{c}/c_0 = \mathbf{d}\mathbf{d}^{\mathrm{T}} \geqq 0 .$$

For the case

(II.21) $$\mathbf{q} = \mathbf{q}_0 = -\mathbf{f}\mathbf{F}^+\mathbf{Y} + \mathbf{z}$$

the difference $\hat{c} - \tilde{c}$ is zero, but this is the case when

(II.22) $$\mathbf{w} = \mathbf{f}\mathbf{F}^+ + [\mathbf{z} - \mathbf{f}\mathbf{F}^+\mathbf{Y}]\,\mathbf{C}^+ = \mathbf{w}_c$$

is identical with (40). Zero difference will occur also in the case

(II.23) $$\mathbf{q} = \mathbf{q}_0 + \mathbf{q}_1 \mathbf{S}_1^{\mathrm{T}} ,$$

where $\boldsymbol{q}_1$ is any $1 \times m$ vector. But the addition of the term $\boldsymbol{q}_1 \boldsymbol{S}_1^{\mathrm{T}}$ does not change the estimator $\hat{\boldsymbol{w}}$ (II.16),

$$\boldsymbol{S}_1^{\mathrm{T}} \boldsymbol{G}^+ = \boldsymbol{0} . \tag{II.24}$$

Therefore, there exists only one estimator satisfying the constraint and minimizing the penalty.

A. III. Choice of the constraint of the best conditional estimator

For an arbitrary estimator $\boldsymbol{w}$, the error square $\langle e_x^2 \rangle$ equals (23). For the estimator satisfying (41) this value is

$$\langle e_x^2 \rangle_0 = \langle z_x^2 \rangle - \langle z_x \boldsymbol{y}_x^{\mathrm{T}} \rangle \langle \boldsymbol{y}_x \boldsymbol{y}_x^{\mathrm{T}} \rangle^+ \langle \boldsymbol{y}_x z_x \rangle . \tag{III.1}$$

Using factorization

$$\langle \boldsymbol{y}_x \boldsymbol{y}_x^{\mathrm{T}} \rangle^+ = \boldsymbol{R} \boldsymbol{R}^{\mathrm{T}} \tag{III.2}$$

and a row vector

$$\boldsymbol{c} = [\boldsymbol{w} \langle \boldsymbol{y}_x \boldsymbol{y}_x^{\mathrm{T}} \rangle - \langle z_x \boldsymbol{y}_x^{\mathrm{T}} \rangle] \boldsymbol{R} , \tag{III.3}$$

we obtain

$$\langle e_x^2 \rangle - \langle e_x^2 \rangle_0 = \boldsymbol{c} \boldsymbol{c}^{\mathrm{T}} \geqq 0 . \tag{III.4}$$

Thus, the error $\langle e_x^2 \rangle$ reaches the minimum (III.1) for the estimator satisfying the constraint (42).

Substituting (7), (34) and (35) into (41) and (III.1) and taking into account (33) we obtain (42), where

$$\boldsymbol{X}^+ = \boldsymbol{Q}_1^{\mathrm{T}} \boldsymbol{D}_x^{-1} \boldsymbol{S}_1^{\mathrm{T}} . \tag{III.5}$$

A. IV. On the formula of minimum penalty estimator

The inverse of the matrix $\boldsymbol{M}_s$ (I.3) is

$$\boldsymbol{M}_s^{-1} = \left\| \begin{array}{c|c} \boldsymbol{M}_r^{-1} & -\boldsymbol{M}_r^{-1} \boldsymbol{M}_0 \boldsymbol{D}_2^{-1} \\ -\boldsymbol{D}_2^{-1} \boldsymbol{M}_0^{\mathrm{T}} \boldsymbol{M}_r^{-1} & \boldsymbol{D}_2^{-1} + \boldsymbol{D}_2^{-1} \boldsymbol{M}_0^{\mathrm{T}} \boldsymbol{M}_r^{-1} \boldsymbol{M}_0 \boldsymbol{D}_2^{-1} \end{array} \right\| , \tag{IV.1}$$

where the $m \times m$ matrix $\boldsymbol{M}_r$ is defined by (46). After substitution of (I.4), (IV.1), (31), (35) and (36) into (38) using notation (45) and (46) one obtains (43). As shown in A.I., the matrix $\boldsymbol{M}_s$ is regular for all positive values of the parameter r. This implies the nonsingularity of the matrix $\boldsymbol{M}_r$ for the same condition. Limits for an increasing r are

$$\lim_{r \to \infty} \boldsymbol{M}_r^{-1} = 0_{m \times m} \tag{IV.2}$$

and

(IV.3) $$\lim_{r\to\infty} (r\boldsymbol{M}_r^{-1}) = \boldsymbol{D}_1^{-1}\,,$$

(IV.4) $$\lim_{r\to\infty} \boldsymbol{v}_r = \|p\boldsymbol{Q}_1^{\mathrm{T}}\boldsymbol{D}_x^{-1} \mid \boldsymbol{u}_2\|\,,$$

and

(IV.5) $$\lim_{r\to\infty} \boldsymbol{K}_r = \left\| \begin{array}{c|c} \boldsymbol{E} & -\boldsymbol{M}_0\boldsymbol{D}_2^{-1} \\ \boldsymbol{0} & \boldsymbol{D}_2^{-1} \end{array} \right\|.$$

Direct substitution of (7), (14), (16), (31), (35), (36) and (41) into (40) results in the same as (43) with (IV.4) and (IV.5).

