Journal of Mathematical Systems, Estimation, and Control
Vol. 5, No. 2, 1995, pp. 1–17

Minimax Estimation of Statistically Uncertain Systems Under the Choice of a Feedback Parameter*

B.I. Anan'ev†

Abstract

This paper considers minimax problems for estimating a function of an unobservable variable from the measurement of another one statistically connected with the first. It is supposed that the statistically uncertain parameter presenting in the joint distribution may depend on the realizations of the above variables. This supposition extends the range of the uncertain parameter and leads to estimates that are simply realizable particularly in multistage and dynamical systems. The case when the parameter depends only on the observable variable is discussed separately. The proposed schemes are applied to the filtering of statistically uncertain Markov sequences and to the investigation of discrete and continuous time Kalman filtering with uncertain parameters.

Key words: minimax estimates, statistically uncertain systems, feedback parameter

AMS Subject Classifications: 93E05, 93E10, 93E11, 62C20

1 Introduction

In many problems of statistics, filtering, and control [1-5] it is necessary to estimate some function $h(x)$ of the unobservable variable x on the basis of observations of another value y. In this connection, the joint distribution $P_\theta(dx, dy)$ of the random variables x and y may depend on an unknown but bounded parameter $\theta \in \Theta$. One possible approach [6-10] to the solution

*Received December 22, 1993; received in final form December 28, 1993. Summary appeared in Volume 5, Number 2, 1995.

†This work was supported by the Russian Fund of Fundamental Researches under Project 93-011-16041.

of the problem consists in the choice of a certain class Δ of estimates $\delta(y)$, for example, those linear in y or nonlinear but complying with some conditions of measurability and integrability, and the determination of an optimal minimax estimate:

$$\sup_{\theta} E_{\theta}\|h(x) - \delta(y)\|^2 \to \min_{\delta(\cdot)} . \tag{1.1}$$

Under certain conditions of regularity the optimal minimax solution to the problem exists and may be constructed [8]. The method of the solution includes the randomization of the set Θ, the conversion of (1.1) to a supremum over all probability measures on Θ, and the application of various minimax theorems. The above approach when applied to the dynamical systems meets considerable difficulties connected with increasing dimension of the defining correlations. The latter is typical not only for the nonlinear case but also for the linear one when only linear estimates are used [9].

There are other approaches to estimation under conditions of statistical uncertaity. In paper [11] a recurrent procedure of estimation has been constructed. Paper [12] deals with stochastic approximation procedures for the guaranteed filtering scheme. We also refer to paper [13] as well.

In problem (1.1) the parameter θ is in no way connected with the realizations of random variables x and y. It stays constant (but unknown) while the random experiment is repeated to obtain values x and y. This paper offers some statements of the problem of minimax estimation other than (1.1) when the parameter θ is formed on the basis of realizations of x, y or only of y. These statements expand the possibilities of the parameter θ and they are more rough with respect to the statistician who selects the estimate $\delta(y)$. However these new problems turn out to be more convenient when applied to estimating the states and parameters of dynamical systems. This assertion is justified in the present work by examples of filtering of Markov sequences and by the investigation of extending the Kalman-Bucy scheme to uncertain parameters.

The importance of minimax estimation methods can be motivated by the fact that in many practical engineering problems the detailed information on the dynamics of the process and its statistics turns out to be unavailable [14 ,16]. The approach presented in this paper assures numerical robustness for respective approximation schemes. Other results concerned with robustness of estimation can be found in paper [15].

2 The One-stage Minimax Estimation Problems

Consider a static situation. Let X, Y be separable metrizable spaces, Θ be a Polish space. These may be interpreted as the state, measurement, and parameter spaces respectively. In the spaces X and Y, Borel σ-finite measures $\mu(\mathrm{dx})$ and $\nu(\mathrm{dx})$ are given. The joint distribution of variables x and y, i.e. Borelian probability measure $P_\theta(dx, dy)$ depending on uncertain parameter $\theta \in \Theta$, is supposed to be absolutely continuous with respect to the product measure $\mu \times \nu$. According to the Radon-Nikodim theorem, one has the formula

$$P_\theta(dx, dy) = p(\theta, x, y)\mu(dx)\nu(dy). \tag{2.1}$$

The non-negative function $p(\theta, x, y)$ is assumed as satisfying one of the following assumptions:

(i) $p(\theta, x, y)$ is continuous in θ for every fixed x, y, Borelian in x, y for every fixed θ, and bounded from above, i.e. it is a Caratheodory type function;

(ii) $p(\theta, x, y)$ is upper semi-continuous in θ, x, y, bounded from above, and the set Θ is compact.

