Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Minimax Identification of Linear Systems by Probability Criterion

A. R. Pankov, A. S. Popov, E. P. Platonov, and K. V. Siemenikhin

Abstract— **The problem of linear estimation of states and parameters of the multidimensional statistically indeterminate system by minimax probability criterion is considered. *A priori* information about the distribution law of the model random parameters and disturbances is defined by certain constraints on the first- and second-order moments. The explicit form of the minimax probability estimator and the sharp guaranteed bound for the probability functional are provided. The analytic expression of the "worst-case" distribution of the random model parameters is presented.**

I. Introduction

In this paper, we wish to investigate the problem of linear estimation for the multidimensional statistically indeterminate model (MSI-model) with mixed-type uncertainties. The model under consideration contains uncertain nonrandom parameters and random disturbances with partially known probability distributions. The special feature of the estimation problem statement is that we use a probability criterion to optimize the estimation procedure. The problem of minimax estimation of the MSI-model states and parameters has a long-standing history [1]–[3]. During the recent period of time, the most general results concerning the mean-square minimax estimation were obtained for both static and dynamic MSI-models with parameters and disturbances of the mixed-type nature [4]–[10]. It should be emphasized that when using the mean-square criterion, we have no real possibility to determine the probabilistic characteristics of the estimate error in the multivariate case even though the observation noise is Gaussian. For example, we cannot calculate the exact value of the probability of the following random event: the norm of the estimate error exceeds some *a priori* fixed level. In practice, the joint distribution law of random parameters is usually unknown and its most important characteristics such as expectation and covariance are defined only approximately. The last makes the estimation problem with probability criterion even more involved. Hence, the problem of minimax estimation of MSI-models using the probability and related criteria is actual, since its solution provides an opportunity to calculate directly the guaranteed values of some probabilistic characteristics of the estimate obtained. At present, the probability and quantile criteria are rather intensively used to solve the stochastic programming problems with both complete [11] and incomplete [12]–[16] *a priori* statistical information. Also the information theoretic criterion should be mentioned as an alternative to the ordinary mean-square criterions. For the purposes of identification of MSI-models, probabilistic criteria have at present limited application. We can indicate interesting methods developed for filtering with non-Gaussian disturbances [17], [18], applications to the linear regression analysis [19], and some others [5], [12]. The main objective of this paper is to elaborate the minimax methods for statistical estimation in MSI-models using the probability optimization criterion. The similar methods for MSI-models were earlier obtained with the help of mean-square optimization via convex analysis and duality theory of minimax problems [6], [7]. The paper is organized as follows: in Section II the MSI-model is described; in Section III the minimax probability criterion and the corresponding estimator are introduced; in Sections IV and V the minimax estimators obtained by the probability criterion for different levels of *a priori* uncertainty, are derived; in Section VI it is shown that the obtained guaranteed values for the probability functional are sharp [20], [21], and the corresponding "worst-case" distributions of the MSI-model random parameters are also presented.

This work was supported by the program "Universities of Russia" (grant number UR-03-01-005)

The authors are with Probability Theory Department, Applied Mathematics and Physics Faculty of Moscow Aviation Institute (Volokolamskoe shosse 4, GSP-3, Moscow A-80, Russia 125993)

A. R. Pankov: arpankov@bk.ru
A. S. Popov: alexlaru@mail.ru
E. N. Platonov: eplatonov@mtu-net.ru
K. V. Siemenikhin: siemenkv@mail.ru

II. Statistically Indeterminate Model

The following notation will be used in the sequel: $\mathsf{E}\{\cdot\}$ and $\mathsf{cov}\{\cdot,\cdot\}$ are the expectation vector and covariance matrix; $\mathrm{col}[x, y] = \begin{pmatrix} x \\ y \end{pmatrix}$; $|\cdot|$ and $\langle \cdot, \cdot \rangle$ are the Euclidean norm and inner product on $\mathbb{R}^m$; A^+, $A^\top$, $\mathrm{tr}[A]$, and $\|A\|$ are the pseudoinverse, transpose, trace and spectral norm of a matrix A; $A > 0$ ($A \geq 0$) means that the matrix A is symmetric and positively (semi)definite; $\arg\min_{x \in X} f(x)$ is the set of all minima of $f(x)$ on X.

