MINIMAX PARAMETER ESTIMATION FOR SINGULAR LINEAR MULTIVARIATE MODELS WITH MIXED UNCERTAINTY

Alexei R. Pankov * **Konstantin V. Siemenikhin** **,[1]

* *Probability Theory Dep., Moscow Aviation Inst. (Volokolamskoe sh., 4, Moscow A-80, Russia, 125993)*
arpankov@bk.ru
** siemenkv@mail.ru

Abstract: The problem of minimax estimation is considered for the linear multivariate statistically indeterminate observation model with mixed uncertainty. It is shown that in the regular case the minimax estimate is defined explicitly via the solution of the dual optimization problem. For the singular models, the method of dual optimization is developed by means of using the technique of Tikhonov regularization. Several particular cases which are widely used in practice are also examined.

Keywords: parameter estimation, minimax techniques, uncertain linear systems, duality, convex programming, singular systems, regularization, Kalman filters.

1. INTRODUCTION

In this paper, the problem of minimax estimation by the mean-square criterion is studied for linear multivariate statistically indeterminate models with both stochastic and deterministic uncertain parameters and disturbances. Following (Kurzhanski and Tanaka, 1989), such type of systems will be referred to as ones with *mixed uncertainty*. During the recent period of time, very broad class of statistically indeterminate models has been studied using the minimax approach. Nevertheless, the majority of the models under consideration can be divided into two classes:

a) the models involving only random variables with partially known nondegenerate distributions (Verdú and Poor, 1984; Anan'ev, 1995; Soloviov, 2000);
b) the models with uncertain bounded nonrandom parameters and disturbances (Kurzhanski and Tanaka, 1989; El Ghaoui and Lebret, 1997; Matasov, 1998).

Under stochastic uncertainty, one of the major techniques for constructing minimax estimates is the method of dual optimization. This straightforward and efficient algorithm consists of two steps:

1) to find the *least favorable* joint distribution of the random parameters;
2) to compute the *optimal* estimate designed for the obtained worst-case characteristics.

The necessary and sufficient conditions for the method described above to lead to the minimax estimate are obtained in (Pankov and Siemenikhin, 2000). The standard situation in which such conditions are fulfilled is provided by so-called *regular* models (Verdú and Poor, 1984). However, deterministic models and ones with mixed uncertainty are *singular*, since they contain singular probability distributions.

The main contribution of this paper is to extend the approach of dual optimization to singular lin-

[1] The work is supported in part by Russian Foundation for Basic Research under Grant No 02-01-00361

ear multivariate models with mixed uncertainty. This aim is achieved by means of the Tikhonov regularization techniques (Tikhonov and Arsenin, 1977; Hanke and Hansen, 1993; El Ghaoui and Lebret, 1997; Pankov and Siemenikhin, 2002). Combination of the methods of dual optimization and Tikhonov regularization provides a unified approach to designing efficient algorithms of minimax robust identification for linear multivariate systems with mixed *a priori* uncertainty.

2. STATISTICALLY INDETERMINATE MULTIVARIATE MODEL

The following notation will be used: $\|x\| = \sqrt{x^\top x}$, $x \in \mathbb{R}^m$; O, I are the null and identity matrices, resp.; $\ker[A]$, $\mathrm{im}[A]$, $\mathrm{tr}[A]$, $\sigma[A]$, $\|A\|$, and $\|A\|_2$ are the kernel, image, trace, spectrum, spectral norm, and Euclidean norm of a matrix A, resp.; $\mathrm{diag}[A, B] = \begin{pmatrix} A & O \\ O & B \end{pmatrix}$; A^+ is the Moor–Penrose pseudoinverse; $A > O$ $(A \geq O)$ means that the matrix A is symmetric positively (semi)definite.

Consider the following *statistically indeterminate linear multivariate* model:

$$\begin{cases} x = \Phi_0\rho_0 + \Phi\rho_1, \\ y = \Psi_0\rho_0 + \Psi\rho_1, \end{cases} \tag{1}$$

where $x \in \mathbb{R}^m$ is the vector to be estimated given the observation vector $y \in \mathbb{R}^n$; $\rho = \mathrm{col}[\rho_0, \rho_1]$ is the vector of parameters and disturbances; Φ_0, Φ, Ψ_0, and Ψ are some given matrices.

Concerning the subvector $\rho_0 \in \mathbb{R}^p$ there is no *a priori* information, while $\rho_1 \in \mathbb{R}^q$ is supposed to have partially known moment characteristics: $\mathsf{E}\{\rho_1\} \in \mathcal{M}$, $\mathsf{cov}\{\rho_1\} \in \mathcal{R}$. The set $\mathcal{M}$ is compact and centrally symmetric and $\mathcal{R}$ is a compact set of symmetric positive-semidefinite matrices.