Further on the symbol , stands for the set of all Borelian mappings $\gamma : XY \to \Theta$, where XY is the Cartesian product of the respective spaces. The word "estimate" will be used for any Borelian mapping $\delta : Y \to R^d$, where d is a fixed integer. Consider the functional

$$J(f) = \sup_{\gamma \in ,} \int_{XY} \|f(x, y\|^2 p(\gamma, x, y)\mu(dx)\nu(dy) \tag{2.2}$$

for every Borelian function $f : XY \to R^d$. From now on, symbol $\|\cdot\|$ is the Euclidean norm in R^d. The functional (2.2) will take finite non-negative values or $+\infty$ and is correctly defined for every f in view of the measurability of the functional composition [17, Theorem 6.1]. Let us formulate the following:

Problem A: Find an estimate δ^0 which minimizes the deviation between a function $h(x)$ and the estimate according to functional (2.2), i.e.

$$J(h - \delta^0) = \min_{\delta(\cdot)} J(h - \delta).$$

It is rather evident that functional $J(h-\delta)$ is not finite for all estimates δ. Therefore, we introduce

Definition 2.1: *An estimate δ is said to be admissible for Problem A if $J(h - \delta) < \infty$.*

The set of all admissible estimates will be denoted by Δ. In order to ensure the non-emptines of set Δ, assume

Condition 2.1: $J(h) < \infty$.

Under Condition 2.1 we have $0 \in \Delta$. The following assertion gives a direct characteristic of the set Δ.

Lemma 2.1: *Under Condition 2.1 the estimate δ is admissible if and only if $J(\delta) < \infty$.*

Proof: If $\delta \in \Delta$ then $J(\delta) = J(\delta - h + h) \leq 2(J(\delta - h) + J(h))$. This fact follows from the elementary inequality $\|x + y\|^2 \leq 2(\|x\|^2 + \|y\|^2)$. The last inequality also implies the inverse statement. Q.E.D.

Furthermore, it follows from Lemma 2.1 that the set Δ is a linear space.

Lemma 2.2: *Let the Borelian mapping $f : XY \to R^d$ satisfy $J(f) < \infty$. Then we have*

$$J(f) = \int \|f(x,y)\|^2 \overline{p}(x,y)\mu(dx)\nu(dy), \tag{2.3}$$

where

$$\overline{p}(x,y) = \sup_{\theta \in \Theta} p(\theta, x, y). \tag{2.4}$$

Proof: First note that function (2.4) is Borelian in x, y due to either the continuity of $p(\theta, x, y)$ in θ and the separability of the space Θ or the semi-continuity of that function [18]. If condition (i) holds, consider the set-valued mapping

$$T_\epsilon(x,y) = \{\theta : p(\theta, x, y) > (1-\epsilon)\overline{p}(x,y)\}, \qquad \epsilon > 0, \tag{2.5}$$

when $\overline{p}(x,y) > 0$. If $\overline{p}(x,y) = 0$, we set $T_\epsilon(x,y) = \Theta$. According to [17, Theorem 6.2], the mapping (2.5) is Borelian in x, y and hence $\overline{T}_\epsilon(x,y)$ (closure in Θ) is weakly measurable. Therefore, there exists a Borelian selector $\theta_\epsilon(x,y) \in \overline{T}_\epsilon(x,y)$ [17, Theorem 5.1] for which we have $p(\theta_\epsilon(x,y), x, y) \geq (1-\epsilon)\overline{p}(x,y)$. As ϵ is arbitrary, it follows from this that inequality (2.3) is true. If condition (ii) holds then there exists a Borelian function $\theta(x,y)$ which delivers the maximum in (2.4) [18, Lemma 7.20]. Q.E.D.

Combining the assertions of lemmas 2.1 and 2.2, one may state that

$$\Delta = L_2^d(Y, \mathcal{B}_y, Q), \tag{2.6}$$

where $\mathcal{B}_y$ is Borelian σ-algebra in Y. Q is defined by

$$Q(dy) = q(y)\nu(dy), \; q(y) = \int \overline{p}(x,y)\mu(dx). \tag{2.7}$$

In other words, under Condition 2.1 the set of admissible estimates for Problem A represents the Hilbert space of square integrable Borelian mappings with respect to the measure Q defined by (2.7). Admissible estimates δ_1, δ_2 which differ on the set of Q-measure zero may be considered identical since $J(h - \delta_1) = J(h - \delta_2)$. Generally, the function $q(y)$ in (2.7) may not be ν-a.e. finite, and measure Q is not bound to be σ-finite on Y. However, the latter is σ-finite on the set $\tilde{Y} = \{y : q(y) < \infty\}$, and also any admissible estimate $\delta(y) = 0$ on the set $Y \backslash \tilde{Y}$ (ν-a.e.) by virtue of equality (2.3). A solution to Problem A may be formulated in the following way.

Theorem 2.1 *Under condition (2.2), there exists an optimal estimate* δ^0 *for Problem A. It is (mod-*Q*)-unique and defined by the equality*

$$\delta^0(y) = \int h(x)\overline{p}(x,y)\mu(dx)/q(y), \tag{2.8}$$

where function $q(y)$ *is given by (2.7).*

The proof of this theorem arises from the convexity of the functional $J(h-\delta)$ and its Frechet differentiability [19]. As a property of the estimate (2.8), note the inequality

$$J(\delta^0) < J(h) \tag{2.9}$$

which holds if $h(x) \neq \text{const}$. That can be directly derived with the help of the Cauchy-Schwarz inequality.