A. V. The error of transformation of a useful component of data

The errors (48) and (49) can be obtained from (43)–(46) using (15), (17), (31) and (34). Substituting (IV.2), (IV.3), (33) and (31) into (48) and (49) one gets (51) and (52). The difference of values (49) and (52) can be shown by substitution to have the form

(V.1) $$\langle e_{xr}^2\rangle - \lim_{r\to\infty} \langle e_{xr}^2\rangle = \boldsymbol{k}\boldsymbol{k}^{\mathrm{T}}\,,$$

where

(V.2) $$\boldsymbol{k} = p\boldsymbol{Q}_1^{\mathrm{T}}(\boldsymbol{E} - \boldsymbol{D}_x(r\boldsymbol{M}_r^{-1})\,\boldsymbol{D}_x) - (\boldsymbol{u}_1 - \boldsymbol{u}_2\boldsymbol{D}_2^{-1}\boldsymbol{M}_0^{\mathrm{T}})\,\boldsymbol{M}_r^{-1}\,,$$

from which (53) follows.

A. VI. The error of the estimation

The estimation error defined in (54) follows from (43) when using (15)–(18) and (36). For the limit case $r = 0$ we obtain from (44)–(46), (18) and (54)

(VI.1) $$(\boldsymbol{v}_r\boldsymbol{K}_r)_{r=0} = \boldsymbol{u}\boldsymbol{M}^{-1}$$

which substituted into (54) gives (56).

Denoting

(VI.2) $$\boldsymbol{h} = (\boldsymbol{v}_r\boldsymbol{K}_r - \boldsymbol{u}\boldsymbol{M}^{-1})\,\boldsymbol{T}$$

and

(VI.3) $$\boldsymbol{T}\boldsymbol{T}^{\mathrm{T}} = \boldsymbol{M}\,,$$

we get from (54)–(56)

(VI.4) $$\langle e_0^2\rangle_r - \langle e_0^2\rangle_{r=0} = \boldsymbol{h}\boldsymbol{h}^{\mathrm{T}} \geqq 0\,.$$

Substituting (55) into (47) one obtains

(VII.1) $$\mathbf{w}_c = (p\mathbf{Q}_1^{\mathrm{T}}\mathbf{D}_x^{-1}\mathbf{J} + u\mathbf{S}^{\mathrm{T}}\mathbf{S}_2\mathbf{D}_2^{-1}\mathbf{S}_2^{\mathrm{T}}\mathbf{S})\,\mathbf{S}^{\mathrm{T}}$$

where

(VII.2) $$\mathbf{J} = \|\mathbf{E}\mid -\mathbf{M}_0\mathbf{D}_2^{-1}\|\,.$$

It can be proved by substitution that

(VII.3) $$\mathbf{JT} = (\mathbf{T}^{-1}\mathbf{I})^{+}$$

and

(VII.4) $$\mathbf{S}^{\mathrm{T}}\mathbf{S}_2\mathbf{D}_2^{-1}\mathbf{S}_2^{\mathrm{T}}\mathbf{S} = (\mathbf{T}^{\mathrm{T}})^{-1}(\mathbf{E} - \mathbf{T}^{-1}\mathbf{IJT})\,\mathbf{T}^{-1}$$

where

(VII.5) $$\mathbf{I} = \mathbf{S}^{\mathrm{T}}\mathbf{S}_1 = \left\|\begin{matrix}\mathbf{E}_{m\times m}\\ \mathbf{0}_{(s-m)\times m}\end{matrix}\right\|$$

and where the matrix

(VII.6) $$\mathbf{TT}^{\mathrm{T}} = \mathbf{M}$$

is the same as in (18), the matrix $\mathbf{T}$ being upper triangular. It follows from (VII.3), (VII.5) and from the definition of Penrose pseudo-inverse that

(VII.7) $$\mathbf{Q}_1^{\mathrm{T}}\mathbf{D}_x^{-1}\mathbf{J} = \mathbf{Q}_1^{\mathrm{T}}\mathbf{D}_x^{-1}(\mathbf{T}^{-1}\mathbf{I})^{+}\,\mathbf{T}^{-1} = (\mathbf{T}^{-1}\mathbf{S}^{\mathrm{T}}\mathbf{X})^{+}\,\mathbf{T}^{-1}$$

and

(VII.8) $$\mathbf{S}^{\mathrm{T}}\mathbf{S}_2\mathbf{D}_2^{-1}\mathbf{S}_2^{\mathrm{T}}\mathbf{S} = (\mathbf{T}^{\mathrm{T}})^{-1}(\mathbf{E} - (\mathbf{T}^{-1}\mathbf{S}^{\mathrm{T}}\mathbf{X})(\mathbf{T}^{-1}\mathbf{S}^{\mathrm{T}}\mathbf{X})^{+})\,\mathbf{T}^{-1}\,.$$

After substitution of (VII.7) and (VII.8) into (VII.1) one obtains (58).