Let P_ρ denote the probability distribution of the vector ρ. The stated above assumptions mean that $\mathsf{P}_\rho \in \mathcal{P}$, where $\mathcal{P}$ is the *set of feasible distributions*:

$$\mathcal{P} = \{\mathsf{P}_\rho\colon\ \rho = \mathrm{col}[\rho_0, \rho_1],\ \mathsf{E}\{\|\rho\|^2\} < \infty,\ \mathsf{E}\{\rho_1\} \in \mathcal{M},\ \mathsf{cov}\{\rho_1\} \in \mathcal{R}\}. \tag{2}$$

Note that partitioning the model parameters into the structural and disturbing ones can be done explicitly by an appropriate choosing of the matrices Φ_0, Φ, Ψ_0, Ψ. Furthermore, all probability distributions are allowed to be singular. This makes it possible not to separate the random and deterministic variables. Thus, model (1) is the most general one and the majority of linear observation models can be considered as particular cases of (1).

3. MINIMAX ESTIMATION PROBLEM

Consider a linear estimate $\widetilde{x} = Fy$, $F \in \mathcal{F}$, of x given the vector y, where $\mathcal{F} \subseteq \mathbb{R}^{m\times n}$ is some prespecified set of estimators. Then, $\widetilde{x} = Fy$ will be referred to as an *admissible* estimate. The accuracy of $\widetilde{x}$ is measured by the m.s.e. criterion:

$$\mathfrak{D}(F, \mathsf{P}_\rho) = \mathsf{E}\{\|Fy - x\|^2\}, \quad \mathsf{P}_\rho \in \mathcal{P}. \tag{3}$$

Definition 1. An estimate $\hat{x} = \hat{F}y$ is called *minimax* if

$$\hat{F} \in \widehat{\mathcal{F}} = \arg\min_{F\in\mathcal{F}} \sup_{\mathsf{P}_\rho\in\mathcal{P}} \mathfrak{D}(F, \mathsf{P}_\rho). \tag{4}$$

The *optimal guaranteed value* of the m.s.e. criterion is equal to $\hat{J} = \inf_{F\in\mathcal{F}} \sup_{\mathsf{P}_\rho\in\mathcal{P}} \mathfrak{D}(F, \mathsf{P}_\rho)$.

Since (3) depends only on $\mathsf{E}\{\rho\}$ and $\mathsf{cov}\{\rho\}$, (4) can be reduced to the following minimax problem:

$$\hat{F} \in \widehat{\mathcal{F}} = \arg\min_{F\in\mathcal{F}_0} \max_{K\in\mathcal{K}} J(F, K), \tag{5}$$

where the auxiliary functional

$$J(F, K) = \sup_{\mathsf{P}_\rho}\{\mathfrak{D}(F, \mathsf{P}_\rho)\colon\ \mathsf{E}\{\rho_1\rho_1^\top\} = K\} \tag{6}$$

will henceforth be used instead of $\mathfrak{D}(\cdot)$,

$$\mathcal{K} = \mathrm{co}\{K\colon\ K = \mu_\rho\mu_\rho^\top + R_\rho,\ \mu_\rho \in \mathcal{M},\ R_\rho \in \mathcal{R}\}$$

is the convex hull of the set of matrices $\mathsf{E}\{\rho_1\rho_1^\top\}$ such that $\mathsf{P}_\rho \in \mathcal{P}$, and

$$\mathcal{F}_0 = \{F \in \mathcal{F}\colon\ F\Psi_0 = \Phi_0\} \tag{7}$$

is the class of admissible estimators, which are unbiased w.r.t. ρ_0. In what follows, the set $\mathcal{F}_0$ is supposed to be nonempty, convex, and closed.

Problems (4) and (5) are equivalent due to

$$\sup_{\mathsf{P}_\rho\in\mathcal{P}} \mathfrak{D}(F, \mathsf{P}_\rho) = \sup_{K\in\mathcal{K}} J(F, K) \quad \forall\, F \in \mathbb{R}^{m\times n},$$

where $J(F, K) = \mathrm{tr}\big[(F\Psi - \Phi)K(F\Psi - \Phi)^\top\big]$ if $F\Psi_0 = \Phi_0$ and $J(F, K) = +\infty$ otherwise.

Specify the regular and singular situations.

Definition 2. The statistically indeterminate model (1) and the corresponding minimax estimation problem (4) are *regular* if

$$\Psi K \Psi^\top > O \quad \forall\, K \in \mathcal{K}, \tag{8}$$

otherwise (1), (4) are *singular*.

Condition (8) means that any feasible covariance $\mathsf{cov}\{y\} = \Psi\,\mathsf{cov}\{\rho_1\}\Psi^\top$ of the observation vector is a nonsingular matrix.

As it is known (Verdú and Poor, 1984), for singular observation models there may exist several minimax estimates. The last makes it reasonable to introduce the next concept.

Definition 3. The minimax estimate $\hat{x}^{(o)} = \hat{F}^{(o)}y$ is called *normal* if $\|\hat{F}^{(o)}\|_2 \leq \|\hat{F}\|_2 \ \forall \hat{F} \in \widehat{\mathcal{F}}$.