If parameter θ is known precisely and the set Θ consists of one point, then the estimate (2.8) is the well-known Bayesian that coincides with the conditional expectation of the random value $h(x)$ with respect to y. When using the terminology of game theory, it is possible to say that in Problem A the first player or statistician when choosing his estimate $\delta(y)$ is discriminated because the second player chooses his strategy $\theta = \gamma(x,y)$ on the basis of the realizations of both y and x. To equalize the chances we restrict the possibilities of the second player by strategies $\theta = \gamma(y)$. Namely, consider the set $\Gamma_1 \subset \Gamma$ of all Borelian mappings $\gamma : Y \to \Theta$. Instead of (2.2) the functional

$$J_1(f) = \sup_{\gamma \in \Gamma_1} \int_{XY} \|f(x,y)\|^2 p(\gamma,x,y)\mu(dx)\nu(dy) \tag{2.10}$$

will be used. We have $J_1(f) \leq J(f)$ for all f.

In what follows one has to introduce an additional

Condition 2.2: The function $\overline{p}(x,y)$ given by (2.4) is integrable with respect to the measure $\mu \times \nu$.

This condition is necessary for the correctness of the passage to the limit under the integral sign. For functional (2.10) consider the following:

Problem B: One has to find an estimate δ_0 which minimizes the deviation according to functional (2.10), i.e.

$$J_1(h-\delta_0) = \min_{\delta(\cdot)} J_1(h-\delta).$$

A wider than Δ class of admissible estimates is introduced as follows:

Definition 2.2: *An estimate δ is said to be admissible for Problem B if $J_1(h-\delta) < \infty$.*

The set of all admissible estimates for Problem B will be denoted by Δ_1. Under Condition 2.1 this class is non-void by virtue of inclusion $0 \in \Delta \subset \Delta_1$. The next lemma is analogous to 2.1.

Lemma 2.3: *Under conditions 2.1 and 2.2 the estimate δ is admissible for the Problem B ($\delta \in \Delta_1$) if and only if $J_1(\delta) < \infty$.*

This lemma implies that set Δ_1 in just the same way as Δ is a linear space. To clarify its structure we prove

Lemma 2.4: *Let conditions 2.1 and 2.2 hold and estimate δ be admissible for Problem B. Then*

$$J_1(h-\delta) = \int j(h-\delta(y);y)\nu(dy), \tag{2.11}$$

$$j(h-\delta;y) = \sup_\theta \int \|h(x)-\delta\|^2 p(\theta,x,y)\mu(dx). \tag{2.12}$$

Proof: The integrand in (2.12) is dominated (ν-a.e.) by the function $2(\|h(x)\|^2 + \|\delta\|^2)\cdot \overline{p}(x,y)$ that is integrable with respect to measure $\mu(dx)$ by virtue of suppositions. By Lebesgue's convergence theorem the integral (2.12) represents a continuous in θ function if condition (i) holds. Hence function $j(h-\delta(y);y)$ is Borelian and ν-a.e. finite. Consider the set-valued mapping

$$T_\epsilon(y) = \{\theta : \int \|h(x)-\delta(y)\|^2 p(\theta,x,y)\mu(dx) > (1-\epsilon) j(h-\delta(y);y)\}.$$

Continuing the arguments as in Lemma 2.2, we complete the proof. If condition (ii) holds then by [18, Proposition 7.31] one may assert that the integral in (2.12) represents an upper semi-continuous in t, y and δ function. Further, the same reasoning as in Lemma 2.2 leads us to the aim. Q.E.D.

Corollary 2.1: *Under the conditions of Lemma 2.4 we have*

$$J_1(\delta) = \int \|\delta(y)\|^2 q_1(y)\nu(dy), \tag{2.13}$$

$$q_1(y) = \sup_\theta \int p(\theta,x,y)\mu(dx). \tag{2.14}$$

In order to prove this it is sufficient to set $h(x) = 0$ in (2.11), (2.12). It follows from Lemmas 2.3 and 2.4 and Corollary 2.1 that

$$\Delta_1 = L_2^d(Y, \mathcal{B}_y, Q_1), \tag{2.15}$$

where the measure

$$Q_1(dy) = q_1(y)\nu(dy) \tag{2.16}$$

is completely finite. The admissible estimates δ_1, δ_2 which differ on a set of Q_1-measure zero may be identified since $J_1(h-\delta_1) = J_1(h-\delta_2)$.

Let us pass to the solution of Problem B.

Theorem 2.2 *Let conditions 2.1 and 2.2 be fulfilled. Then the estimate $\delta_0(y)$ solves Problem B if and only if the equality*

$$j(h-\delta_0(y);y) = \min_{\delta\in R^d} j(h-\delta;y) \tag{2.17}$$

holds ν-a.e. This estimate $\delta_0(y)$ exists.