Consider the following multidimensional linear observation model:

$$\begin{cases} x = a\theta + b\xi, \\ y = A\theta + B\xi, \end{cases} \tag{1}$$

where $x \in \mathbb{R}^m$ is a vector to be estimated given the observation vector $y \in \mathbb{R}^n$; $\theta \in \mathbb{R}^p$ is the unknown vector of nonrandom parameters; $\xi \in \mathbb{R}^q$ is the vector of random parameters and/or disturbances such that $\mathsf{E}\{|\xi|^2\} < \infty$; the matrices a, b, A, and B are supposed to be known and have appropriate dimension.

Concerning θ there are no *a priori* restrictions, i.e., $\theta \in \mathbb{R}^p$.

The vector ξ is supposed to have zero mean and partially known covariance matrix:

$$V = \mathsf{cov}\{\xi, \xi\} \in \mathcal{V},$$

where $\mathcal{V}$ is a given set of positively semidefinite $q \times q$ matrices.

So, $\rho = \mathrm{col}[\theta, \xi]$ is the uncertain-stochastic vector of the model (1) parameters. Denote the distribution law of ρ by P_ρ. It should be noted that P_ρ is multidimensional and singular.

Thus, the assumptions stated above can be represented in the form:

$$\mathsf{P}_\rho \in \mathcal{P},$$

where $\mathcal{P}$ is the set of all admissible distributions P_ρ:

$$\mathcal{P} = \{\mathsf{P}_\rho \colon \rho = \mathrm{col}[\theta, \xi],\ \theta = \mathsf{E}\{\theta\} \in \mathbb{R}^p, \quad \mathsf{E}\{\xi\} = 0,\ \mathsf{cov}\{\xi, \xi\} \in \mathcal{V}\}. \quad (2)$$

III. Minimax Linear Estimation by Probability Criterion

In this paper, we consider only linear estimates $\widetilde{x}$ of x given the observation vector y:

$$\hat{x} = Fy, \quad F \in \mathbb{R}^{m \times n}.$$

Suppose that any linear estimate is admissible.

In order to optimize the estimator F we introduce the following probability functional:

$$J_\gamma(F, \mathsf{P}_\rho) = \mathsf{P}\{|Fy - x| \geq \gamma\}, \quad (3)$$

where $\gamma > 0$ is given, $F \in \mathbb{R}^{m \times n}$, and $\mathsf{P}_\rho \in \mathcal{P}$.

So, we are going to choose F by minimizing (3). Evidently, given a $\gamma > 0$, the problem of F optimization with respect to (3) has a non-trivial solution only if there exists an estimator $\overline{F}$ such that $J_\gamma(\overline{F}, \mathsf{P}_\rho) < 1$.

Since the exact distribution P_ρ is unknown, we shall use the minimax approach to obtain the desired optimal estimator $\hat{F}$.

Definition 1: The estimator $\hat{F}$ is called minimax with respect to the probability criterion if

$$\hat{F} \in \arg \min_{F \in \mathbb{R}^{m \times n}} \sup_{\mathsf{P}_\rho \in \mathcal{P}} J_\gamma(F, \mathsf{P}_\rho). \quad (4)$$

IV. Minimax Estimation with The Known Covariance

In this section we shall describe a particular solution of problem (4).

Denote $V = \mathsf{cov}\{\xi, \xi\}$ and consider the set $\mathcal{F}_0$ of unbiased estimators:

$$\mathcal{F}_0 = \{F \colon F \in \mathbb{R}^{m \times n},\ FA = a\}.$$

It is known that $\mathcal{F}_0 \neq \varnothing$ if and only if the following observability condition holds [22]:

$$aA^+A = a. \quad (5)$$

Introduce the following additional notation:

$$D(F, V) = \left(\mathrm{tr}\left[(FB - b)V(FB - b)^\top\right]\right)^{1/2}, \quad (6)$$

$$F(V) = F_0 + (b - F_0B)VB^\top(QBVB^\top Q)^+, \quad (7)$$

$$I(V) = \mathrm{tr}\left[(F_0B - b) \times (V - VB^\top(QBVB^\top Q)^+BV)(F_0B - b)^\top\right], \quad (8)$$

where $F_0 = aA^+$, $Q = I - AA^+$.

By $\mathcal{P}(V)$ denote the following subset of $\mathcal{P}$:

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

By definition, $\mathcal{P}(V)$ contains all admissible distributions P_ρ with the fixed covariance V of ξ.

Proposition 1: Assume that condition (5) *holds and there exists a matrix $\overline{F}$ such that $D(\overline{F}, V) < \gamma$, where $V \geq O$. Then, for any $F \in \mathbb{R}^{m \times n}$ we have*

$$\sup_{\mathsf{P}_\rho \in \mathcal{P}(V)} J_\gamma(F, \mathsf{P}_\rho) = \begin{cases} \gamma^{-2} D^2(F, V) & \text{if} \quad F \in \mathcal{F}_0, \\ 1 & \text{if} \quad F \notin \mathcal{F}_0. \end{cases} \quad (9)$$

The proof of Proposition 1 is given in the Appendix.

