On Minimax Identification: Method of Dual Optimization[1]

A.R. Pankov, E.N. Platonov, K.V. Siemenikhin

Institute for Information Transmission Problems, Russian Academy of Sciences
B. Karetny per. 19, Moscow, GSP-4, 101447, Russia

e-mail: `lucky1@orc.ru`

Abstract

The problem of minimax affine identification of the linear uncertain-stochastic multivariate model is considered. The minimax optimization problem together with the corresponding dual one are stated and examined. The necessary and sufficient conditions for the minimax affine estimate to exist and to be determined analytically via the dual problem solution are given. The algorithm of minimax stochastic estimation for the infinite-dimensional model given a finite number of observations is also considered. The numerical method for the minimax estimation is described, and the results of computer modeling are presented.

1 Introduction

In the report, we consider a problem of linear estimation of parametric functions in the linear multidimensional regression model with statistically indeterminate parameters and disturbances. The practical importance of the model under consideration is well known since typical linear models which are usually used for the purposes of parameter estimation, filtering, and control [1–7] can be reduced to the general uncertain-stochastic model (US-model) [8], which is examined in this paper. Our model covers several particular types of linear stochastic models:

- — models with unknown nonrandom unbounded parameters with purely random disturbances (classical regression models) [1, 7];
- — static and dynamic stochastic models with random parameters and disturbances [2];
- — models with unknown nonrandom but bounded parameters and disturbances (set-membership models) [3];
- — models with random parameters and disturbances with partially known probabilistic characteristics (statistically indeterminate stochastic models) [4].
- — dynamic models described by stochastic linear differential equations in the separable Hilbert space given a finite number of observed functionals [9].

Traditionally the aforementioned models are examined separately by means of methods and approaches of different nature [5]. From our results it follows that there exists a unified method for linear identification of US-models based on the minimax approach and the duality theory [6–8]. In this report we state the minimax-stochastic criterion for solving the identification problem for the US-model, and provide the necessary and sufficient conditions for the minimax affine estimate to exist. We also give the explicit expression for the minimax estimator and provide the necessary and sufficient conditions for the last to be determined analytically via the solution of the dual optimization problem. We also consider some numerical algorithms of the convex nonlinear programming which make possible to solve the dual optimization problem and hence to derive the minimax affine estimate. It is shown that even in the case, when the minimax affine estimating operator can not be expressed as a function of the dual problem solution, the last makes possible to simplify significantly the original minimax optimization problem. The corresponding examples and the results of the numerical experiments are presented. We also examine the minimax estimation problem for the purely stochastic infinite-dimensional linear model with finite-dimensional observations. In this case, the minimax estimator is obtained using a solution of the corresponding dual optimization problem.

2 Model description

We shall use the following notation: $\mathbf{E}_{\mathbf{P}}\{\cdot\}$ is the mathematical expectation with respect to the distribution $\mathbf{P}$; $\mathrm{co}[\mathcal{K}]$ is the convex hull of the set $\mathcal{K}$; I is the identity matrix; $\mathrm{rg}\{A\}$ and $\mathrm{tr}\{A\}$ are the range and the trace of the matrix A, respectively; $A \leq B$ means

---

[1] The work is supported by Russian Foundation for Basic Research under Contract 99-01-01088

that $B - A$ is positive-semidefinite; A^+ is the Moor–Penrose pseudoinverse [1] of A; A^* is the transposed matrix; $\mathrm{col}[x_1, \dots, x_n] = (x_1^*, \dots, x_n^*)^*$; $\|x\| = \sqrt{x^* x}$; $\arg\min_{x \in X} g(x)$ ($\arg\max_{x \in X} g(x)$) is the set of all minimum (maximum) points of the functional $g(x)$ on X.

Let us consider the following US-model:

$$\begin{cases} x = \Phi_0 \theta_0 + \Phi \theta_1, \\ y = \Psi_0 \theta_0 + \Psi \theta_1, \end{cases} \tag{1}$$