Proof: The equality (2.12) may be written out in more detail as

$$j(h-\delta;y) = \sup_{\theta}\{a(\theta,y) - 2\delta'\, b(\theta,y) + \|\delta\|^2 c(\theta,y)\}, \tag{2.18}$$

where symbol ' means the transposition, and functions

$$\begin{aligned} a(\theta,y) &= \int \|h(x)\|^2 p(\theta,x,y)\mu(dx), \\ b(\theta,y) &= \int h(x)p(\theta,x,y)\mu(dx), \\ c(\theta,y) &= \int p(\theta,x,y)\mu(dx), \end{aligned} \tag{2.19}$$

are ν-a.e. finite and continuous in θ by virtue of the assumptions of the theorem if (i) holds. Under assumption (ii) the functions (2.19) are upper semi-continuous in θ, y [18, Proposition 7.33]. Furthermore, these are bounded in θ for every y (ν-a.e). Therefore, the function $j(h-\delta;y)$ is ν-a.e. continuous in δ, and minimum in (2.17) will be a Borelian function in y. By the implicit function theorem [17, Theorem 7.2] there exists a Borelian function $\delta_0(y)$ satisfying the equality (2.17) ν-a.e. It is clear that this function solves Problem B. Conversely, if $\delta_0(y)$ is a solution to Problem B, then in view of (2.11) and the elementary properties of the Lebesgue integral the equality (2.17) should be fulfilled. Q.E.D.

Using equality (2.17), one may give a more exact characteristic of the optimal solution to Problem B. We introduce the set (Borelian in y)

$$A(y) = \{a(\theta,y), b(\theta,y), c(\theta,y) : \theta\in\Theta\} \subset R^{d+2}. \tag{2.20}$$

The set (2.20) is bounded for $y \in Y$ (ν-a.e.). Therefore, the closure $\overline{A}(y)$ will be compact ν-a.e. Applying the minimax theorem from [22], the following proposition is derived.

Theorem 2.3 *Under the assumptions of Theorem 2.2 the estimate $\delta_0(y)$ solves Problem B if and only if*

$$\delta_0(y) = (\sum_{i=1}^{d+1} \lambda_i^0 b_i^0)/(\sum_{i=1}^{d+1} \lambda_i^0 c_i^0), \tag{2.21}$$

where the coefficients $\lambda_i^0 > 0$, $\sum_{i=1}^{d+1} \lambda_i^0 = 1$ are defined by the probability measure

$$\lambda_0 = \sum_{i=1}^{d+1} \lambda_i^0 \epsilon_i \qquad (\epsilon_i - \textit{Dirac measure}) \tag{2.22}$$

on the compact set $\overline{A}(y)$. The pair $\lambda_0, \delta_0(y)$ forms a saddle point for the minimax problem of minimization in δ of expression (2.18).

It should be noted that unit Dirac measures ϵ_i on $\overline{A}(y)$ together with coefficients λ_i^0 are defined from the solution to the extremal problem

$$\sum_{i=1}^{d+1} \lambda_i a_i - \| \sum_{i=1}^{d+1} \lambda_i b_i \|^2 / (\sum_{i=1}^{d+1} \lambda_i c_i) \rightarrow \max_{\lambda}. \tag{2.23}$$

By the implicit function theorem [17, Theorem 7.1] the coefficients λ_i^0, a_i^0, b_i^0, c_i^0 of the extremal measure (2.22) may be chosen Borelian in y.

In the further sections, we consider some applications of the above results for the one-stage problems to the filtering of dynamical processes.

3 Minimax Filtering of Markov Sequences in Metric Spaces

Let the spaces X, Y and the measures $\mu(dx), \nu(dy)$ be the same as in the previous section. Given in XY is a Markov sequence with transient probabilities

$$P_{\theta_k}(dx_k, dy_k \mid x_{k-1}, y_{k-1}) = p_k(\theta_k, x_k, y_k \mid x_{k-1}, y_{k-1}) \mu(dx_k) \nu(dy_k),$$
$$k = 1, 2, \ldots, \qquad P_{\theta_0}(dx_0, dy_0) = p_0(\theta_0, x_0, y_0) \mu(dx_0) \nu(dy_0). \tag{3.1}$$

In equalities (3.1) the functions are supposed to be complying with conditions of type (i) or (ii). The spaces Θ_k are assumed to be Polish, and with p_k semi-continuous these spaces are compact. Further we use the notations

$$y^k = (y_0, \ldots, y_k), \qquad \theta^k = (\theta_0, \ldots, \theta_k) \tag{3.2}$$

for the sequences of measurements and parameters up to the number k. For every discrete instant k one has to estimate the function $h(x_k)$.

Consider a problem of type A:

$$\sup_{\gamma^k} \int_{X^k Y^k} \|h(x_k) - \delta(y^k)\|^2 p_k(\theta_k, x_k, y_k \mid x_{k-1}, y_{k-1}) \mu(dx_k) \nu(dy_k) \ldots p_0(\theta_0, x_0, y_0) \mu(dx_0) \nu(dy_0) \to \min_{\delta(\cdot)} . \tag{3.3}$$

In order to achive the maximal result for the second player it is sufficient to use strategies $\theta_i = \gamma_i(x_i, y_i, x_{i-1}, y_{i-1})$ for all $i = 0, 1, \ldots, k$. Let us introduce the functions

$$q_k(x_k, y_k, x_{k-1}, y_{k-1}) = \sup_{\theta_k} p_k(\theta_k, x_k, y_k \mid x_{k-1}, y_{k-1}),$$
$$r_k(x^k, y^k) = q_k(x_k, y_k, x_{k-1}, y_{k-1}) \cdots q_0(x_0, y_0), \tag{3.4}$$

and measures

$$\mu^k(dx^k) = \mu(dx_k) \cdots \mu(dx_0), \qquad \nu^k(dy^k) = \nu(dy_k) \cdots \nu(dy_0). \tag{3.5}$$

Then due to the results of the previous section, the function

$$\delta^0(y^k) = \int h(x_k) r_k(x^k, y^k) \mu^k(dx^k) / \int r_k(x^k, y^k) \mu^k(dx^k) \tag{3.6}$$

will serve as an optimal estimate if

$$\int \|h(x_k)\|^2 r_k(x^k, y^k) \mu^k(dx^k) \nu^k(dy^k) < \infty \tag{3.7}$$

for all $k \geq 0$.