In other words, $\hat{F}^{(o)}$ is of the minimal Euclidean norm over all minimax estimators.

The following theorem describes the main features of the minimax optimization problem (5).

Theorem 1. Assume $\exists\, K_0 \in \mathcal{K}$: $\Psi K_0 \Psi^\top > O$.

1) Then, the set $\widehat{\mathcal{F}}$ of minimax estimators is nonempty, convex, and compact. Moreover

$$\|\hat{F}\|_2^2 \leq \frac{\left(\sqrt{\mathrm{tr}[\Phi K_0 \Phi^\top]} + \sqrt{\hat{J}}\right)^2}{\min \sigma[\Psi K_0 \Psi^\top]} \quad \forall \hat{F} \in \widehat{\mathcal{F}}. \quad (9)$$

2) There exists a unique normal minimax estimator $\hat{F}^{(o)}$.
3) The following duality relation holds:

$$\hat{J} = \min_{F \in \mathcal{F}_0} \max_{K \in \mathcal{K}} J(F,K) = \max_{K \in \mathcal{K}} \inf_{F \in \mathcal{F}_0} J(F,K). \quad (10)$$

4) Under the regularity condition (8), the minimax estimator $\hat{F}$ is uniquely determined and can be found as follows:

$$\{\hat{F}\} = \arg \min_{F \in \mathcal{F}_0} J(F, \hat{K}), \quad (11)$$

where $\hat{K}$ is an arbitrary solution to the problem

$$\hat{K} \in \arg \max_{K \in \mathcal{K}} \underline{J}(K), \quad \underline{J}(K) = \inf_{F \in \mathcal{F}_0} J(F,K). \quad (12)$$

The next definition is motivated by equality (10).

Definition 4. The maximin problem (12) is called *dual* w.r.t. the minimax one (4).

Thus, $\hat{K}$ describes a least favorable combination of the moment characteristics involved in (1). Nevertheless, in general, a distribution $\mathsf{P}_\rho \in \mathcal{P}$ such that $\hat{K} = \Psi\, \mathsf{E}\{\rho_1 \rho_1^T\}\, \Psi^\top$ may not exist.

The third assertion of Theorem 1 means that $\hat{F}$ is a minimax estimator and $\hat{K}$ is a dual solution iff the pair $(\hat{F}, \hat{K})$ forms a saddle point for the game $(J, \mathcal{F}_0, \mathcal{K})$: for every $F \in \mathcal{F}_0$ and $K \in \mathcal{K}$ $J(\hat{F}, K) \leq J(\hat{F}, \hat{K}) \leq J(F, \hat{K})$.

The last part of Theorem 1 describes the way of finding minimax estimates using the dual optimization approach. According to the last, the minimax estimate is sought as a solution of the *linear-optimal estimation problem* (11) by the m.s.e. criterion with the least favorable moment characteristics (12). For standard classes of estimators, the optimal estimate $\hat{x} = \hat{F}y$ and the dual functional $\underline{J}(\cdot)$ have the explicit representation.

So, in the regular case, the method of dual optimization can be directly applied to finding the minimax estimate. In the singular case, algorithm (11)–(12) yields the minimax estimate if $\Psi \hat{K} \Psi^\top > O$ or $\mathrm{im}[\Psi K \Psi^\top] \subseteq \mathrm{im}[\Psi \hat{K} \Psi^\top]$ for any $K \in \mathcal{K}$ (Pankov and Siemenikhin, 2000).

It should be noted that if the uncertainty set $\mathcal{K}$ contains a maximal element $\overline{K}$ (i.e., $K \leq \overline{K}$ for all $K \in \mathcal{K}$), then the solution of (12) is trivial: $\hat{K} = \overline{K}$, whence (11) is the minimax estimator.

Even if the least favorable matrix $\hat{K}$ can be obtained asymptotically, the method of minimax estimation based on the dual optimization turns to be valid and possesses the robust property. Furthermore, the deviation of the approximate estimator from the minimax one can be majorized by the computation error of the dual solution.

Theorem 2. Under the regularity condition (8), given a sequence $\{K^s\} \subset \mathcal{K}$, the estimators $\{F^s\}$ are supposed to be defined as follows:

$$\{F^s\} = \arg \min_{F \in \mathcal{F}_0} J(F, K^s). \quad (13)$$

Then $\|F^s - \hat{F}\|_2^2 \leq (\hat{J} - \underline{J}(K^s)) / \min_{K \in \mathcal{K}} \sigma[\Psi K \Psi^\top]$ for all s and the sequence $\{F^s\}$ converges to the minimax estimator $\hat{F}$ whenever $\lim_{s \to \infty} \underline{J}(K^s) = \hat{J}$.

Thus, for finding the dual solution one may use any numerical procedure for which the convergence w.r.t. $\underline{J}(\cdot)$ is fulfilled.