where $x \in \mathbf{R}^m$ is the vector to be estimated in a non-parametric way given the observation vector $y \in \mathbf{R}^n$; $\theta = \mathrm{col}[\theta_0, \theta_1]$ is the vector of the model (1) parameters and disturbances. It is supposed that the sub-vectors $\theta_0 \in \mathbf{R}^p$ and $\theta_1 \in \mathbf{R}^q$ differ with respect to the amount of available *a priori* information, namely, θ_0 is unbounded unknown vector, some components of which are unknown deterministic while the others are random with unknown distributions; θ_1 is statistically indeterminate random vector with some mean value $m_\theta = \mathbf{E}\{\theta_1\}$ and covariance matrix $R_\theta = \mathbf{cov}(\theta_1, \theta_1)$. It is supposed that $m_\theta \in \mathcal{M}$ and $R_\theta \in \mathcal{R}$, where $\mathcal{M}$ and $\mathcal{R}$ are some known sets which describe the rate of *a priori* statistical uncertainty of the θ_1 distribution. Concerning $\mathcal{M}$ and $\mathcal{R}$, we assume the set $\mathcal{M}$ to be compact and centrally symmetric with respect to some known point $m_0 \in \mathbf{R}^q$ (i.e. if $u = m_0 + \Delta u \in \mathcal{M}$ then $u' = m_0 - \Delta u \in \mathcal{M}$), $\mathcal{R}$ is supposed to be a compact set of symmetric $q \times q$-dimensional positive-semidefinite matrices. The matrices Φ_0, Φ, Ψ_0, Ψ are known and have appropriate dimensions.

Let $\mathbf{P}_\theta$ denote the generalized* probabilistic distribution of the vector θ in (1). The stated above assumptions mean that $\mathbf{P}_\theta \in \mathcal{P}$, where $\mathcal{P}$ is a set of all possible distributions:

$$\begin{aligned} \mathcal{P} = \{\mathbf{P}_\theta : \theta = \mathrm{col}[\theta_0, \theta_1],\ m_\theta = \mathbf{E}\{\theta_1\} \in \mathcal{M}, \\ R_\theta = \mathbf{cov}(\theta_1, \theta_1) \in \mathcal{R}\}. \end{aligned} \tag{2}$$

Note that model (1) possesses the following specific features:

- — we do not distinguish the structural model parameters (for example, parameters of the signal model) and the disturbing parameters (such as uncertain and random observation disturbances) since the corresponding partitioning can be done explicitly by an appropriate choosing of the matrices Φ_0, Φ, Ψ_0, Ψ;
- — we assume the structural model parameters and the disturbances to be dependent in general;
- — all probabilistic distributions may be singular: in particular, any covariance matrices from $\mathcal{R}$ are not supposed to be strictly positive (this allows us not to separate the random and deterministic variables).

3 Problem statement

Let $\mathcal{F}$ be some *a priori* choosen class of estimators of x given the observations y. Then for any $\mathbf{F} \in \mathcal{F}$ $\widetilde{x} = \mathbf{F}y$ is some admissible estimate of x. The estimate $\widetilde{x}$ accuracy is measured by the following mean-square criterion:

$$\Im(\mathbf{F}, \mathbf{P}_\theta) = \mathbf{E}_{\mathbf{P}_\theta} \left\{ \|\mathbf{F}y - x\|^2 \right\}, \tag{3}$$

where $\mathbf{P}_\theta$ is any possible distribution of θ from $\mathcal{P}$. Since $\mathbf{P}_\theta$ is not known, we can not calculate the exact value of $\Im(\mathbf{F}, \mathbf{P}_\theta)$. Therefore, we are going to find the estimating operator $\widehat{\mathbf{F}}$ that is optimal in a minimax sense.

Definition 1 *The estimating operator* $\widehat{\mathbf{F}}$ *and the corresponding estimate* $\hat{x} = \widehat{\mathbf{F}}y$ *are called the minimax ones if*

$$\widehat{\mathbf{F}} \in \arg\min_{\mathbf{F} \in \mathcal{F}} \sup_{\mathbf{P}_\theta \in \mathcal{P}} \Im(\mathbf{F}, \mathbf{P}_\theta). \tag{4}$$

In order (4) to possess sense it is necessary that

$$\exists \mathbf{F} \in \mathcal{F} : \quad \sup_{\mathbf{P}_\theta \in \mathcal{P}} \Im(\mathbf{F}, \mathbf{P}_\theta) < \infty. \tag{5}$$

Definition 2 *The vector* x *admits* [1] *the estimate defined by the operators from* $\mathcal{F}$ *given the observations* y *if* (5) *is fulfilled.*

Further, we study the case in which $\mathcal{F}$ is the class of all affine transformations, i.e., $\mathbf{F} = (F, f) \in \mathcal{F}$ means that $\mathbf{F}y = Fy + f$, where $F \in \mathbf{R}^{m \times n}$ defines a linear transformation and $f \in \mathbf{R}^m$ is a slide.

4 Main results

4.1 Common structure of minimax problem

The first theorem concerns the existence of the minimax affine estimate.

