NUMERICAL METHOD OF CONSTRUCTION THE QUADRATIC DISCRIMINATION CRITERION MINIMIZING THE MAXIMUM OF VARIATION

V.Y. SHELEKHOVA, Y.P. YURACHKOVSKI*

The problem of adequate regression function choice is considered. The concept of normalized quadratic discrimination criterion (NQDC) is introduced. Numerical method based on the theory of linear operators eigenvalues disturbances for the purpose of construction the NQDC with minimum of its maximum variation is offered. Examples illustrating offered method applications are given.

1. Introduction

Discrimination criteria are intended to choose from a given finite set of known functions (with unknown parameters) the one which coincides with unknown "true" dependence kept in the sample of observations and distorted with additive noise. The value of discrimination criterion is used as a measure of distance from any function to the "true" dependence. The quality of discrimination criterion can be compared by means of different approaches based on the various definitions of discrimination criterion optimality. Here we suppose, that the criterion minimizing the maximum of variation is the optimal one.

2. Assumptions and Problem Statement

Let's suppose that the output of a random experiment, carried out in some points $v \in V \subset \mathbb{R}^m$ is the observed random value $y(v,\omega)$, $\omega \in \Omega$, where Ω is the space of elementary events. We also suppose that random values $y(v,\omega)$ are independent and identically distributed and they have an unknown mathematical expectation $\overline{y}(v)$, unknown independent from v finite variation σ^2 and equal to zero coefficients of asymmetry and excess:

*Cybernetics Institute of Ukraine Academy of Sciences, 252028 Kiev, Ukraine–USSR

$$Ey(v,\omega) = \overline{y}(v) \tag{1}$$

$$E(y(v,\omega) - \overline{y}(v))^2 = Dy(v,\omega) = \sigma^2 < \infty \tag{2}$$

$$E(y(v,\omega) - \overline{y}(v))^3/\sigma^3 = 0 \tag{3}$$

$$E(y(v,\omega) - \overline{y}(v))^4/\sigma^4 - 3 = 0 \tag{4}$$

where E means operator of mathematical expectation.

The number of experiments we designate as n and the points in which the experiment was made are $w_1, ..., w_n$. Let $f_\alpha(\cdot,\cdot)$, $f_\beta(\cdot,\cdot)$ be two functions determined on the sets $V \times \Theta_\alpha$, $V \times \Theta_\beta$ correspondingly. Furthermore, let these functions be linear by the parameters $\Theta_\alpha \in \Theta_\alpha$, $\Theta_\beta \in \Theta_\beta$ correspondingly, i.e.

$$f_\alpha(\cdot, \Theta_\alpha) = \Theta_\alpha^T \phi(\cdot) \tag{5}$$

$$f_\beta(\cdot, \Theta_\beta) = \Theta_\beta^T \phi(\cdot) \tag{6}$$

where

$$\phi(\cdot) = [\phi_1(\cdot), ..., \phi_{m_\alpha}(\cdot)]^T \quad \phi(\cdot) = [\phi_1(\cdot), ..., \phi_{m_\beta}(\cdot)]^T$$

Function $f_\alpha(\cdot,\cdot)$, $f_\beta(\cdot,\cdot)$ are called structures of regression models or simply structures and instead of $f_\alpha(\cdot,\cdot)$, $f_\beta(\cdot,\cdot)$ we shall write f_α, f_β accordingly.

Definition 1. Structure f_γ is adequate to random value $y(\cdot,\omega)$ on the set $V_0 \subset V$ if there exists such a value $\Theta_\gamma^0 \times \Theta_\gamma$ that $f_\gamma(v, \Theta_\gamma^0) = \overline{y}(v)$, $v \in V_0$.

In terms defined above our problem is to find an adequate (on some set) structure among f_α and f_β if one of them is adequate.

Function $\overline{y}(v)$ is unknown, therefore a definition can't be applied for solving this problem. Below we shall use statistical estimations of structure which may be adequate.