The iterative algorithm presented below possesses the desired property.

Algorithm 1. 1) Take $K^0 \in \mathcal{K}$ and put $s = 0$.
2) Solve the quadratic minimization problem (13).
3) Solve the following problem of linear programming: $\widetilde{K}^s \in \arg \max_{K \in \mathcal{K}} J(F^s, K)$.
4) If $\max_{K \in \mathcal{K}} J(F^s, K) = J(F^s, K^s)$, put $\hat{F} = F^s$ and $\hat{K} = K^s$ and terminate the iterative process.
5) Solve the one-dimensional maximization problem $\gamma^s \in \arg \max_{\gamma \in [0,1]} \underline{J}((1-\gamma)K^s + \gamma \widetilde{K}^s)$.
6) Put $K^{s+1} = (1-\gamma^s)K^s + \gamma^s \widetilde{K}^s$, increase s by 1, and go to step 2.

The convergence of Algorithm 1 is stated in the following theorem.

Theorem 3. Let the regularity condition (8) be fulfilled and the sequences $\{F^s\}$, $\{K^s\}$ be generated by Algorithm 1. Then, $\{F^s\}$ converges to the minimax estimator $\hat{F}$ and $\{K^s\}$ converges to the set $\widehat{\mathcal{K}} = \arg \max_{K \in \mathcal{K}} \underline{J}(K)$ of dual solutions, i.e.,

$$\lim_{s \to \infty} \|F^s - \hat{F}\|_2 = 0, \qquad \lim_{s \to \infty} \inf_{\hat{K} \in \widehat{\mathcal{K}}} \|K^s - \hat{K}\| = 0.$$

Furthermore, the convergence holds also w.r.t. the functionals $\overline{J}(F) = \max_{K \in \mathcal{K}} J(F,K)$ and $\underline{J}(K)$: $\lim_{s \to \infty} \overline{J}(F^s) = \lim_{s \to \infty} \underline{J}(K^s) = \hat{J}$.

4. MINIMAX ESTIMATION FOR SINGULAR MODELS

In this section, the Tikhonov regularization method will be applied to the singular case of the minimax problem (5).

Introduce the regularized criterion

$$J^\varepsilon(F,K) = J(F,K) + \varepsilon\,\|F\|_2^2, \quad \varepsilon > 0. \quad (14)$$

Then, the *regularized minimax problem* has the form

$$\hat F^\varepsilon \in \arg\min_{F\in\mathcal{F}_0}\max_{K\in\mathcal{K}} J^\varepsilon(F,K), \quad (15)$$

where $\hat F^\varepsilon$ is said to be the *regularized minimax estimator*. The optimal guaranteed value of $J^\varepsilon(\cdot)$ is equal to $\hat J^\varepsilon = \min\limits_{F\in\mathcal{F}_0}\max\limits_{K\in\mathcal{K}} J^\varepsilon(F,K)$.

Note that problem (15) is regular, since it corresponds to the following regular observation model:

$$\begin{cases} x = \Phi_0\rho_0 + \Phi\rho_1, \\ y^\varepsilon = \Psi_0\rho_0 + \Psi\rho_1 + \varepsilon\eta, \end{cases} \quad (16)$$

where $\mathsf{P}_\rho \in \mathcal{P}$ and the random vector η is supposed to be normalized ($\mathsf{E}\{\eta\} = 0$, $\mathsf{cov}\{\eta\} = I$) and independent of ρ. Indeed, any feasible covariance of the observation vector is nonsingular:

$$\mathsf{cov}\{y^\varepsilon\} = \Psi\,\mathsf{cov}\{\rho_1\}\Psi^\top + \varepsilon I > O. \quad (17)$$

Since problem (15) is regular, one can use the method of dual optimization. To this end, introduce the functional $\underline{J}^\varepsilon(K) = \inf\limits_{F\in\mathcal{F}_0} J^\varepsilon(F,K)$ and consider the *regularized dual problem*

$$\hat K^\varepsilon \in \arg\max_{K\in\mathcal{K}} \underline{J}^\varepsilon(K). \quad (18)$$

The following result explains how to obtain a minimax estimate using the regularization technique.

Theorem 4. Under the conditions of Theorem 1, the following assertions are valid:

1) The regularized minimax estimator $\hat F^\varepsilon$ can be found in the form $\{\hat F^\varepsilon\} = \arg\min\limits_{F\in\mathcal{F}_0} J^\varepsilon(F,\hat K^\varepsilon)$, where $\hat K^\varepsilon$ is an arbitrary solution of (18).
2) $\{\hat F^\varepsilon\}$ converges to the normal minimax estimator $\hat F^{(o)}$: $\|\hat F^\varepsilon - \hat F^{(o)}\|_2 \to 0$ as $\varepsilon \downarrow 0$.
3) The optimal guaranteed values of the original and regularized criteria satisfy the following relation: $\hat J \le \hat J^\varepsilon \le \hat J + \varepsilon\|\hat F^{(o)}\|^2\ \forall\,\varepsilon > 0$.