In formula (3.6) the integration is carried out with respect to some conditional probability measure posessing a density concerned $\mu(dx_k)$. This density may be given recursively. In fact, we set

$$q_0(x_0 \mid y^0) = q_0(x_0, y_0) / \int q_0(x_0, y_0) \mu(dx_0).$$

Putting in (3.6) $k = 1$ and dividing the numerator and the denominator by the denominator of the last equality, we have

$$q_1(x_1 \mid y^1) = \int q_1(x_1, y_1, x_0, y_0) q_0(x_0 \mid y^0) \mu(dx_0) / \int \int q_1(x_1, y_1, x_0, y_0) q_0(x_0 \mid y^0) \mu(dx_0) \mu(dx_1).$$

Continuing the process, one can define the densities for all $k > 1$

$$q_k(x_k \mid y^k) = \int q_k(x_k, y_k, x_{k-1}, y_{k-1}) q_{k-1}(x_{k-1} \mid y^{k-1}) \mu(dx_{k-1}) / \int \int q_k(x_k, y_k, x_{k-1}, y_{k-1}) q_{k-1}(x_{k-1} \mid y^{k-1}) \mu(dx_{k-1}) \mu(dx_k). \tag{3.8}$$

By using densities (3.8), the equality (3.6) may be rewritten as

$$\delta^0(y^k) = \int h(x_k) q_k(x_k \mid y^k) \mu(dx_k). \tag{3.9}$$

Summarizing the reasonings, we come to the conclusion.

Theorem 3.1 *Let condition (3.7) be fulfilled for all $k = 0, 1, \ldots$. Then the optimal estimate in problem (3.3) is defined by formula (3.9) where the conditional densities are determined according to (3.8).*

A problem of type B will be considered below for the linear-Gaussian case.

4 The Discrete-time Kalman Filter with Uncertain Values of the Moments of the Disturbances

Consider a multi-stage system

$$\begin{aligned} x_k &= A_k x_{k-1} + v_k + \xi_k, \qquad x_{-1} = 0, \\ y_k &= C_k x_k + w_k + \eta_k, \end{aligned} \tag{4.1}$$

where $x_k \in R^n$, $y_k \in R^m$; ξ_k and η_k are independent sequences of Gaussian "white noises" with zero means and covariance matrices

$$cov(\xi_k, \xi_k) = Q_k, \qquad cov(\eta_k, \eta_k) = R_k \tag{4.2}$$

which, along with vectors v_k, w_k, are supposed to be uncertain. Assume that the uncertain values belong to compact sets. Minimax filtering problems for system (4.1) have been investigated in many papers [8-11,13]. However in this work, we will apply the approaches stated above.

If a problem of type A is formulated for system (4.1), then our reasoning corresponds to the previous section. Here the transient densities from (3.1) with respect to Lebesgue measures have the form

$$\begin{aligned} p_k(\theta_k, x_k, y_k \mid x_{k-1}, y_{k-1}) &= N_n(x_k - A_k x_{k-1} - v_k; Q_k) \\ \cdot N_m(y_k - w_k - C_k x_k; R_k), & \\ k = 0, 1, \ldots, \qquad \theta_k &= (v_k, w_k, R_k, Q_k), \end{aligned} \tag{4.3}$$

where $N_n(x; Q)$ is the density of a Gaussian distribution in R^n with zero mean and covariance matrix Q. In general, this density is not continuous in x, Q, but it will be upper semi-continuous. Note that recursive defined densities (3.8) made of (4.3) will not be Gaussian yet.

Now let us consider a problem of type B for system (4.1) in more detail. Here for stage k the functions (2.19) are taken to be as follows:

$$\begin{aligned} a(\theta^k, y^k) &= \int \|x_k\|^2 p(x_k \mid y^k) dx_k p(y^k), \\ b(\theta^k, y^k) &= \int x_k p(x_k \mid y^k) dx_k p(y^k), \\ c(\theta^k, y^k) &= p(y^k). \end{aligned} \tag{4.4}$$

It is known from the Kalman theory of filtering that the conditional density of the phase state and the unconditional density of the signal are

$$\begin{aligned} p(x_k \mid y^k) &= N_n(x_k - \hat{x}_k; P_k), \\ p(y^k) &= N_m(y_k - \overline{y}_k; \Omega_k) \cdots N_m(y_0 - \overline{y}_0; \Omega_0). \end{aligned} \tag{4.5}$$

In formulas (4.5) we have

$$\begin{aligned} \hat{x}_k = E(x_k \mid y^k), &\qquad P_k = cov(x_k, x_k \mid y^k), \\ \Omega_k = cov(y_k, y_k \mid y^k), &\qquad \overline{y}_k = E(y_k \mid y^{k-1}), \end{aligned} \tag{4.6}$$

where the matrices are non-random. It is obvious that $p(y^k)$ may be given recursively:

$$p(y^k) = N_m(y_k - \overline{y}_k; \Omega_k) p(y^{k-1}). \tag{4.7}$$

For given parameters θ^k the values (4.6) are also determined recursively by the well-known Kalman filter equations [1] if $R_k > 0$.