Let $y_i = (w_i, \omega)$, $i = 1, ..., n$,

$$Y = [y_1, ..., y_n]^T$$

$$\overline{Y} = [\overline{y}(w_1), ..., \overline{y}(w_n)]^T$$

Definition 2. Value

$$cr(f_\gamma) = \boldsymbol{Y}^T \boldsymbol{F}_\gamma \boldsymbol{Y} \tag{7}$$

is called normalized quadratic discrimination criterion (NQDC) if for matrix F_γ the following properties are satisfied:

1) $\boldsymbol{F} = \boldsymbol{F}^T$

2) $\boldsymbol{F} \geq 0$

3) if structure f_γ is adequate on the set $\{w_1, ..., w_n\}$ then $\overline{\boldsymbol{Y}}^T \boldsymbol{F}_\gamma \overline{\boldsymbol{Y}} = 0$,

4) $tr\boldsymbol{F}_\gamma = 1$.

Statement 1. If structure f_β is adequate on the set $\{w_1, ..., w_n\}$, then for any $\boldsymbol{F}_\alpha$ and $\boldsymbol{F}_\beta$

$$E\ cr(f_\alpha) - E\ cr(f_\beta) \geq 0 \tag{8}$$

Proof. It's easy to obtain the following equality

$$E\ cr(f_\gamma) = E[\boldsymbol{Y}^T \boldsymbol{F}_\beta \boldsymbol{Y}] = \overline{\boldsymbol{Y}}^T F_\gamma \overline{\boldsymbol{Y}} + \sigma^2 \tag{9}$$

for any γ. In addition $\overline{\boldsymbol{Y}}^T \boldsymbol{F}_\beta \overline{\boldsymbol{Y}} = 0$ and $\overline{\boldsymbol{Y}}^T \boldsymbol{F}_\alpha \overline{\boldsymbol{Y}} \geq 0$. Statement (8) follows from (9) and these inequalities.

So we see that

$$\hat{f} = \arg \min_{f_\gamma \in \{f_\alpha, f_\beta\}} cr(f_\gamma) \tag{10}$$

may be used as estimation of the adequate structure on the set $\{w_1, ..., w_n\}$. Matrix $\boldsymbol{F}_\gamma$ in (7) an in (10) is not unique. Arbitrariness in its choice we shall use to optimize discrimination criterion quality.

Let

$$\boldsymbol{X} = [x_{ij}] = [\phi_j(w_i)] \quad \boldsymbol{Z} = [z_{ij}] = [\phi_j(w_i)]$$

The criterion $d(\alpha/\beta)$ of discrimination criterion quality assumes a maximum of variation of random value $cr(f_\alpha)$ calculated in supposition that structure f_β is adequate on the set $\{w_1, ..., w_n\}$ and $\|\boldsymbol{\Theta}_\beta^0\|^2 = \boldsymbol{\Theta}_\beta^{0T} \boldsymbol{\Theta}_\beta^0 = \eta^2 > 0$, i.e.

$$d(\alpha/\beta) = \max_{\overline{Y} \in \mathcal{Y}_\beta} D\ cr(f_\alpha)$$

where

$$\mathcal{Y}_\beta = \{\overline{Y} : \overline{Y} = \boldsymbol{Z}\boldsymbol{\Theta}_\beta,\ \|\boldsymbol{\Theta}_\beta\|^2 = \eta^2 > 0\}$$

We define also

$$F_D^*(\alpha/\beta) = \arg\min_{\{F_\alpha\}} d(\alpha/\beta)$$

Our main purpose is to create the numerical algorithm of matrix $\boldsymbol{F}_D^*(\alpha/\beta)$ construction.

