On Minimax Estimation in Uncertain-Stochastic Models With Probability Criteria*

Alexei Pankov and Konstantin Siemenikhin

Probability Theory Dep., Moscow Aviation Inst.

In this work we study the problem of optimal estimation in the multivariate uncertain-stochastic observation model by minimax criterion with generalized probabilistic risk functions. The most general results in this area are obtained using the mean-square error as loss (Martin and Mintz, 1983; Kurzhanski, 1989; Soloviov, 1993; Matasov, 1998; El Ghaoui and Galafiore, 2001; Pankov and Siemenikhin, 2007). Nevertheless, the statistical references based on the mean-square error could lead to non-adequate decisions if the exact joint distribution of random parameters differs from the Gaussian law. At the same time, given *a priori* statistical information in terms of restrictions on the moment characteristics, one can find the tight bounds of various non-mean-square risk functions at linear decision rules (Karlin and Studden, 1966). This makes possible to suggest efficient optimization procedures for designing linear estimation algorithms, which are optimal in a minimax sense. The practical and theoretical interests motivate the following question: whether linear estimators are minimax-optimal over the class of all measurable decision rules given fixed second-order moments of random parameters? For various linear uncertain-stochastic systems this problem has been investigated in detail using the mean-square risk.

In this work we are going to show that there exists a linear operator that is minimax over the family of all unbiased estimators for the broad class of risk functions monotonous with respect to the euclidian norm of the estimation error. In addition, we treat three kinds of estimation criteria based on expectation, probability, and quantile risk functions.

Assume that the system state $X \in \mathbb{R}^m$ is to be estimated from the random observation vector $Y \in \mathbb{R}^n$. Let *a priori* information concerning the observation model (X, Y) be defined by the family $\mathcal{P}$ of the distributions P of the augmented vector $Z = \mathrm{col}[X, Y]$:

$$\mathcal{P} = \{\mathsf{P}\colon \mathsf{E}Z = 0,\ \mathsf{cov}\{Z, Z\} \in \mathcal{R}\}, \tag{1}$$

where $\mathcal{R}$ is a given convex compact set of symmetric positively semidefinite matrices $R = \begin{pmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{pmatrix} \in \mathbb{R}^{p\times p}$ $(p = m + n)$.

The accuracy of any feasible estimate $\widetilde{X} = \varphi(Y)$ is measured by the certain criterion $\mathfrak{D}(\varphi, \mathsf{P})$ whenever P is known to be the true joint distribution of X and Y. The criterion $\mathfrak{D}(\cdot)$ is supposed to have the following structure:

*The work is supported in part by Russian Foundation for Basic Research under Grants No 05-01-00508 and 05-08-17963

a) there exists a functional $\mathfrak{J}(\cdot)$ such that $\mathfrak{D}(\varphi, \mathsf{P}) = \mathfrak{J}\big(\mathsf{P}^{\|X-\varphi(Y)\|^2}\big)$, where $\mathsf{P}^{\|X-\varphi(Y)\|^2}$ is the distribution of the squared euclidian norm of estimation error;

b) $\mathfrak{J}(\pi_1) \le \mathfrak{J}(\pi_2)$ whenever $\pi_1,\ \pi_2 \in \mathcal{P}_+$ satisfy $\pi_1(s,+\infty) \le \pi_2(s,+\infty)$ $\forall\, s \ge 0$, where $\mathcal{P}_+$ denotes the family of all distributions on $[0,\infty)$;

c) $\gamma(\mu) = \sup\{\mathfrak{J}(\pi) \mid \pi \in \mathcal{P}_+,\ \int s\,\pi(ds) = \mu\} = \sup\{\mathfrak{J}(\pi_{t,\mu}) \mid t \ge 1\}\ \forall\, \mu \ge 0$, where $\pi_{t,\mu}$ is the two-point distribution: $\pi_{t,\mu}\{0\} = 1 - 1/t$, $\pi_{t,\mu}\{\mu t\} = 1/t$.

d) $\gamma(\mu)$ is left continuous in $\mu > 0$.