Theorem 1 *The vector* x *admits the affine estimate given* y *and the minimax affine estimate exists if and only if*

$$\Phi_0 = \Phi_0 \Psi_0^+ \Psi_0. \tag{6}$$

The stated above condition (6) for the minimax estimate to exist means that the class of unbiased (with respect to θ_0) estimators $\mathcal{F}_0 = \{F : F \in \mathbf{R}^{m \times n}, F\Psi_0 = \Phi_0\}$ is not empty.

---

*i.e., it may contain singular components

In particular, condition (6) is fulfilled whenever the matrix Ψ_0 is of the full range.

In order to formulate the following result we have to introduce the auxiliary functional $J(F,K)$ defined for all $F \in \mathbf{R}^{m\times n}$ and $K \in \mathcal{K}$, where

$$\mathcal{K} = \{K : K = (m_\theta - m_0)(m_\theta - m_0)^* + R_\theta,\\ m_\theta \in \mathcal{M},\, R_\theta \in \mathcal{R}\}.$$

By definition, put

$$J(F,K) = \sup\left\{\Im((F,(\Phi - F\Psi)m_0),\mathbf{P}_\theta) :\\ \mathbf{E}\{(\theta_1 - m_0)(\theta_1 - m_0)^*\} = K\right\}.$$

Now we can show that the problem of minimax affine estimation can be reduced to the problem of minimax linear estimation.

Theorem 2 *Let* (6) *hold, then*

1) *the affine estimate $\hat{x} = \widehat{\mathbf{F}}y = \Phi m_0 + \hat{F}(y - \Psi m_0)$ is minimax if and only if*

$$\hat{F} \in \arg\min_{F\in\mathcal{F}_0}\sup_{K\in\mathcal{K}} J(F,K), \tag{7}$$

where $\mathcal{F}_0 = \{F : F \in \mathbf{R}^{m\times n},\, F\Psi_0 = \Phi_0\}$,

$$J(F,K) = \begin{cases} \operatorname{tr}\{(F\Psi - \Phi)K(F\Psi - \Phi)^*\}, \\ \text{if } \; F \in \mathcal{F}_0,\, K \in \operatorname{co}[\mathcal{K}], \\ +\infty, \text{if } \; F \notin \mathcal{F}_0; \end{cases}$$

2) *the following duality condition is fulfilled:*

$$\min_{\mathbf{F}\in\mathcal{F}}\sup_{\mathbf{P}_\theta\in\mathcal{P}} \Im(\mathbf{F},\mathbf{P}_\theta) = \max_{K\in\operatorname{co}[\mathcal{K}]} \underline{J}(K), \tag{8}$$

where $\underline{J}(K) = \inf_{F\in\mathcal{F}_0} J(F,K)$.

Note that dimension of the minimax optimization problem (7) is usually very large since it equals $m \times n$, where n is the number of observations. So, the direct numerical solution of (7) seems to be practically impossible.

4.2 Method of dual optimization

One of the major technique for designing a minimax solution is based on using a solution of the dual optimization problem. For the minimax problem under consideration, this method seems to be relevant since the duality condition (8) holds. In this case, the dual optimization problem is as follows

$$\overline{K} \in \arg\max_{K\in\operatorname{co}[\mathcal{K}]} \underline{J}(K). \tag{9}$$

The subsequent results show how the minimax affine estimate can be defined using the dual solution $\overline{K}$.

Theorem 3 *Let condition* (6) *hold, then*

$$\underline{J}(K) = \operatorname{tr}\{(F_0\Psi - \Phi)[K - P(K)\Psi K](F_0\Psi - \Phi)^*\}, \tag{10}$$

where $F_0 = \Phi_0\Psi_0^+$, $Q = I - \Psi_0\Psi_0^+$, $P(K) = K\Psi^*(Q\Psi K\Psi^*Q)^+$, and the following results are valid:

1) *there exists a solution $\overline{K}$ of the dual optimization problem* (9)*;*

2) *the affine estimate $\hat{x} = \widehat{\mathbf{F}}y = \Phi m_0 + \hat{F}(y - \Psi m_0)$ is minimax if*

$$\hat{F} = \widetilde{F} + \hat{H},$$
$$\widetilde{F} = F_0 + (\Phi - F_0\Psi)P(\overline{K}), \tag{11}$$

and $\hat{H}$ is an arbitrary solution of the auxiliary minimax optimization problem

$$\hat{H} \in \arg\min_{H\in\mathcal{H}}\sup_{K\in\mathcal{K}} J(\widetilde{F} + H, K), \tag{12}$$

where $\mathcal{H} = \{H \in \mathbf{R}^{m\times n} : H\overline{P} = H\}$, $\overline{P} = Q[I - \Psi P(\overline{K})]$;