A. VIII. The best linear estimate

The penalty for an arbitrary estimate of the type

(VIII.1) $$z = \mathbf{wy} + b$$

obtained after substitution of (VIII.1) into (59) is

(VIII.2) $$c/c_0 = c'/c_0 + [\langle z_0\rangle + r\langle z_x\rangle - \mathbf{w}(\langle\mathbf{y}\rangle + r\langle\mathbf{y}_x\rangle) - (1 + r)\,b]^2 : (1 + r)$$

where

(VIII.3) $$c'/c_0 = \langle z_0^2\rangle + r\langle z_x^2\rangle + \mathbf{w}(\langle\mathbf{yy}^{\mathrm{T}}\rangle + r\langle\mathbf{y}_x\mathbf{y}_x^{\mathrm{T}}\rangle)\,\mathbf{w}^{\mathrm{T}} - 2\mathbf{w}(\langle\mathbf{y}z_0\rangle + r\langle\mathbf{y}_x z_x\rangle) - [\langle z_0\rangle + r\langle z_x\rangle - \mathbf{w}(\langle\mathbf{y}\rangle + r\langle\mathbf{y}_x\rangle)]^2/(1 + r)$$

 is the penalty value for the case

$$\langle z_0 \rangle + r\langle z_x \rangle - \mathbf{w}(\langle \mathbf{y} \rangle + r\langle \mathbf{y}_x \rangle) - (1 + r)\, b = 0 . \tag{VIII.4}$$

This equation characterizes the optimum choice of the constant b for an arbitrary vector $\mathbf{w}$, as follows from (VIII.2),

$$c/c_0 \geqq c'/c_0 . \tag{VIII.5}$$

But from (VIII.4) we obtain (62). Substituting (62) into (VIII.3) and using (64)–(67) one gets the equation

$$c/c_0 = \langle (\hat{z}_0 - \mathbf{w}\hat{\mathbf{y}})^2 \rangle + r\langle (\hat{z}_x - \mathbf{w}\hat{\mathbf{y}}_x)^2 \rangle \tag{VIII.6}$$

which is formally the same as the penalty (26) with the square errors (22) and (23) minimization of which gave the best unconditional estimator $\mathbf{w}_r$ (38). Therefore, the best choice $\mathbf{w}_0$ of the vector $\mathbf{w}$ minimizing (VIII.6) is also given by (38) after substitutions corresponding to (63)–(67).

A. IX. Errors of the best linear estimate

Mean errors (68) and (69) follow by the substitution of (62) into (60) and (61). Eq. (70) may be obtained from the definitions (59)–(61) using the linear estimator (63) with the constant term (62).

(Received December 18, 1970.)

REFERENCES

[1] M. Blum: An extension of the minimum mean square prediction theory for sampled input signals. IRE Trans. *IT-2* (1956), 176–184.

[2] P. Kovanic: Optimum digital operators. I.F.I.P. Congress 1968, Edinburgh, 5–10 Aug. 1968.

[3] H. P. Deull, P. L. Odell: A note concerning a generalization of the Gauss-Markov theorem. Texas Journal of Science *27*, (1966), 1, 21–24.

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

[5] И. Д. Крутько: Статистическая динамика импульсных систем (Statistical dynamics of impulse systems). Советское радио, Москва 1963.

[6] R. J. Hansen, Ch. L. Lawson: Extensions and Applications of the Householder Algoritmus for Solving Linear Least Squares Problems. Mathematics of Computation *23* (1969), 108, 787–812.

Odhad s nejmenším penále

PAVEL KOVANIC

V článku se zavádí zobecněný typ lineárního odhadu, minimalizující vážený součet středních čtverců celkové chyby odhadu a chyby zpracování užitečné složky dat. Minimalizovaná veličina se nazývá penále; její první složka odráží vliv rušivých náhodných složek dat a druhá složka charakterizuje nepřesnosti, s nimiž by byla požadovaná transformace realisována při vymizení rušivých složek. Ukazuje se, že odhad s nejmenším penále zahrnuje jako zvláštní případy jak tzv. nestranný odhad s minimální disperzí (odhad Zadeha a Ragazziniho), tak i nepodmíněný odhad s minimální disperzí (odhad Semjonova), jakož i další známé typy lineárních odhadů. Zobecněný odhad dovoluje využít apriorní informaci (znalost korelačních funkcí) ke snížení chyby odhadu. Výběr poměrných vah obou složek penále umožňuje přizpůsobit vlastnosti odhadu předpokládané aplikaci.

Ing. Pavel Kovanic, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Vyšehradská 49, Praha 2.