3. Method and Mathematical Substantiation

Let $\boldsymbol{N}$ be such a matrix that

$$\boldsymbol{N}^T\boldsymbol{N} = \boldsymbol{I}$$

$$\boldsymbol{N}\boldsymbol{N}^T = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$$

and

$$\boldsymbol{N}^T\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{N} = \boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{U}^T \tag{11}$$

is the spectral factorization in which

$$\boldsymbol{\Gamma} = \operatorname{diag}(\gamma_1, ..., \gamma_n) = \begin{bmatrix} \gamma_1 & & 0 \\ & \ddots & \\ 0 & & \gamma_n \end{bmatrix}, \quad \gamma_1 \le ... \le \gamma_n$$

It is easy to prove (Yurachkovski) that any matrix $\boldsymbol{F}_\alpha$ may be represented in the form

$$\boldsymbol{F}_\alpha = \boldsymbol{N}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{N}^T \tag{12}$$

where $\boldsymbol{A}$ is such arbitrary matrix that

$$tr\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{I}$$

As it was proved by Seber (1980)

$$\boldsymbol{D}\ cr(f_\alpha) = \boldsymbol{D}[\boldsymbol{Y}^T\boldsymbol{F}_\alpha\boldsymbol{Y}] = 2\sigma^4 tr\boldsymbol{F}_\alpha^2 + 4\sigma^2\overline{\boldsymbol{Y}}^T\boldsymbol{F}_\alpha\overline{\boldsymbol{Y}}$$

therefore

$$d(\alpha/\beta) = \max_{\overline{\boldsymbol{Y}}\in\mathcal{Y}_\beta} \boldsymbol{D}\ cr(f_\alpha) = 2\sigma^4 tr\boldsymbol{F}_\alpha^2 + 4\sigma^2\overline{\boldsymbol{Y}}^T\boldsymbol{F}_\alpha\overline{\boldsymbol{Y}} \tag{13}$$

$$= 2\sigma^4 tr\boldsymbol{F}_\alpha^2 + 4\sigma^2 \max_{\overline{\boldsymbol{Y}}\in\mathcal{Y}_\beta} \overline{\boldsymbol{Y}}^T\boldsymbol{F}_\alpha\overline{\boldsymbol{Y}} \tag{14}$$

$$= 2\sigma^4 tr\boldsymbol{F}_\alpha^2 + 4\sigma^2 \max_{\|\boldsymbol{\Theta}_\beta\|^2=\eta^2} \boldsymbol{\Theta}^T\boldsymbol{Z}^T\boldsymbol{F}_\alpha\boldsymbol{Z}\boldsymbol{\Theta} \tag{15}$$

$$= 2\sigma^4 tr\boldsymbol{F}_\alpha^2 + 4\sigma^2\eta^2\lambda_{-1}[\boldsymbol{Z}^T\boldsymbol{F}_\alpha^2\boldsymbol{Z}] \tag{16}$$

where $\lambda_{-1}[C]$ designates the greatest eigenvalue of matrix $\boldsymbol{C}$.

From equalities (11), (12) and $\lambda_{-1}[\boldsymbol{C}\boldsymbol{C}^T] = \lambda_{-1}[\boldsymbol{C}^T\boldsymbol{C}]$ follows that

$$d(\alpha/\beta) = 2\sigma^4 tr(\boldsymbol{A}\boldsymbol{A}^T)^2 + 4\sigma^2\eta^2\lambda_{-1}[\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{A}^T].$$

Let $h = \sigma^2/(2\eta^2)$ then

$$\boldsymbol{F}_D^*(\alpha/\beta) = \boldsymbol{N}(\boldsymbol{A}\boldsymbol{A}^T)_D^*\boldsymbol{N}^T$$

where

$$(\boldsymbol{A}\boldsymbol{A}^T)_D^* = \arg \min_{\{\boldsymbol{A}\boldsymbol{A}^T:tr\boldsymbol{A}\boldsymbol{A}^T=1\}} \left(\lambda_{-1}[\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{A}^T] + h\cdot tr(\boldsymbol{A}\boldsymbol{A}^T)^2\right)$$