Note that the minimax solution $\hat F^\varepsilon$ can be obtained using the iterative algorithm described in Section 3. However, if the convergence provided by Algorithm 1 is not finite, the error of computing $\hat F^\varepsilon$ may be significant. This may lead to the unstable behavior of the sequence $\{\hat F^\varepsilon\}$. In order to overcome this obstacle one can also use the method of Tikhonov regularization of ill-posed optimization problems (Tikhonov and Arsenin, 1977).

The theorem below shows how to compute the minimax estimator in a stable manner.

Theorem 5. Under the conditions of Theorem 1, suppose the following assumptions to be fulfilled:

a) the parameter ε_ν tends to zero as $\nu \to \infty$;
b) $\{K_\nu\}$ is a given sequence such that $K_\nu \in \mathcal{K}$ and $\delta_\nu = \hat J^{\varepsilon_\nu} - \underline{J}^{\varepsilon_\nu}(K_\nu) \to 0$ as $\nu \to \infty$;
c) $\{F_\nu\} = \arg\min\limits_{F\in\mathcal{F}_0} J^{\varepsilon_\nu}(F,K_\nu)$, $\nu = 1,2,\ldots$

Then,

1) $\lim\limits_{\nu\to\infty}\inf\limits_{\hat F\in\widehat{\mathcal{F}}}\|F_\nu - \hat F\|_2 = 0$ if $\sup\limits_\nu \delta_\nu/\varepsilon_\nu^3 < \infty$;

2) $\lim\limits_{\nu\to\infty}\|F_\nu - \hat F^{(o)}\|_2 = 0$ if $\lim\limits_{\nu\to\infty}\delta_\nu/\varepsilon_\nu^3 = 0$.

In the both cases $\lim\limits_{\nu\to\infty}\overline{J}(F_\nu) = \hat J$.

The first part of Theorem 5 describes the situation, when one can claim the convergence of $\{F_\nu\}$ to the set $\widehat{\mathcal{F}}$ of minimax estimators. The second part provides the sufficient condition for $\{F_\nu\}$ to converge to the normal minimax estimator $\hat F^{(o)}$.

5. BASIC PARTICULAR CASES

Suppose that there are no *a priori* constraints on estimators, i.e., any linear estimate $\widetilde{x} = Fy$, $F \in \mathbb{R}^{m\times n}$, is assumed to be admissible. Thus,

$$\mathcal{F} = \mathbb{R}^{m\times n} \quad \text{and} \quad \mathcal{F}_0 = \{F\colon F\Psi_0 = \Phi_0\}. \quad (19)$$

The assumption $\mathcal{F}_0 \ne \varnothing$ is equivalent to the identifiability condition:

$$\Phi_0 = \Phi_0\Psi_0^+\Psi_0, \quad \text{or} \quad \ker[\Psi_0] \subseteq \ker[\Phi_0]. \quad (20)$$

Consider the linear-optimal estimation problem

$$\arg\min_{F\in\mathcal{F}_0} J(F,K), \quad K \in \mathcal{K}. \quad (21)$$

Under the notation $K_x = \Phi K\Phi^\top$, $K_{xy} = \Phi K\Psi^\top$, $K_y = \Psi K\Psi^\top$, one can claim that if $K_y > O$, then

$$\widetilde{F}(K_{xy},K_y) = K_{xy}K_y^{-1} + (\Phi_0 - K_{xy}K_y^{-1}\Psi_0)\left(\Psi_0^\top K_y^{-1}\Psi_0\right)^+\Psi_0^\top K_y^{-1}$$

is the unique solution of (21) and

$$\underline{J}(K_x,K_{xy},K_y) = \operatorname{tr}\big[K_x - K_{xy}K_y^{-1}K_{xy}^\top + (\Phi_0 - K_{xy}K_y^{-1}\Psi_0)\left(\Psi_0^\top K_y^{-1}\Psi_0\right)^+ \times (\Phi_0 - K_{xy}K_y^{-1}\Psi_0)^\top\big]$$

is the minimum of (21).

Theorem 6. Under the previous notation, suppose that (19) and (20) are fulfilled.