From the results of [6], [7] it immediately follows that the following optimization problem has an explicit solution:

$$F(V) \in \arg \min_{F \in \mathcal{F}_0} D^2(F, V), \quad (10)$$

and the minimal value of $D^2(F, V)$ can be also represented analytically:

$$\min_{F \in \mathcal{F}_0} D^2(F, V) = I(V),$$

where $F(V)$ and $I(V)$ are defined by (7), (8). Hence, from Proposition 1 and relations (9)–(10) we derive the following result.

Proposition 2: Under condition (5) *and notation* (6)–(8)*, we suppose that $I(V) < \gamma^2$.*

Then,

1) *the estimator $F(V)$ is minimax, i.e.,*

$$F(V) \in \arg \min_{F \in \mathbb{R}^{m \times n}} \sup_{\mathsf{P}_\rho \in \mathcal{P}(V)} J_\gamma(F, \mathsf{P}_\rho);$$

2) *the guaranteed value of $J_\gamma(\cdot)$ is as follows:*

$$\hat{J}_\gamma = \min_{F \in \mathbb{R}^{m \times n}} \sup_{\mathsf{P}_\rho \in \mathcal{P}(V)} J_\gamma(F, \mathsf{P}_\rho) = \gamma^{-2} I(V).$$

V. Minimax Estimation: The General Case

Now let us assume that the covariance matrix V belongs to some given set $\mathcal{V}$ of positively semidefinite matrices.

From the results of the previous section we obtain

$$\sup_{\mathsf{P}_\rho \in \mathcal{P}} J_\gamma(F, \mathsf{P}_\rho) = \sup_{V \in \mathcal{V}} \gamma^{-2} D^2(F, V)$$

for all sufficiently large γ, where $\mathcal{P}$ is defined by (2). Hence, the minimax estimator (4) is a solution of the following optimization problem:

$$\hat{F} \in \arg \min_{F \in \mathcal{F}_0} \sup_{V \in \mathcal{V}} D^2(F, V). \quad (11)$$

From the results of [6] it follows that the estimator that is minimax with respect to the mean-square-error criterion $D(\cdot)$ can be obtained using the dual optimization problem:

$$\hat{V} \in \arg\max_{V \in \mathcal{V}} I(V). \tag{12}$$

Here, we suppose that $\mathcal{V}$ is a convex compact set and $\hat{V}$ is an arbitrary solution of (12).

Due to [6], [7], we can claim that the solution $\hat{F} = F(\hat{V})$ of the minimization problem (10) with the least favorable covariance $\hat{V}$ satisfies (11) whenever the following regularity condition is fulfilled:

$$B\hat{V}B^\top > O. \tag{13}$$

Now we can summarize the results obtained above.

Proposition 3: Suppose that (5) *and* (13) *hold,* $\hat{V}$ *is provided by* (12) *and* $I(\hat{V}) < \gamma^2$.

Then,

1) *the estimator* $F(\hat{V})$ *is minimax, i.e.,*

$$F(\hat{V}) \in \arg\min_{F \in \mathbb{R}^{m\times n}} \sup_{\mathsf{P}_\rho \in \mathcal{P}} J_\gamma(F, \mathsf{P}_\rho),$$

where $F(\cdot)$ *is defined by* (7)*;*

2) *the guaranteed value of* $J_\gamma(\cdot)$ *is the following:*

$$\hat{J}_\gamma = \min_{F \in \mathbb{R}^{m\times n}} \sup_{\mathsf{P}_\rho \in \mathcal{P}} J_\gamma(F, \mathsf{P}_\rho) = \gamma^{-2} I(\hat{V}).$$

In the general case, the solution $\hat{V}$ of the dual problem (12) can be obtained numerically by applying the iterative algorithm derived in [6], [7].

Note that if $\mathcal{V}$ contains the maximal element $\overline{V}$, i.e., $\overline{V} \geq V$ for all $V \in \mathcal{V}$, then $\hat{V} = \overline{V}$ and (13) can be omitted.