or

$$\boldsymbol{F}_D^*(\alpha/\beta) = \boldsymbol{N}\boldsymbol{U}(\boldsymbol{B}\boldsymbol{B}^T)_D^*\boldsymbol{U}^T\boldsymbol{N}^T$$

where

$$(\boldsymbol{B}\boldsymbol{B}^T)_D^* = \arg \min_{\{\boldsymbol{B}\boldsymbol{B}^T:tr\boldsymbol{B}\boldsymbol{B}^T=\boldsymbol{I}\}} \left(\ \lambda_{-1}[\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{\Gamma}\boldsymbol{B}\boldsymbol{B}^T] \ + \ htr(\boldsymbol{B}\boldsymbol{B}^T)^2\ \right)$$

and

$$\boldsymbol{B}\boldsymbol{B}^T = \boldsymbol{U}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{U}^T$$

We shall accomplish the search of $(\boldsymbol{B}\boldsymbol{B}^T)^*_D$ by means of the gradient projection method. This method consists of two procedures:

1) finding the direction of the greatest decreasing of the minimized function

$$\Phi(\boldsymbol{B}) = \lambda_{-1}[\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{\Gamma}\boldsymbol{B}\boldsymbol{B}^T] + htr(\boldsymbol{B}\boldsymbol{B}^T)^2$$

2) projection of obtained direction vector on the permissible domain

$$\mathcal{B} = \{\boldsymbol{B}\boldsymbol{B}^T : tr\boldsymbol{B}\boldsymbol{B}^T = \boldsymbol{I}\}$$

The direction of the greatest decreasing of function $\Phi(\boldsymbol{B})$ is given by matrix $-\partial\Phi(\boldsymbol{B})/\partial\boldsymbol{B}$. It is easy to see that

$$\frac{\partial\Phi(\boldsymbol{B})}{\partial\boldsymbol{B}} = \left[\frac{\partial\Phi(\boldsymbol{B})}{\partial b_{ij}}\right] = \left[\frac{\partial\lambda_{-1}[\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{\Gamma}\boldsymbol{B}\boldsymbol{B}^T]}{\partial b_{ij}}\right] + h\left[\frac{\partial tr(\boldsymbol{B}\boldsymbol{B}^T)^2}{\partial b_{ij}}\right]$$

If multiplicity of $\lambda_{-1}[\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{\Gamma}\boldsymbol{B}\boldsymbol{B}^T]$ is one then

$$\frac{\partial\lambda_{-1}[\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{\Gamma}\boldsymbol{B}\boldsymbol{B}^T]}{\partial b_{ij}} =$$

$$\lim_{\varepsilon\to 0}\frac{\lambda_{-1}[(\boldsymbol{B}+\varepsilon G_{ij})(\boldsymbol{B}+\varepsilon G_{ij})\boldsymbol{\Gamma}(\boldsymbol{B}+\varepsilon G_{ij})(\boldsymbol{B}+\varepsilon G_{ij})^T] + \lambda_{-1}[\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{\Gamma}\boldsymbol{B}\boldsymbol{B}^T]}{\varepsilon}$$

where G_{ij} is a matrix whose element with number (i,j) is equal to one and the rest of the elements are equal to zero.

According to the eigenvalue perturbation theory the last limit may be easily calculated. It is equal to

$$\boldsymbol{v}^T_{-1}(G_{ij}\boldsymbol{B}^T\boldsymbol{\Gamma}\boldsymbol{B}\boldsymbol{B}^T + \boldsymbol{B}G^T_{ij}\boldsymbol{\Gamma}\boldsymbol{B}\boldsymbol{B}^T + \boldsymbol{B}^T\boldsymbol{\Gamma}G_{ij}\boldsymbol{B}^T + \boldsymbol{B}\boldsymbol{B}^T\boldsymbol{\Gamma}\boldsymbol{B}G^T_{ij})\boldsymbol{v}_{-1}$$

where $\boldsymbol{v}_{-1}$ is matrix $\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{\Gamma}\boldsymbol{B}\boldsymbol{B}^T$ eigenvector corresponding to the greatest eigenvalue. Thus

$$\frac{\partial\lambda_{-1}[\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{\Gamma}\boldsymbol{B}\boldsymbol{B}^T]}{\partial b_{ij}} = 2\boldsymbol{v}_{-1}\boldsymbol{v}^T_{-1}\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{\Gamma}\boldsymbol{B} + 2\boldsymbol{\Gamma}\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{v}_{-1}\boldsymbol{v}^T_{-1}\boldsymbol{B} \qquad (17)$$