1) Then, the sequence of estimators

$$\hat{F}^\varepsilon = \widetilde{F}(\hat{K}^\varepsilon_{xy}, \hat{K}^\varepsilon_y + \varepsilon I) \tag{22}$$

converges to the normal minimax estimator $\hat{F}^{(o)}$ as $\varepsilon \downarrow 0$ if $\hat{K}^\varepsilon_{xy} = \Phi \hat{K}^\varepsilon \Psi^\top$, $\hat{K}^\varepsilon_y = \Psi \hat{K}^\varepsilon \Psi^\top$, and

$$\hat{K}^\varepsilon \in \arg\max_{K \in \mathcal{K}} \underline{J}(K_x, K_{xy}, K_y + \varepsilon I). \tag{23}$$

2) Let the regularity condition (8) be fulfilled. Then, $\hat{F} = \widetilde{F}(\hat{K}_{xy}, \hat{K}_y)$ is the minimax estimator if $\hat{K}_{xy} = \Phi \hat{K} \Psi^\top$, $\hat{K}_y = \Psi \hat{K} \Psi^\top$, and

$$\hat{K} \in \arg\max_{K \in \mathcal{K}} \underline{J}(K_x, K_{xy}, K_y). \tag{24}$$

Example 5.1. (*Regression with unbounded parameters*). Consider the following linear regression model:

$$\begin{cases} x = A\theta, \\ y = B\theta + \xi, \end{cases} \tag{25}$$

where $\theta \in \mathbb{R}^p$ is the *a priori* unbounded nonrandom vector and $\xi \in \mathbb{R}^n$ is the random observation noise with zero mean and covariance $R \in \mathcal{R}_\xi$, where $\mathcal{R}_\xi$ is a compact set of positive semidefinite matrices. A and B are some given matrices.

Model (25) is widely used in applications and is a particular case of the general model (1), since the probability distribution P_ρ of the vector $\rho = \mathrm{col}[\theta, \xi]$ belongs to the set

$$\mathcal{P} = \{\mathsf{P}_\rho \colon \rho = \mathrm{col}[\theta, \xi],\ \mathsf{E}\{\theta\} = \theta \in \mathbb{R}^p,\ \mathsf{E}\{\xi\} = 0,\ \mathsf{cov}\{\xi\} \in \mathcal{R}_\xi\}. \tag{26}$$

Corollary 1. Let $\mathcal{F} = \mathbb{R}^{m \times n}$ and $AB^+B = A$. Denote $\widetilde{F}(R) = A\left(B^\top R^{-1} B\right)^+ B^\top R^{-1}$ and $\underline{J}(R) = \mathrm{tr}\left[A\left(B^\top R^{-1} B\right)^+ A^\top\right]$.

1) Then, the sequence of $\widetilde{F}(\hat{R}^\varepsilon + \varepsilon I)$ converges to the normal minimax estimator as $\varepsilon \downarrow 0$ if $\hat{R}^\varepsilon \in \arg\max_{R \in \mathcal{R}_\xi} \underline{J}(R + \varepsilon I)$.

2) Let $R > O$ for all $R \in \mathcal{R}_\xi$. Then, the estimator $\widetilde{F}(\hat{R})$ is minimax if $\hat{R} \in \arg\max_{R \in \mathcal{R}_\xi} \underline{J}(R)$.

Since $\widetilde{F}(\hat{R})$ is a least-squares estimator, the result stated in Corollary 1 can be treated as the minimax version of the Gauss–Markov theorem.

Example 5.2. (*Regression with bounded parameters*). Here, the observation model (25) is considered under the same assumptions except for the conditions on θ: θ is now supposed to be *a priori* bounded, i.e., $\theta \in \Theta$, where Θ is a given centrally symmetric compact subset of $\mathbb{R}^p$. Hence,

$$\mathcal{P} = \{\mathsf{P}_\rho \colon \rho = \mathrm{col}[\theta, \xi],\ \mathsf{E}\{\theta\} = \theta \in \Theta,\ \mathsf{E}\{\xi\} = 0,\ \mathsf{cov}\{\xi\} \in \mathcal{R}_\xi\} \tag{27}$$

is the class of feasible distributions.

Corollary 2. Let (27) and $\mathcal{F}_0 = \mathcal{F} = \mathbb{R}^{m \times n}$ hold. Denote $\widetilde{F}(T, R) = ATB^\top(BTB^\top + R)^{-1}$, and $\underline{J}(T, R) = \mathrm{tr}\left[A(T - TB^\top(BTB^\top + R)^{-1}BT)A^\top\right]$, where $T \in \mathcal{T}$, $R \in \mathcal{R}_\xi$, and $\mathcal{T} = \mathrm{co}\{\theta\theta^\top \colon \theta \in \Theta\}$.

1) The sequence of estimators $\widetilde{F}(\hat{T}^\varepsilon, \hat{R}^\varepsilon + \varepsilon I)$ converges to the normal minimax estimator as $\varepsilon \downarrow 0$ if $(\hat{T}^\varepsilon, \hat{R}^\varepsilon) \in \arg\max_{T \in \mathcal{T}, R \in \mathcal{R}_\xi} \underline{J}(T, R + \varepsilon I)$.

2) Let $R > O$ for all $R \in \mathcal{R}_\xi$. The estimator $\widetilde{F}(\hat{T}, \hat{R})$ is minimax if $(\hat{T}, \hat{R}) \in \arg\max_{T \in \mathcal{T}, R \in \mathcal{R}_\xi} \underline{J}(T, R)$.