3) $\sup_{\mathbf{P}_\theta\in\mathcal{P}} \Im(\widehat{\mathbf{F}},\mathbf{P}_\theta) = J(\hat{F},\overline{K}) = J(\widetilde{F},\overline{K}) = \underline{J}(\overline{K})$, *i.e., $\widehat{\mathbf{F}}$ is a guaranteeing on $\mathcal{P}$ affine estimator;*

4) *if $\overline{K} \in \mathcal{K}$, then for every $F \in \mathcal{F}_0$ and $K \in \mathcal{K}$*

$$J(\hat{F},K) \le J(\hat{F},\overline{K}) \le J(F,\overline{K}),$$

i.e., $(\hat{F},\overline{K})$ is a saddle point of the functional $J(\cdot)$ on $\mathcal{F}_0 \times \mathcal{K}$.

The presented above theorem is essentially based on the fact that the operator $\widetilde{F}$ (11) is designed according to the least squares method (LSM), namely

$$\widetilde{F} \in \arg\min_{F\in\mathcal{F}_0} J(F,\overline{K}). \tag{13}$$

Moreover, $\widetilde{F}$ has the least Euclidean norm in comparison with other solutions of problem (13). Motivated by this, the operator $\widetilde{F}$ will henceforth be referred to as the *LSM-operator*. Due to (11) we can see that $\widetilde{F}$ is a function of the dual solution $\overline{K}$ (9). Therefore, in the case

$$\hat{H} = O \in \arg\min_{H\in\mathcal{H}}\sup_{K\in\mathcal{K}} J(\widetilde{F} + H, K) \tag{14}$$

we obtain the analytical expression for the minimax estimator as soon as the dual problem has been solved. The following theorem provides the necessary and sufficient conditions for (14) to be fulfilled.

Theorem 4 *Under the conditions of Theorem 3, suppose that $\widetilde{F}$ is the LSM-operator* (11). *Then $\hat{x} = \widetilde{\mathbf{F}}y = \Phi m_0 + \widetilde{F}(y - \Psi m_0)$ is the minimax affine estimate if and only if there exists a matrix $\widetilde{K} \in \arg\max_{K\in\operatorname{co}[\mathcal{K}]} J(\widetilde{F},K)$ such that*

$$(\widetilde{F}\Psi - \Phi)\widetilde{K}\Psi^*\overline{P} = O. \tag{15}$$

Note that checking (15) often can be done without any difficulties since the functional $J(\widetilde{F}, K)$ is linear with respect to K. In the general case we can suggest several sufficient conditions for (15) to be valid.

Corollary 1 *The LSM-operator $\widetilde{F}$ defines the minimax affine estimate $\hat{x} = \widetilde{\mathbf{F}}y$ if one of the following conditions holds:*

1) $(\widetilde{F}\Psi - \Phi)K\Psi^*\overline{P} = O$ *for all* $K \in \mathcal{K}$;
2) $\Psi K\Psi^* \le \Psi\overline{K}\Psi^*$ *for all* $K \in \mathcal{K}$;
3) *the matrix* $\Psi\overline{K}\Psi^*$ *is strictly positive.*

The following well-known situations are often encountered in applications [4, 7, 8]:

i) $\Psi K\Psi^*$ is strictly positive for all $K \in \mathcal{K}$;
ii) the set $\mathcal{K}$ contains a maximal element $\overline{K}$, namely $\overline{K} \in \mathcal{K}$ and $K \le \overline{K}$ for all $K \in \mathcal{K}$.

Obviously, in both these cases the LSM-operator $\widetilde{F}$ defines the minimax affine estimate since the conditions of Corollary 1 hold.

It should be mentioned that even in the case, when (15) is not fulfilled, the decomposition $\hat{F} = \widetilde{F} + \hat{H}$ is of its practical importance since the set $\mathcal{H}$ in (12) has dimension $m \times n_1$, where $n_1 = \mathrm{rg}[\overline{P}] \le n$. Therefore in the case $n_1 \ll n$, we see that the auxiliary minimax optimization problem (12) is much more simple than the original one (7). In the case (15), it can be shown that $n_1 = 0$ and the solution $\hat{H}$ is of no use.