It is easy to show that

$$h\frac{\partial tr(\boldsymbol{B}\boldsymbol{B}^T)^2}{\partial b_{ij}} = 4h\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{B} \tag{18}$$

From (13), (14) we obtain that

$$\frac{\partial\Phi(\boldsymbol{B})}{\partial\boldsymbol{B}} = 2\boldsymbol{v}_{-1}\boldsymbol{v}_{-1}^T\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{\Gamma}\boldsymbol{B} + 2\boldsymbol{\Gamma}\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{v}_{-1}\boldsymbol{v}_{-1}^T\boldsymbol{B} + 4h\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{B} \tag{19}$$

Denote $\boldsymbol{B}_k\boldsymbol{B}_k^T$ the approximation of k–th iteration. Then in procedure 1 of $(k+1)$–th iteration consists in calculation of

$$\widetilde{\boldsymbol{B}}_{k+1} = \boldsymbol{B}_k - s\frac{\partial\Phi(\boldsymbol{B})}{\partial\boldsymbol{B}}\Big|_{\boldsymbol{B}=\boldsymbol{B}_k} = \boldsymbol{B}_k - s\Phi'(\boldsymbol{B}_k)$$

where $s > 0$ is the step of $(k+1)$–th iteration.

Procedure 2 consists of projecting the matrix $\widetilde{\boldsymbol{B}}_{k+1}\widetilde{\boldsymbol{B}}_{k+1}^T$ on domain $\mathcal{B}$, i.e.

$$\boldsymbol{B}_{k+1}\boldsymbol{B}_{k+1}^T = \widetilde{\boldsymbol{B}}_{k+1}\widetilde{\boldsymbol{B}}_{k+1}^T/tr\widetilde{\boldsymbol{B}}_{k+1}\widetilde{\boldsymbol{B}}_{k+1}^T =$$

$$= \frac{(\boldsymbol{B}_k - s\Phi'(\boldsymbol{B}_k))(\boldsymbol{B}_k - s\Phi'(\boldsymbol{B}_k))^T}{tr(\boldsymbol{B}_k - s\Phi'(\boldsymbol{B}_k))(\boldsymbol{B}_k - s\Phi'(\boldsymbol{B}_k))^T} \tag{20}$$

If multiplicity of eigenvalue $\lambda_{-1}[\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{\Gamma}\boldsymbol{B}\boldsymbol{B}^T]$ is not one formula (13) is not correct and hence formula (16) does not lead to the step of gradient projection method. If the only multiplicity of $\lambda_{-1}[\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{\Gamma}\boldsymbol{B}\boldsymbol{B}^T]$ isn't one after one or more steps performed by formula (16) algorithm comes to the point where multiplicity of eigenvalue is one and the using of formula (14) becomes well–grounded.

4. Algorithm Description

The gradient search is used in the algorithm. Therefore, the speed of the alghorithm depends considerably on the initial approximation of matrix $\boldsymbol{B}$ (or $\boldsymbol{B}\boldsymbol{B}^T$), direction of motion and step at every iteration, and halt criteria. *Initial approximation of matrix* $\boldsymbol{B}\boldsymbol{B}^T$ we obtain as a weighted sum of two matrix, i.e.