Example 5.3. (*Purely stochastic observation model*). Consider the problem of minimax estimation under the following assumptions:

$$\mathsf{E}\{\rho\} = 0 \quad \text{and} \quad \mathsf{cov}\{\rho\} \in \mathcal{R}, \tag{28}$$

where $\rho = \mathrm{col}[x, y]$ and $\mathcal{R}$ is a given convex compact set of positive-semidefinite matrices $R = \begin{pmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{pmatrix}$. Then, the observation model is called *purely stochastic*, since it does not contain any indeterminate nonrandom variables.

Corollary 3. Suppose that $\mathcal{F} = \mathbb{R}^{m \times n}$ and the set $\mathcal{P}$ is defined by condition (28). Denote $\underline{J}(R_x, R_{xy}, R_y) = \mathrm{tr}\left[R_x - R_{xy} R_y^{-1} R_{xy}^\top\right]$.

1) The sequence of matrices $\hat{R}^\varepsilon_{xy}(\hat{R}^\varepsilon_y + \varepsilon I)^{-1}$ converges to the normal minimax estimator as $\varepsilon \downarrow 0$ if $\hat{R}^\varepsilon \in \arg\max_{R \in \mathcal{R}} \underline{J}(R_x, R_{xy}, R_y + \varepsilon I)$.

2) Let $R_y > O$ $\forall R \in \mathcal{R}$. The estimator $\hat{R}_{xy} \hat{R}_y^{-1}$ is minimax if $\hat{R} \in \arg\max_{R \in \mathcal{R}} \underline{J}(R_x, R_{xy}, R_y)$.

The result of Corollary 3 can be strengthened as follows. The linear-minimax estimate $\hat{F}$ and the Gaussian distribution $\widehat{\mathsf{P}}_\rho$ with zero mean and covariance $\hat{R}$ form a saddle point for the game $(\mathfrak{D}, \mathcal{B}, \mathcal{P})$, i.e., $\mathfrak{D}(\hat{F}, \mathsf{P}_\rho) \le \mathfrak{D}(\hat{F}, \widehat{\mathsf{P}}_\rho) \le \mathfrak{D}(F, \widehat{\mathsf{P}}_\rho)$ for all $F \in \mathcal{B}$ and $\mathsf{P}_\rho \in \mathcal{P}$, where $\mathcal{B}$ is the class of all Borelean transformation.

The last makes it possible to claim that the linear-minimax estimate is also minimax over the class of all nonlinear transformations and the Gaussian distribution is least favorable if its covariance is a solution of the dual problem. So, the results above form the regularized minimax version of the theorem on normal correlation.

6. MINIMAX FILTERING

Consider the problem of minimax filtering for the following discrete-time statistically indeterminate dynamic system (Verdú and Poor, 1984):

$$\begin{cases} x_t = \varphi_t \theta_t,\ \theta_t = a_t \theta_{t-1} + b_t \xi_t,\ \theta_0 = \gamma, \\ y_t = A_t \theta_t + B_t \xi_t,\ t = 1, \dots, N, \end{cases} \tag{29}$$

where $\{x_t\}$ is the $m \times 1$ process to be estimated given the $n \times 1$ observation process $\{y_t\}$; θ_t is the $p \times 1$ state vector; $\{\xi_t\}$ is the zero-mean $q \times 1$ white noise uncorrelated with the zero-mean initial state γ; a_t, b_t, A_t, B_t, and φ_t are some known matrices.

The uncertainty in covariances $R_\gamma = \mathsf{cov}\{\gamma\}$ and $R_t = \mathsf{cov}\{\xi_t\}$ is described by the two alternatives:

$$\text{a)} \quad R_\gamma \in \mathcal{R}_\gamma, \quad R_1 = \ldots = R_N \in \mathcal{R}_\xi, \tag{30}$$

$$\text{b)} \quad R_\gamma \in \mathcal{R}_\gamma, \quad R_t \in \mathcal{R}_\xi, \quad t = 1, \ldots, N, \tag{31}$$

where $\mathcal{R}_\gamma$ and $\mathcal{R}_\xi$ are prespecified convex and compact sets of positive-semidefinite matrices.

Cases a) and b) define the set $\mathcal{R}$ of feasible covariances $R = \mathrm{diag}[R_\gamma, R_1, \ldots, R_N]$ of the vector $\rho = \mathrm{col}[\gamma, \xi_1, \ldots, \xi_N]$.

Given the observation process y, the linear estimate $\widetilde{x} = Fy$ is an admissible filter iff $\widetilde{x}$ is non-anticipative, namely, $\widetilde{x}_t = \sum_{s=1}^{t} F_{ts} y_s$, $F_{ts} \in \mathbb{R}^{m \times n}$. In other words, the vector x_t is to be estimated given the observations $y^{(t)} = \mathrm{col}[y_1, \ldots, y_t]$ that are available up to the current moment t. Thus, the set $\mathcal{F}$ of admissible estimators consists of the block matrices $F = \{F_{ts}\}_{t,s=1,\ldots,N}$ of the lower triangle form: $F_{ts} = O$ for all $t < s$.