The aim of this work is to find an estimate $\hat{X} = \hat{\varphi}(Y)$, which is optimal in a minimax sense. Namely, we say that $\hat{X} = \hat{\varphi}(Y)$ is *minimax* on the class of estimators Φ w.r.t. $\mathfrak{D}(\cdot)$ if

$$\hat{\varphi} \in \operatorname*{arg\,min}_{\varphi\in\Phi} \sup_{\mathsf{P}\in\mathcal{P}} \mathfrak{D}(\varphi, \mathsf{P}), \tag{2}$$

In addition, we are going to construct the *least favorable distribution*

$$\widehat{\mathsf{P}} \in \operatorname*{arg\,max}_{\mathsf{P}\in\mathcal{P}} \inf_{\varphi\in\Phi} \mathfrak{D}(\varphi, \mathsf{P}). \tag{3}$$

Relations (2), (3) are fulfilled if the pair $(\hat{\varphi}, \widehat{\mathsf{P}})$ forms a saddle point for the game $(\mathfrak{D}, \Phi, \mathcal{P})$: $\mathfrak{D}(\hat{\varphi}, \mathsf{P}) \le \mathfrak{D}(\hat{\varphi}, \widehat{\mathsf{P}}) \le \mathfrak{D}(\varphi, \widehat{\mathsf{P}})\ \forall\, \varphi \in \Phi\ \forall\, \mathsf{P} \in \mathcal{P}$.

The first result provides the tight bound of the estimation criterion on (1) at any linear estimator F.

Theorem 1. *For any $F \in \mathbb{R}^{m\times n}$, $R \in \mathbb{R}^{p\times p}$, $R \succeq O$, we have*

$$\max_{\mathsf{P}\in\mathcal{P}} \mathfrak{D}(F, \mathsf{P}) = \gamma(\max_{R\in\mathcal{R}} Q(F, R)),$$

where $Q(F,R) = \mathsf{E}_{\mathsf{P}}\big\{\|X - FY\|^2\big\} = \mathrm{tr}[R_x - 2R_{xy}F^* + FR_yF^*]\ \forall\, \mathsf{P} \in \mathcal{P}(R)$.

Thus, the minimax estimation problem with any criterion described above is reduced to the mean-square case. In other words, $\hat{F}$ is minimax on the class of linear estimators w.r.t. $\mathfrak{D}(\cdot)$ if and only if

$$\hat{F} \in \operatorname*{arg\,min}_{F\in\mathbb{R}^{m\times n}} \max_{R\in\mathcal{R}} Q(F, R). \tag{4}$$

Following the approach proposed in (Pankov and Siemenikhin, 2007), the linear minimax estimator can be found by the rule:

1) if the least favorable covariance $\hat{R} \in \operatorname*{arg\,max}_{R\in\mathcal{R}} \mathrm{tr}\Delta_R$ satisfies the condition $\hat{R}_y \succ O$, where $\Delta_R = R_x - R_{xy}R_y^+R_{yx}$, then $\hat{F} = \hat{R}_{xy}\hat{R}_y^{-1}$ is a solution of (4);

2) in general, (4) can be obtained as the limit of regularized estimators $\hat{F} = \lim_{\alpha\downarrow 0} \hat{R}^{\alpha}_{xy}(\hat{R}^{\alpha}_y + \alpha I)^{-1}$, where $\hat{R}^{\alpha} \in \operatorname*{arg\,max}_{R\in\mathcal{R}} \mathrm{tr}\big[R_x - R_{xy}(R_y + \alpha I)^{-1}R_{yx}\big]$.

Let us describe the construction of the following random vector $\zeta \in \mathbb{R}^p$:

1) $\zeta = \mathrm{col}[\xi, \eta]$, $\xi \in \mathbb{R}^m$, $\eta \in \mathbb{R}^n$, $\xi = \varepsilon + R_{xy}R_y^+\eta$, $\eta = \mathrm{I}\{\varepsilon = 0\}\upsilon$;

2) $\varepsilon \in \mathbb{R}^m$, $\|\varepsilon\|^2 \sim \pi_{t,\mathrm{tr}\Delta_R}$, $\mathsf{E}\varepsilon = 0$, $\mathsf{cov}\{\varepsilon, \varepsilon\} = \Delta_R$;

3) $\upsilon \in \mathbb{R}^n$, $\mathsf{E}\upsilon = 0$, $\mathsf{cov}\{\upsilon, \upsilon\} = \frac{t}{t-1}R_y$;

4) ε and υ are independent,

where $t > 1$, $R \succeq O$, and $\mathrm{I}\{\dots\}$ denotes the indicator of a random event.