4.3 Minimax estimation for purely stochastic models

Now let us consider the problem of minimax estimation for the statistically indeterminate stochastic model. Suppose that the *a priori* information about the model parameters is formulated in term of the second-order moment characteristics of the vector $z = \mathrm{col}[x, y]$:

$$\mathbf{E}\{z\} = m_z, \quad \mathbf{cov}(z, z) \in \mathcal{R}, \tag{16}$$

where the mean $m_z = \mathrm{col}[m_x, m_y]$ and the set $\mathcal{R}$ are known. Concerning $\mathcal{R}$, we assume that $\mathcal{R}$ is a convex compact set of symmetric positive-semidefinite matrices $R = \begin{pmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{pmatrix}$. Clearly, the stated assumptions lead to a particular type of the US-model (1) if we take

$$\theta_0 = 0,\ \theta_1 = z,\ \Phi_0 = O,\ \Psi_0 = O,\ \Phi = \left(\ I_m\ ;\ O\ \right),$$

$$\Psi = \left(\ O\ ;\ I_n\ \right),\ \mathcal{K} = \mathrm{co}[\mathcal{K}] = \mathcal{R}.$$

The following result may be referred to as the minimax generalization of the theorem on normal correlation.

Theorem 5 *Let $\mathcal{F} = \mathcal{B}$ be the class of all Borel transformations $\mathbf{R}^n \to \mathbf{R}^m$, and $\mathcal{P}$ be the class of distributions $\mathbf{P}_z$ satisfying* (16)*. Then the following results are valid:*

1) *there exists a solution of the dual optimization problem*
$$\overline{R} \in \arg\max_{R \in \mathcal{R}} \underline{J}(R),$$
where $\underline{J}(R) = \mathrm{tr}\{R_x - R_{xy}R_y^+R_{yx}\}$;
2) *the affine estimate*
$$\hat{x} = \widehat{\mathbf{F}}y = m_x + \left(\widetilde{F} + \hat{H}\right)(y - m_y) \tag{17}$$
is minimax, where $\widetilde{F} = \overline{R}_{xy}\overline{R}_y^+$ is the LSM-operator, $\hat{H}$ is a solution of the auxiliary minimax problem
$$\hat{H} \in \arg\min_{H \in \mathcal{H}} \sup_{K \in \mathcal{K}} J(\widetilde{F} + H, K),$$
$$J(F, R) = \mathrm{tr}\{R_x - 2R_{xy}F^* + FR_yF^*\},$$
$$\mathcal{H} = \left\{H \in \mathbf{R}^{m \times n} : H\overline{R}_y = O\right\};$$
3) *if $\overline{\mathbf{P}}_z$ is the normal distribution with the mean m_z and the covariance $\overline{R}$, then*
$$\Im(\widehat{\mathbf{F}}, \mathbf{P}_z) \le \Im(\widehat{\mathbf{F}}, \overline{\mathbf{P}}_z) \le \Im(\mathbf{F}, \overline{\mathbf{P}}_z) \tag{18}$$
for all $\mathbf{F} \in \mathcal{F},\ \mathbf{P}_z \in \mathcal{P}$.

Relation (18) means that the affine estimating operator $\widehat{\mathbf{F}}$ (17) and the normal distribution $\overline{\mathbf{P}}_z$ form a saddle point of the criterion $\Im(\cdot)$ on the set $\mathcal{F} \times \mathcal{P}$. Hence, the affine estimate (17) is minimax on the class of all measurable estimates, and the normal distribution $\overline{\mathbf{P}}_z$ is least favourable on the class of all distributions with characteristics (16).

5 Numerical algorithms for dual problem solution

It can be shown that the functional $\underline{J}(K)$ is concave and upper semicontinuous on the convex compact set $\mathrm{co}[\mathcal{K}]$. So, the solution $\overline{K}$ can be determined using the standard methods of convex nonlinear programming such as the following method of subgradient projecting:

$$\widetilde{K}_s = K_{s-1} - \gamma_s\nabla_{s-1},\ K_s = \mathrm{Proj}_{\mathrm{co}[\mathcal{K}]}(\widetilde{K}_s),\ s = 1, 2, \ldots,$$

where $\gamma_s > 0 : \sum_{s=1}^{\infty} \gamma_s = \infty,\ \sum_{s=1}^{\infty} \gamma_s^2 < \infty$, and ∇_s is an arbitrary subgradient of the functional $-\underline{J}(\cdot)$ at the point K_s.

For the special case, when the matrices $\Psi K \Psi^*$ are strictly positive for every $K \in \mathcal{K}$, the following iterative procedure seems to be very effective.