Due to (4.4)-(4.6), the extremal problem (2.23) will have to be solved with

$$a(\theta^k, y^k) = tr(P_k + \hat{x}_k \hat{x}_k'), \qquad b(\theta^k, y^k) = \hat{x}_k, \tag{4.8}$$

and the estimate (2.21) will be equal to

$$\delta_0(y^k) = (\sum_{i=1}^{n+1} \lambda_i^0 \hat{x}_k p(y^k))/(\sum_{i=1}^{n+1} \lambda_i^0 p(y^k)). \tag{4.9}$$

The problem (2.23) with (4.8) appears to be considerably more easy than in the case when the parameters θ_k do not use the knowledge of y^k [8].

5 The Continuous-time Kalman-Bucy Filter With Uncertain Moments of the Disturbances

Let the linear stochastic differential equations

$$\begin{aligned} dx &= (A(t)x + v(t))dt + \sigma_1(t)d\xi, \\ dy &= (C(t)x + w(t))dt + \sigma(t)d\eta, \\ x(t_0) &= \sigma_0 x_0 + K_0 \overline{x}_0, \qquad y(t_0) = 0 \end{aligned} \tag{5.1}$$

be given. Here the unobservable vector $x \in R^n$, the observable one $y \in R^m$; A, C are continuous matrices; the uncertain vector and matrix functions v, w, σ_1, σ are measurable in t; $\xi(t) \in R^k$, $\eta(t) \in R^m$ are standard Wiener processes with zero means and

$$cov(d\xi, d\xi) = I_k dt, \qquad cov(d\eta, d\eta) = I_m dt;$$

x_0 is Gaussian vector with zero mean and unit covariance matrix. It is supposed that the vector x_0 and the processes ξ, η are mutually independent. Functional parameters v, w, σ_1, σ in each moment of time belong to compact sets:

$$v \in V,\ w \in W,\ \sigma_1 \in \Sigma_1,\ \sigma \in \Sigma. \tag{5.2}$$

Besides, we assume that

$$\sigma(t)\sigma'(t) \geq \epsilon I_m. \tag{5.3}$$

According to the Kalman-Bucy theory, the parameters of the conditional Gaussian distribution of $x(t)$ with measurements $y_0^t = \{y(\tau), 0 \leq \tau \leq t\}$ may be found from equation

$$d\hat{x} = (A(t)\hat{x} + v(t))dt + PC'\sigma'^{-1} d\zeta, \qquad \hat{x}(0) = K_0 \overline{x}_0,$$

where $d\zeta = \sigma^{-1}(dy - (C\hat{x} + w)dt)$ is the standard Wiener innovation process in R^m, $P(t)$ is a solution of the Riccati type differential equation [2,5]. The process $\zeta(t)$ generates a Wiener probability measure $\mu_\zeta(d\zeta)$ in the space $(C_t^m, \mathcal{B}_t^m)$ of continuous m-dimensional functions $\zeta(\tau)$, $\zeta(0) = 0$, given on $[0, t]$, where the Borelian σ-algebra $\mathcal{B}_t^m$ is produced by cylindrical sets [5]. In view of the fact that the observable process $y(t)$ has a stochastic differential $dy = (C\hat{x} + w)dt + \sigma d\zeta$, the Wiener probability measure $\mu_y(dy)$ generated by the observable process is absolutely continuous with respect to standard measure $\mu_\zeta(d\zeta)$, and the equalities

$$\begin{aligned} \mu_y(d\zeta) &= \alpha(t, \zeta)\mu_\zeta(d\zeta), \qquad (5.4)\\ \alpha(t, \zeta) &= exp[\int_0^t (C\hat{x} + w)'\sigma d\zeta - (1/2)\int_0^t (C\hat{x} + w)'\sigma\sigma'(C\hat{x} + w)dt] \end{aligned}$$

will be true [5].

A problem of type B may be formulated as follows. It is necessary to find an estimate $\delta(y_0^t)$ for vector $x(t)$ so that

$$\max_{\theta \in \Theta} \int \|x - \delta(y_0^t)\|^2 N_n(x - \hat{x}; P(t))\alpha(t, y_0^t)\mu_\zeta(dy_0^t)dx \to \min_{\delta(\cdot)}. \tag{5.5}$$

Here the integration is carried out with respect to the joint distribution of y_0^t and $x(t)$. Let the parameter set Θ be the Cartesian product of functional sets of $v(\cdot)$,$w(\cdot)$,$\sigma_1(\cdot)$,$\sigma(\cdot)$ from L_2 complying with constraints (5.2),(5.3).