By $\Pi_{t,R}$ we denote the distribution of the vector ζ introduced above.

The main result of the work is the following.

Theorem 2. *Under the previous notation and assumptions, the following assertions are valid*:

1) $\hat{F}$ *is minimax w.r.t.* $\mathfrak{D}(\cdot)$ *on the class* $\mathcal{B}_0$ *of all unbiased estimators*

$$\mathcal{B}_0 = \{\varphi \mid \varphi\colon \mathbb{R}^n \to \mathbb{R}^m \text{ is a Borel mapping such that } \varphi(0) = 0\};$$

2) *the duality relation is fulfilled*

$$\gamma(\mathrm{tr}\Delta_{\hat{R}}) = \min_{\varphi\in\mathcal{B}_0} \sup_{\mathsf{P}\in\mathcal{P}} \mathfrak{D}(\varphi, \mathsf{P}) = \sup_{\mathsf{P}\in\mathcal{P}} \inf_{\varphi\in\mathcal{B}_0} \mathfrak{D}(\varphi, \mathsf{P}),$$

where sup *in the right-hand side is achieved on* $\{\Pi_{t_k,\hat{R}}\}$ *whenever* $\{t_k\}$ *is a sequence of numbers* $t_k > 1$ *maximizing*

$$\sup_{t\geq 1} \mathfrak{J}(\pi_{t,\mathrm{tr}\Delta_{\hat{R}}}); \tag{5}$$

3) *if in* (5) *the maximum is achieved at some* $\hat{t} > 1$, *then the pair* $(\hat{F}, \Pi_{\hat{t},\hat{R}})$ *is a saddle point for* $(\mathfrak{D}, \mathcal{B}_0, \mathcal{P})$ *and hence* $\Pi_{\hat{t},\hat{R}}$ *is the least favorable distribution on* $\mathcal{P}$.

The results indicated above are applied to several cases: *expected loss*: $\mathfrak{D}_\lambda(\varphi, \mathsf{P}) = \mathsf{E}_\mathsf{P}\{\lambda(\|X - \varphi(Y)\|)\}$ with a convex function $\lambda(t)$; *probability criterion*: $\mathfrak{P}_\delta(\varphi, \mathsf{P}) = \mathsf{P}\{\|X - \varphi(Y)\| \geq \delta\}$, where δ is a given positive number; *quantile of level* $1-\alpha$: $\mathfrak{Q}_\alpha(\varphi, \mathsf{P}) = \max\{s \geq 0\colon\ \mathsf{P}\{\|X - \varphi(Y)\| \geq s\} \geq \alpha\}$, $\alpha \in (0, 1)$.

References

El Ghaoui, L. and G. Galafiore (2001). Robust filtering for discrete-time systems with structured uncertainty. *IEEE Trans. Automat. Control.*

Karlin, C. and W. J. Studden (1966). *Tchebycheff systems: with applications in analysis and statistics.* J. Wiley & Sons. N.Y.

Kurzhanski, A. B. (1989). Identification — a theory of guaranteed estimates. In: *From Data to Model* (J. C. Willems, Ed.). pp. 135–214. Springer-Verlag. N.Y.

Martin, C. and M. Mintz (1983). Robust filtering and prediction for linear systems with uncertain dynamics: A game-theoretic approach. *IEEE Trans. Automat. Control* **AC-28**(9), 888–896.

Matasov, A. I. (1998). *Estimators for uncertain dynamic systems.* Kluwer Acad. Publ. Dordrecht.

Pankov, A. R. and K. V. Siemenikhin (2007). Minimax estimation for singular linear multivariate models with mixed uncertainty. *J. Multivariate Analysis* **98**(1), 145–176.

Soloviov, V. N. (1993). Minimax estimation and the least squares method. *Stochastics and Stochastics Rep.* **42**, 209–223.