1) Choose an arbitrary matrix $K_0 \in \mathrm{co}[\mathcal{K}]$ and put $s = 0$.
2) Define F^s as in (11): $F^s = F_0 + (\Phi - F_0\Psi)P(K_s)$.
3) Solve the linear programming problem: $\widetilde{K}_s \in \arg\max_{K \in \mathrm{co}[\mathcal{K}]} J(F^s, K)$.
4) Calculate $\delta_s = J(F^s, \widetilde{K}_s) - J(F^s, K_s) \geq 0$.
5) If $\delta_s = 0$, then $\overline{K} = K_s$ and the desired solution is obtained, otherwise the process has to be continued.
6) Define the number $\gamma_s \in (0, \overline{\gamma}_s]$ using the condition:

   $$\gamma_s \in \arg\max_{\gamma \in (0, \overline{\gamma}_s]} \underline{J}((1-\gamma)K_s + \gamma \widetilde{K}_s),$$

   where $\overline{\gamma}_s = \max\{\gamma \geq 0 : (1-\gamma)K_s + \gamma\widetilde{K}_s \in \mathrm{co}[\mathcal{K}]\}$.
7) Put $K_{s+1} = (1-\gamma_s)K_s + \gamma_s\widetilde{K}_s$, increase s by 1, and return to Step 2.

It may be shown that the obtained sequence $\{K_s\}$ always converges to the desired solution $\overline{K}$.

6 Minimax identification of infinite-dimensional linear stochastic model

In this section, we shall use the following notation: H_ξ, H_x are separable Hilbert spaces; $\mathcal{L}(H, G)$ is a space of linear bounded operators mapping H into G; $\mathcal{L}_1^+(H)$ is a set of selfadjoint positive-semidefinite nuclear operators mapping a Hilbert space H into itself; $\mathrm{tr}\{A\}$ denotes the trace of $A \in \mathcal{L}_1^+(H)$; A^* is the adjoint operator.

Consider the following observation model:

$$\left\{ \begin{array}{l} x = A\xi, \\ y = B\xi + \varepsilon, \end{array} \right. \tag{19}$$

where $x \in H_x$ is the random element to be estimated given the observation vector $y \in \mathbf{R}^n$. The operators $A \in \mathcal{L}(H_\xi, H_x)$, $B \in \mathcal{L}(H_\xi, \mathbf{R}^n)$ are known. The random element $\xi \in H_\xi$ and the random noise vector $\varepsilon \in \mathbf{R}^n$ are supposed to be independent and centered. Concerning the covariance operators $R_\xi = \mathbf{cov}(\xi, \xi)$, $R_\varepsilon = \mathbf{cov}(\varepsilon, \varepsilon)$, we assume that $R_\xi \in \mathcal{R}_\xi$ and $R_\varepsilon \in \mathcal{R}_\varepsilon$, where the sets $\mathcal{R}_\xi \subset \mathcal{L}_1^+(H_\xi)$, $\mathcal{R}_\varepsilon \subset \mathcal{L}_1^+(\mathbf{R}^n)$ are convex. In addition to this, suppose that $\mathcal{R}_\varepsilon$ is compact.

Let $\mathcal{F} = \mathcal{L}(\mathbf{R}^n, H_x)$ be a class of admissible estimators F. Then for $F \in \mathcal{F}$ the estimate $\widetilde{x} = Fy$ accuracy is measured by the mean-square criterion:

$$J(F, R_\xi, R_\varepsilon) = \mathbf{E}\left\{\|Fy - x\|^2\right\},$$

where $\|\cdot\|$ denotes the corresponding norm in H_x.

As it was assumed above, the estimate $\hat{x} = \hat{F}y$ is called the minimax one if

$$\hat{F} \in \arg\min_{F \in \mathcal{F}} \sup_{R_\xi \in \mathcal{R}_\xi, R_\varepsilon \in \mathcal{R}_\varepsilon} J(F, R_\xi, R_\varepsilon). \tag{20}$$

Denote $\underline{J}(R_\xi, R_\varepsilon) = \inf_{F \in \mathcal{F}} J(F, R_\xi, R_\varepsilon)$, then the optimization problem

$$(\overline{R}_\xi, \overline{R}_\varepsilon) \in \operatorname*{argmax}_{R_\xi \in \mathcal{R}_\xi, R_\varepsilon \in \mathcal{R}_\varepsilon} \underline{J}(R_\xi, R_\varepsilon) \tag{21}$$

may be called *dual* with respect to (20) whenever the following duality relation is fulfilled:

$$\inf_{F \in \mathcal{F}} \sup_{R_\xi \in \mathcal{R}_\xi, R_\varepsilon \in \mathcal{R}_\varepsilon} J(F, R_\xi, R_\varepsilon) = \sup_{R_\xi \in \mathcal{R}_\xi, R_\varepsilon \in \mathcal{R}_\varepsilon} \underline{J}(R_\xi, R_\varepsilon). \tag{22}$$

In this situation, the dual solution $(\overline{R}_\xi, \overline{R}_\varepsilon)$ corresponds to the least favourable case for the model (19) probabilistic characteristics.

The following result provides the explicit expression of the minimax estimate obtained via the solution of the dual optimization problem.