$$\boldsymbol{B}\boldsymbol{B}^T = p_\lambda(\boldsymbol{B}\boldsymbol{B}^T)_\lambda + p_h(\boldsymbol{B}\boldsymbol{B}^T)_h$$

where $(\boldsymbol{BB}^T)_\lambda$ and $(\boldsymbol{BB}^T)_h$ are matrices on which minimum of functionals $\lambda_{-1}[\boldsymbol{BB}^T\boldsymbol{\Gamma}\boldsymbol{BB}^T]$ and $tr[(\boldsymbol{BB}^T)^2]$ are achieved correspondingly. Note that this matrices are diagonal. Elements of matrix $(\boldsymbol{BB}^T)$ depend on n_2 – number of zeroes in $\gamma = \text{diag}\,(\gamma_1,...,\gamma_n) : (\boldsymbol{BB}^T)_\lambda = \text{diag}\,(1/n_z,...,1/n_z,0,...,0)$, if n_z isn't zero ; otherwise, elements $(\boldsymbol{BB}^T)_\lambda$ are calculated from equations

$$\begin{cases} (\boldsymbol{BB}^T)^i_\lambda \gamma_i^{1/2} = (\boldsymbol{BB}^T)^j_\lambda \gamma_j^{1/2} & \text{for every} \quad i,j = 1,...,n \\ \sum_{i=1}^n (\boldsymbol{BB}^T)^i_\lambda = 1 \end{cases}$$

Hence, i–th element of $\text{diag}(\boldsymbol{BB}^T)_\lambda$ is

$$(\boldsymbol{BB}^T)^i_\lambda = \frac{1}{\gamma_i^{1/2}\left(\sum_{j=1}^n 1/\gamma_j^{1/2}\right)}$$

Diagonal elements of matrix $(\boldsymbol{BB}^T)^i_h$ are equal to $1/n$. Weights p_λ and p_h depend on $\{\gamma_i\}$ and h.

Choice of direction and step. At any point of iteration algorithm, motion is conducted towards the antigradient $-\partial\Phi(\boldsymbol{BB})/\partial\boldsymbol{BB}$. The norm of $\partial\Phi(\boldsymbol{BB})/\partial\boldsymbol{BB}$ depends linearly on the values $\{\gamma_i\}$ and h, so that if the value $\|\partial\Phi(\boldsymbol{BB})/\partial\boldsymbol{BB} \cdot \text{step}\|$ isn't very large (or small) it's necessary to normalize values $\{\gamma_i\}$ and h.

The motion is conducted until Φ_k decreases, where Φ_k is the value of functional Φ at the k–th iteration. The increasing of Φ_k is possible either when the step exceeds distance to extremum or when antigradient of functional Φ_k is considerably changed. The last case we shall determine below as a leap over junction of functional Φ. In both cases the motion should be continued, the step decreased beforehand. Note, that the iteration process can be conducted along the functional junction. In this case the rapid step decrease will result in unnecessary growth of calculation time. To obtain available value of eigenvalue $\lambda_{-1}[\boldsymbol{BB}^T\boldsymbol{\Gamma}\boldsymbol{BB}^T]$ Jacobi algorithm is used, and to get corresponding eigenvector $v_{-1}[\boldsymbol{BB}^T\boldsymbol{\Gamma}\boldsymbol{BB}^T]$ power method is used.

Stop criterion. During the iteration process the minimal value of Φ (we denote it Φ_{opt}) and the matrix in which it's achieved are saved. The criterion of the end of calculations consists in unchangeability of Φ_{opt} during N times leaps over junctions.

Let's mark the additional peculiarities of the given algorithm. On some steps of the gradient method it may occur that matrix $\boldsymbol{BB}^T$ becomes

diagonal, in this case the further search will remain within the space of the diagonal matrix. To avoid this it is necessary sometimes to check the diagonality of $\boldsymbol{BB}^T$ and, when diagonality takes place, it is necessary to noise up the matrix elements:

$$(\boldsymbol{BB}^T)^{ij}_{k+1} = (\boldsymbol{BB}^T)^{ij}_k + \varepsilon_k,$$

where ε_k is the monotonously decreasing sequence. Apart from the checking up of diagonality it's necessary to watch if $\boldsymbol{BB}^T$ is a singular matrix. In algorithm a Cholesky triangle factorization procedure for positive definite symmetric matrix is used. Noising up of diagonal elements makes it possible to leave the class of the singular matrix, because for any singular matrix $\boldsymbol{A}(\boldsymbol{A} + \xi I)$ is a nonsingular one for any small $\xi > 0$.