Then Θ is a metric compactum and the densities $N_n(x-\hat{x};P(t))$, $\alpha(t,y_0^t)$ are upper semi-continuous in θ. In accordance with the conclusions of Section 2 we have an optimal estimate for Problem B

$$\delta_0(y_0^t) = (\sum_{i=1}^{n+1} \lambda_i^0 \hat{x}(t)\alpha(t,y_0^t))/(\sum_{i=1}^{n+1} \lambda_i^0 \alpha(t,y_0^t)). \tag{5.6}$$

where numbers λ_i^0 and the parameters for the measure can be found as in (2.23).

6 Examples

1. Let the one-dimensional equations

$$x = v + \xi, \qquad y = x + \eta \tag{6.1}$$

be given where v is an uncertain value satisfying the inequality $\mid v \mid \leq 1$. The disturbances ξ and η are assumed to be uniformly distributed and independent with the same density function

$$p_1(x) = \begin{cases} 1/2, & \mid x \mid \leq 1, \\ 0, & \mid x \mid > 1. \end{cases}$$

Then the joint distribution density for x and y equals

$$p(v,x,y) = p(y \mid x)p(x) = p_1(y-x)p_1(x-v). \tag{6.2}$$

For the problem of type A, calculating the maximum of $p_1(x-v)$ over v, we have

$$\max_v p_1(x-v) = \overline{p}_1(x) = \begin{cases} 1/2, & \mid x \mid \leq 2, \\ 0, & \mid x \mid > 2. \end{cases} \tag{6.3}$$

Therefore, $\overline{p}(x,y) = p_1(y-x)\overline{p}_1(x)$. It follows from formula (2.8) that

$$\begin{aligned} \delta^0(y) &= \int x\overline{p}(x,y)dx / \int \overline{p}(x,y)dx \\ &= (2 \wedge (1+y) + (-2) \vee (y-1))/2, \end{aligned} \tag{6.4}$$

where $x \wedge y = min(x,y)$, $x \vee y = max(x,y)$. It is interesting to note that the estimate (6.4) coincides with the minimax estimate from the deterministic theory [4]. But, this fact is not valid if the dimension of $x > 1$. For a solution to the problem of type B we write the functions (2.19):

$$\begin{aligned} c(v,y) &= (2- \mid v-y \mid)/4; \\ b(v,y) &= c(v,y)(y+v)/2, \\ a(v,y) &= c(v,y)(y^2+v^2+1+yv- \mid y-v \mid)/3. \end{aligned} \tag{6.5}$$

Further one needs to compose expression (2.23) and to find a maximum. According to formula (2.21), we have

$$\delta_0(y) = y/2 + (\alpha_1^0 c(v_1, y)v_1 + \alpha_2^0 c(v_2, y)v_2) \\ /(\alpha_1^0 c(v_1, y) + \alpha_2^0 c(v_2, y))/2. \qquad (6.6)$$

In general, the estimates (6.4) and (6.6) are non-linear in y and do not coincide with each other.

2. Consider a one-dimensional multi-stage system

$$x_k = a_k x_{k-1}, \qquad y_k = x_k + \eta_k,\ k \geq 1, \qquad (6.7)$$

where $a_k \in [1, \overline{a}]$, $\overline{a} > 1$; $\mid x_0 \mid \leq \mu$; the disturbances η_k have a uniform distribution $p_1(x)$ (6.1) and are independent. As x_k has no probability distribution we can act in two ways. First, it is possible to add a "small" noise ξ_k in (6.7) for constructing approximations. Secondly, one may consider a_k and x_0 as parameters for distribution of y_k. Let's choose the first option. We take random values ξ_k with the distribution

$$p_\epsilon(x) = \begin{cases} (2\epsilon)^{-1}, & \mid x \mid \leq \epsilon, \\ 0, & \mid x \mid > \epsilon. \end{cases}$$

as the "small" noise. Using considerations of section 3 and according to (3.4), we obtain

$$\begin{aligned} q_\epsilon(x_k, x_{k-1}) &= \max_a p_\epsilon(x_k - a x_{k-1}), \\ q_\epsilon(x_0) &= \max_{|\alpha| \leq \mu} p_\epsilon(x_0 - \alpha). \end{aligned} \qquad (6.8)$$

Thereupon by formulas (3.8) one can determine the estimates (3.9).

Examining a_k and x_0 as parameters, we need to solve a problem

$$\max_{a_1, \ldots, a_k, x_0} \mid x_k - \delta \mid^2 p_1(y_0 - x_0) \cdots p_1(y_k - x_k) \to \min_\delta. \qquad (6.9)$$

Therefore, for the first stage we have

$$\delta_0(y_0) = ((-\mu) \vee (y_0 - 1) + (y_0 + 1) \wedge \mu)/2,$$

and for the second one the estimate is equal to

$$\delta_0(y^1) = ((-\overline{a}\mu) \vee \overline{a}(y_0 - 1) \vee (y_1 - 1) + \overline{a}\mu \wedge \overline{a}(y_0 + 1) \wedge (y_1 + 1))/2$$

and so on.