Theorem 6 *1) Suppose that the operator A is compact, the set $\mathcal{R}_\xi$ is bounded in $\mathcal{L}_1^+(H_\xi)$ and closed with respect to the weak operator topology* [10]. *Then*

- *a) there exists a solution $(\overline{R}_\xi, \overline{R}_\varepsilon)$ of the dual optimization problem* (21)*;*
- *b) the duality relation* (22) *is fulfilled;*
- *c) the dual functional $\underline{J}(R_\xi, R_\varepsilon)$ has the form*

  $$\underline{J}(R_\xi, R_\varepsilon) = \mathrm{tr}\{A\left(R_\xi - R_\xi B^* \Lambda^+ B R_\xi\right) A^*\},$$

  where $\Lambda = BR_\xi B^* + R_\varepsilon$.

2) If either the operator $B\overline{R}_\xi B^* + \overline{R}_\varepsilon$ is strictly positive or

$$BR_\xi B^* + R_\varepsilon \leq B\overline{R}_\xi B^* + \overline{R}_\varepsilon, \quad \forall R_\xi \in \mathcal{R}_\xi, R_\varepsilon \in \mathcal{R}_\varepsilon,$$

then

- *a) the estimate $\hat{x} = \hat{F}y$ is minimax, where*

  $$\hat{F} = A\overline{R}_\xi B^*(B\overline{R}_\xi B^* + \overline{R}_\varepsilon)^+; \tag{23}$$

- *b) the guaranteed accuracy of the minimax estimate $\hat{x} = \hat{F}y$ is equal to*

  $$\max_{R_\xi \in \mathcal{R}_\xi, R_\varepsilon \in \mathcal{R}_\varepsilon} J(\hat{F}, R_\xi, R_\varepsilon) = \underline{J}(\overline{R}_\xi, \overline{R}_\varepsilon).$$

7 Example

Consider the following uncertain-stochastic observation model:

$$\begin{cases} x = A\xi, \\ y_k = B_k\xi + \varepsilon_k, \quad k = 1, \ldots, N, \end{cases} \tag{24}$$

where $x \in \mathbf{R}^m$ is the vector to be obtained; $\xi \in \mathbf{R}^r$ is the vector of statistically indeterminate parameters; $y_k \in \mathbf{R}^l$ is the vector of the k-th observation, $\varepsilon_k \in \mathbf{R}^l$ is the corresponding vector of the random observation error; N is the total amount of observations; $A \in \mathbf{R}^{m\times r}$, $B_k \in \mathbf{R}^{l\times r}$ are known nonrandom matrices. Concerning ξ and $\{\varepsilon_k\}$ we have the following *a priori* information:

1) $\mathbf{E}\{\xi\} = 0$, $R_\xi = \mathbf{cov}(\xi, \xi) \in \mathcal{R}_\xi$, where $\mathcal{R}_\xi$ is the set of covariance matrices with the elements $\{R_{ij}\}$ that satisfy the restrictions

$$R_{0ij} \le R_{ij} \le R^0_{ij}; \tag{25}$$

2) $\mathbf{E}\{\varepsilon\} = 0$, $\mathbf{cov}(\varepsilon_k, \varepsilon_j) = 0$ if $k \neq j$, $S_\varepsilon = \mathbf{cov}(\varepsilon_k, \varepsilon_k) \in \mathcal{S}_\varepsilon$, where $\mathcal{S}_\varepsilon$ is the set of covariance matrices with the elements $\{S_{ij}\}$ that satisfy the restrictions

$$S_{0ij} \le S_{ij} \le S^0_{ij}; \tag{26}$$

3) $\mathbf{cov}(\xi, \varepsilon_k) = 0, \quad k = 1, \ldots, N.$

The matrices $R_0 = \{R_{0ij}\}$, $R^0 = \{R^0_{ij}\}$, $S_0 = \{S_{0ij}\}$, $S^0 = \{S^0_{ij}\}$ in (25), (26) are supposed to be known and positive-semidefinite.

Obviously, (24) is a particular case of the general US-model (1). Actually, let us take $n = l \cdot N$, $q = r + n$, $\Phi_0 = O$, $\theta_0 = 0$, $\Psi_0 = O$, $\theta_1 = \mathrm{col}[\xi, \varepsilon_1, \ldots, \varepsilon_N] \in \mathbf{R}^q$, $\Phi = (\ A\ ;\ O\ ) \in \mathbf{R}^{m\times q}$, $\Psi = (\ B\ ;\ I\ ) \in \mathbf{R}^{n\times q}$, where $B = \mathrm{col}[B_1, \ldots, B_N] \in \mathbf{R}^{n\times r}$. With these notations (24) coincides with (1). The restrictions on the covariance matrix R_θ of the vector θ_1 obviously follow from (24), (25).