As before, the accuracy of estimates $\widetilde{x} = Fy$, $F \in \mathcal{F}$, is measured by the m.s.e. criterion of the integral form: $\mathfrak{D}(F, \mathsf{P}_\rho) = \mathsf{E}\{\|Fy - x\|^2\} = \sum_{t=1}^{N} \mathsf{E}\{\|\widetilde{x}_t - x_t\|^2\}$. The last does not lead to loss of generality, since the necessary weight matrices for each value of the process can be took into account by an appropriate modification of the matrices $\{\varphi_t\}$.

Definition 5. A filter $\hat{x} = \hat{F}y$ is called minimax if

$$\hat{F} \in \arg\min_{F \in \mathcal{F}} \sup_{\mathsf{P}_\rho \in \mathcal{P}} \mathfrak{D}(F, \mathsf{P}_\rho), \tag{32}$$

where $\mathcal{F}$ is the class of linear non-anticipative estimators and $\mathcal{P}$ is defined by (29)–(31).

Since the finite horizon discrete-time linear system (29) can be reduced to the multivariate observation model (1), the problem of minimax filtering (32) can be treated as a particular case of the general problem of minimax estimation (4). Therefore, all the results obtained in Sections 3, 4 can be applied to designing the minimax filter.

Note that the regularity condition (33) takes the following form: $\forall\, t = 1, \ldots, N$

$$S_t(R) = [A_t b_t + B_t] R_t [A_t b_t + B_t]^\top > O. \tag{33}$$

In particular, the last is fulfilled whenever the observation noise $w_t = B_t \xi_t$ is nondegenerate and uncorrelated with the state noise $v_t = b_t \xi_t$.

If (33) does not hold, then the observation model (29) is singular. In this case, one should consider the regularized version of (29) with the observation process $y_t^\varepsilon = y_t + \varepsilon \eta_t$, where $\{\eta_t\}$ is a standard white noise (i.e., $\mathsf{E}\{\eta_t\} = 0$, $\mathsf{cov}\{\eta_t\} = I$) uncorrelated with γ and $\{\xi_t\}$.

Theorem 7. Given the observation process (29) and $R = \mathrm{diag}[R_\gamma, R_1, \ldots, R_N]$, suppose that the filter $\{\widetilde{x}_t^\varepsilon(R)\}$ and the error covariance $\{P_t^\varepsilon(R)\}$ satisfy the Kalman equations with coefficients as ones for the regularized observation system. Then, the sequence of regularized Kalman filters $\widetilde{x}^\varepsilon(\hat{R}^\varepsilon)$ converges to the normal minimax filter as $\varepsilon \downarrow 0$ if

$$\hat{R}^\varepsilon \in \arg\max_{R \in \mathcal{R}} \sum_{t=1}^{N} \mathrm{tr}\big[\varphi_t P_t^\varepsilon(R) \varphi_t^\top\big].$$

Note that the normal minimax filter has the recursive structure even if it does not coincide with any Kalman filter.

REFERENCES

Anan'ev, B.I. (1995). Minimax estimation of statistically uncertain systems under the choice of a feedback parameter. *J. of Math. Syst., Estimation and Control*, **5**(2), 1–17.

El Ghaoui, L. and H. Lebret (1997). Robust solutions to least-squares problems with uncertain data. *SIAM J. Matrix Anal. Appl.*, **18**(4), 1035–1064.

Hanke, M. and N.J. Hansen (1993). Regularization methods for large-scale problems. *Surveys on Math. for Industry*, **3**, 253–315.

Kurzhanski, A.B. and M. Tanaka (1989). *On a Unified Framework for Deterministic and Stochastic Treatment of Identification Problems.* IIASA, Laxenburg.

Matasov, A.I. (1998). *Estimators for Uncertain Dynamic Systems.* Kluwer, Dordrecht.

Pankov, A.R. and K.V. Siemenikhin (2000). The parametric identification methods for many-dimensional linear models in the presence of *a priori* uncertainty. *Automation and Remote Control*, **61**(5. Part 1), 785–800.

Pankov, A.R. and K.V. Siemenikhin (2002). Minimax estimation in singular uncertain stochastic models. *Automation and Remote Control*, **63**(9), 1410–1425.

Soloviov, V.N. (2000). Towards the theory of minimax-Bayesian estimation. *Theory of Probab. and Its Applications*, **44**(4), 739–754.

Tikhonov, A.N. and V.Ya. Arsenin (1977). *Solutions of Ill-Posed Problems.* Wiley, New York.

Verdú, S. and H.V. Poor (1984). Minimax linear observers and regulators for stochastic systems with uncertain second order statistics. *IEEE TAC*, **29**(6), 499–511.