To increase the speed of calculations eigenvector v_{-1} of the k–th iteration is handed as an initial approximation of v_{-1} at the $(k+1)$-th iteration. When doing this we must bear in mind that the junction of functional means that λ_{-1} is considerably changed. Thus it can occur that $\boldsymbol{v}_{-1}$ from the previous iteration is saved as eigenvector, but it does not corresponds to λ_{-1}. It takes place, for example, for $h = 0$ eigenvectors of matrix $\boldsymbol{BB}^T\boldsymbol{\Gamma}\boldsymbol{BB}^T$ are ortonormal basis $e_1, ..., e_n$. If $\boldsymbol{v}_{-1}$ is e_i at the k–th iteration and leaps over the junction it becomes e_j at the $(k+1)$–th iteration $(i \neq j)$, power method will not detect the error. Thus we must change the initial approximation of $\boldsymbol{v}_{-1}$, when Φ_k increases:

$$\boldsymbol{v}^k_{-1} = \boldsymbol{v}^{k+1}_{-1} + \boldsymbol{\delta}, \qquad \boldsymbol{\delta} = \text{ const}$$

Note, that every time the elements are noised up we must normalsize $\boldsymbol{v}_{-1}$ and $\boldsymbol{BB}^T$ by formulas $\|\boldsymbol{v}_{-1}\| = 1$, $tr\boldsymbol{BB}^T = \boldsymbol{I}$.

In the algorithm we use following values and constants:

$$p_\lambda/p_h = tr\gamma/h, \; gh = \max(\gamma_1, ..., \gamma_n h), \; \text{step}_k/\text{step}_{k+1} = 1.02,$$

$$\{\varepsilon_k\} : \varepsilon_0 = 0.01, \text{ and } \varepsilon_{k+1} = \varepsilon_k/2, \; \xi_k = 10\varepsilon_k, \; \delta = 0.01.$$

5. Examples

We have set up in all the examples the same matrix $\boldsymbol{\Gamma} = \text{ diag}(1, 4, 9, 16)$.

1. $h = 0$.

In this case we obviously obtained the diagonal matrix

$$(\boldsymbol{B}\boldsymbol{B}^T)^*_D = \mathrm{diag}(0.48,\ 0.24,\ 0.16,\ 0.12).$$

2. $h = 0.1$.

$$(\boldsymbol{B}\boldsymbol{B}^T)^*_D = \begin{bmatrix} 0.480478 & 0.000020 & 0.000031 & -0.000068 \\ & 0.239764 & -0.000176 & -0.000087 \\ & & 0.159956 & -0.000018 \\ & & & 0.119803 \end{bmatrix}.$$

3. $h = 100$.

It's easy to show that $(\boldsymbol{B}\boldsymbol{B}^T)^*_D \to \mathrm{diag}(1/n, ..., 1/n)$ when $h \to \infty$. In this example matrix $(\boldsymbol{B}\boldsymbol{B}^T)^*_D$ is close to the diagonal one. Really, the modules of all nondiagonal elements are less than 10^{-6}. In our calculations

$$(\boldsymbol{B}\boldsymbol{B}^T)^*_D = \mathrm{diag}(0.258929,\ 0.258929,\ 0.258929,\ 0.223214).$$

References

Yurachkovski Yu.P., *Analytical construction of optimal quadratic discrimination criterion.*– Soviet Journal of Automation and Information Sciences, v.21, No.1.

Seber J. (1980): *Liner regression analyses.*– Moscow, Mir, (in Russian).