Let $r = l = m = 3$, $N = 25$, then $n = 75$, $q = 78$. We take the following exact values of A and B_k:

$$A = I, \quad B_k = \cos(0.06k)\begin{pmatrix} 2 & 0.5 & -1 \\ -0.4 & 1 & 0.5 \\ 0.3 & 0.6 & 1.5 \end{pmatrix},$$

$k = 1, \ldots, N$. The restrictions on R_ξ and S_ε are as follows:

$$R_0 = \begin{pmatrix} 1.5 & -0.9 & 0.3 \\ -0.9 & 2.2 & 0.8 \\ 0.3 & 0.8 & 2.2 \end{pmatrix} \quad R^0 = \begin{pmatrix} 2.7 & 0.1 & 1 \\ 0.1 & 3 & 1.6 \\ 1 & 1.6 & 3.5 \end{pmatrix}$$

$$S_0 = \begin{pmatrix} 1.2 & -0.3 & -0.5 \\ -0.3 & 1.4 & 0.7 \\ -0.5 & 0.7 & 1.2 \end{pmatrix} \quad S^0 = \begin{pmatrix} 1.9 & -0.1 & -0.2 \\ -0.1 & 1.8 & 1.2 \\ -0.2 & 1.2 & 1.6 \end{pmatrix}.$$

For the given model (24)–(26) condition 3 of Corollary 1 is fulfilled, hence the LSM-operator $\widetilde{F}$ defines the minimax affine estimate. Using the described above numerical algorithm the dual optimization problem has been solved with the results: $\overline{K} = \begin{pmatrix} \overline{R}_\xi & O \\ O & \overline{R}_\varepsilon \end{pmatrix}$, where

$$\overline{R}_\xi = \begin{pmatrix} 2.7 & -0.9 & 1 \\ -0.9 & 3 & 0.8 \\ 1 & 0.8 & 3.5 \end{pmatrix} \quad \overline{S} = \begin{pmatrix} 1.9 & -0.1 & -0.5 \\ -0.1 & 1.8 & 0.7 \\ -0.5 & 0.7 & 1.6 \end{pmatrix},$$

$\overline{R}_\varepsilon = \mathrm{diag}[\overline{S}, \ldots, \overline{S}]$. Using (11) and $\overline{K}$ we obtain the LSM-operator $\widetilde{F}$ and the corresponding minimax estimate $\hat{x} = \widetilde{F}y$ of x. In the Gaussian case with the worst true values of R_ξ and S_ε, namely $R_\xi = R_0$, $S_\varepsilon = S_0$ we have obtained

$$\hat{x} = (2.86, -1.25, -0.28)^*.$$

The true value of the vector x was $(2.71, -0.98, -0.22)^*$, and the guaranteed value of the criterion $\underline{J}(\overline{K}) = 8.73$.

References

[1] A. Albert, *Regression and the Moor–Penrose pseudoinverse*, Academic Press, New York, 1972.

[2] D.E. Catlin, "Estimation of random states in general linear models", *IEEE Trans. Autom. Control*, vol. AC-36, 2, pp. 348–352, 1991.

[3] A.B. Kurzhanski, "Identification — a theory of guaranteed estimates", *From Data to Model*, Springer-Verlag, Berlin, 1989.

[4] S. Verdú and H.V. Poor, "On minimax robustness: a general approach and applications", *IEEE Trans. Inform. Theory*, vol. IT-30, 2, pp. 328–340, 1984.

[5] L. Ljung, "Issues in system identification", *Report LiTH-ISY-I-125*, Linköping University, Sweden, 1990.

[6] T. Başar, and P. Bernhard, *H_∞-optimal control and related minimax design problem. A game theoretic approach*, Boston, Birkhauser, 1991.

[7] V.N. Soloviov, "Minimax estimation and the least squares method", *Stochastics and Stochastics Reports*, vol. 42, pp. 209–223, 1993.

[8] A.R. Pankov, and K.V. Siemenikhin, "Minimax identification of generalized uncertain-stochastic linear model", *Automation and Remote Control*, vol. 59, 11 (part 2), 1998.

[9] A.R. Pankov, and A.V. Borisov, "Conditionally-minimax filtering and control in infinite dimensional stochastic systems", *Proceeding of the 34-th IEEE CDC, New Orleans, USA*, vol. 1, pp. 87–92, 1995.

[10] N. Dunford, and J.T. Schwartz, *Linear Operators I*, New York: Wiley Interscience, 1967.