3. Given are continuous-time one-dimensional equations

$$\dot{x} = v_i, \qquad dy = x dt + d\eta \qquad (6.10)$$

where v_i, $i = 1, 2$ represents the uncertain deterministic functional parameter, η is a standard Wiener process, x_0 is the Gaussian standard value independent of η. From formulas of section 5 we have

$$\begin{aligned} d\hat{x} &= v_i dt + d\zeta/(t+1), \qquad \hat{x}(0) = 0, \\ d\zeta &= dy - \hat{x}dt, \qquad \alpha(t,\zeta) = exp(\int_0^t \hat{x} d\zeta - (1/2)\int_0^t \hat{x}^2 ds). \end{aligned}$$

It is well-known [5] that values like $\alpha(t, \zeta)$ are square-integrable martingales satisfying the equation $d\alpha = 1 + \alpha\hat{x}d\zeta$, $\alpha(0) = 1$. Therefore, the minimax problem (5.5) has the form

$$\max_{i=1,2}[(1/(t+1) + \hat{x}_i^2)\alpha_i - 2\hat{x}_i\delta\alpha_i + \delta^2\alpha_i] \to \min_\delta .$$

7 Conclusions

In this paper we have given two new approaches to minimax estimating problems for statistically uncertain systems. These approaches use the supposition that uncertain parameters may "choose" their values on the basis of the knowledge of measurements and the phase vector. It has been demonstrated that the arising minimax problems are finite dimensional as opposed to those that come from statistical Wald theory [6-9]. Moreover, for dynamic systems, the dimension of the problems under consideration does not grow with time t or with index k tending to infinity.

References

[1] R.E. Kalman. A new approach to linear filtering and prediction problems, *J. Basic Engr. ASME Trans.* **82D** (1960), 35-45.

[2] R.E. Kalman and R. Bucy. New results in linear prediction and filtering theory, *J. Basic Engr. ASME Trans.* **83D** (1961), 95-100.

[3] N.N. Krasovskii. *Theory of Control of Motion.* Moscow: Nauka, 1968.

[4] A.B. Kurzhanskii. *Control and Observation Under Conditions of Uncertainty.* Moscow: Nauka, 1977.

[5] R.Sh. Liptser and A.N. Shiryayev. *Statistics of Random Processes* **1,2**. New York: Springer-Verlag, 1978.

[6] A. Wald. *Statistical Decision Functions.* New York: J. Wiley & Sons, 1950.

[7] C.J. Martin and M. Mintz. Robust filtering and prediction for linear systems with uncertain dynamics: a game-theoretic approach, *IEEE Transactions on Automatic Control* **28** (1983), 888-896.

[8] B.I. Anan'ev. On minimax state estimates for multistage statistically uncertain systems, *Problems of Control and Information Theory* **18(1)** (1989), 27-41.

[9] B.I. Anan'ev. Minimax linear filtering of multistage processes with uncertain distributions of random disturbances, *Automation and Remote Control* **10** (1993), in Russian.

[10] H.V. Poor and D.P. Loose. Minimax state estimation for linear stochastic systems with noise uncertainty, *IEEE Transactions on Automatic Control* **26** (1981), 902-906.

[11] I.Ya. Katz and A.B. Kurzhanskii. Minimax estimation for multistage systems, *Dokl. Akad. Nauk SSSR* **221(3)** (1975), 535-538, in Russian.

[12] A.B. Kurzhanskii. Stochastic filtering approximation of estimation problems for systems with uncertainty, *Stochastics* **23** (1988), 104-130.

[13] B.I. Anan'ev and A.B. Kurzhanskii. The nonlinear filtering problem for a multistage system with statistical uncertainty, *Second IFAC Symposium on Stochastic Control*, Vilnius, USSR, Pt. 1, (1986), 205-210.

[14] B.Ts. Bakhshiyan, R.R. Nazirov and P.Ye. El'yasberg. *Determination and Correction of Motion.* Moscow: Nauka, 1980.

[15] B.T. Polyak and Ya.Z. Tsypkin. Robust identification, *Automatica* **16(1)** (1980), 53-63.

[16] I.A. Boguslavskii. *Applied Problems of Filtering and Control.* Moscow: Nauka, 1983.

[17] C.J. Himmelberg. Measurable relations. *Fundamenta mathematicae. T.* **87(1)** (1975), 53-72.

[18] D.P. Bertsekas and S.E. Shreve. *Stochastic Optimal Control. The Discrete Time Case.* New York: Academic Press, 1978.

[19] V.M. Alekseyev, M. Tikhomirov and S.V. Fomin. *Optimal Control.* Moscow: Nauka, 1979.

[20] A.A. Desalu, L.A. Gould and F.C. Schweppe. Dynamic estimation of air pollution. *IEEE Transactions on Automatic Control* **19(6)** (1974), 904-910.

[21] A.B. Kurzhanskii. Identification - A Theory of Guaranteed Estimates in *From Data to Model* (Jan C. Willems, ed.). New York: Springer-Verlag, 1989, 135-214.

[22] D.H. Blackwell and M.A. Girshick. *Theory of Games and Statistical Decisions.* New York: J. Wiley & Sons, 1954.

Institute of Mathematics and Mechanics RAN, 16 Kovalevskaya Street, 620219, GSP-384, Ekaterinburg, Russia

Communicated by A. Kurzhanski