VI. The "Worst-Case" Distribution

Since $\hat{F} \in \mathcal{F}_0$, the error of the minimax estimate $\hat{x} = \hat{F}y$ takes the form:

$$\varepsilon = \hat{x} - x = (\hat{F}B - b)\xi. \tag{14}$$

Hence, $\mathsf{E}\{\varepsilon\} = 0$ and

$$R_\varepsilon(V) = \mathsf{cov}\{\varepsilon, \varepsilon\} = (\hat{F}B - b)V(\hat{F}B - b)^\top$$

if $\mathsf{cov}\{\xi, \xi\} = V$.

From Proposition 1 and [7] we derive

$$\sup_{\mathsf{P}_\varepsilon \in \mathcal{P}_\varepsilon} \mathsf{P}\{|\varepsilon| \geq \gamma\} = \sup_{V \in \mathcal{V}} \gamma^{-2} D(\hat{F}, V) = \gamma^{-2} I(\hat{V}),$$

where $\mathcal{P}_\varepsilon$ denotes the set of m-dimensional distributions with zero mean and covariance $R_\varepsilon(V)$, $V \in \mathcal{V}$.

Therefore, $\hat{\mathsf{P}}_\varepsilon$ is the "worst-case" distribution of the estimate error ε if $\hat{\mathsf{P}}_\varepsilon\{u\colon |u| \geq \gamma\} = \gamma^{-2} I(\hat{V})$.

In the following proposition, we present the explicit form of the estimate error with the "worst-case" distribution $\hat{\mathsf{P}}_\varepsilon$.

Proposition 4: Under the condition $0 < I(\hat{V}) < \gamma^2$*, put*

$$\varepsilon = \gamma\, I(\hat{V})^{-1/2}\, \delta_0\, \eta, \quad \eta = \sum_{i=1}^m \sqrt{r_i}\, \delta_i\, e^{(i)}, \tag{15}$$

where $\{e^{(i)}\}$ *are the eigenvectors of the matrix* $R_\varepsilon(\hat{V})$ *such that* $|e^{(i)}| = 1$*;* $\{r_i\}$ *are the respective eigenvalues;* $\delta_0, \ldots, \delta_m$ *are independent random variables distributed as follows:*

$$\mathsf{P}\{\delta_0 = 1\} = \gamma^{-2} I(\hat{V}),\ \mathsf{P}\{\delta_0 = 0\} = 1 - \gamma^{-2} I(\hat{V}),$$
$$\mathsf{P}\{\delta_i = \pm 1\} = 1/2,\ i = 1, \ldots, m.$$

Then the distribution of (15) *is "worst case", i.e.,*

$$\begin{gathered} \mathsf{E}\{\varepsilon\} = 0, \quad \mathsf{cov}\{\varepsilon, \varepsilon\} = R_\varepsilon(\hat{V}), \\ \mathsf{P}\{|\varepsilon| \geq \gamma\} = \gamma^{-2} I(\hat{V}). \end{gathered} \tag{16}$$

The proof of Proposition 4 is given in the Appendix.

For the particular case: $x \in \mathbb{R}^1$ (i.e., $m = 1$), the "worst-case" distribution of ε is presented in Table I.

TABLE I
The Univariate "Worst-Case" Distribution of The Minimax Estimate Error $\varepsilon = \hat{x} - x$

u	$-\gamma$	0	γ
$\hat{\mathsf{P}}_\varepsilon\{u\}$	$\frac{I(\hat{V})}{2\gamma^2}$	$1 - \frac{I(\hat{V})}{\gamma^2}$	$\frac{I(\hat{V})}{2\gamma^2}$

The next result shows that there always exists a "worst-case" distribution $\hat{\mathsf{P}}_\xi$ that generates the desired distribution $\hat{\mathsf{P}}_\varepsilon$ of the estimate error (14).

Proposition 5: Let $\varepsilon \in \mathbb{R}^m$ *be a random vector with*

$$\mathsf{E}\{\varepsilon\} = H\mu, \quad \mathsf{cov}\{\varepsilon, \varepsilon\} = HVH^\top,$$

where $\mu \in \mathbb{R}^q$*,* $H \in \mathbb{R}^{m\times q}$*,* $V \in \mathbb{R}^{q\times q}$*, and* $V \geq O$*.*

If the random vector ξ *is determined by the following expression:*

$$\xi = G\varepsilon + (I - GH)\delta, \quad G = VH^\top(HVH^\top)^+, \tag{17}$$

where $\delta \in \mathbb{R}^q$ *is an arbitrary random vector such that*

$$\mathsf{E}\{\delta\} = \mu, \quad \mathsf{cov}\{\delta, \delta\} = V, \quad \textit{and} \quad \mathsf{cov}\{\delta, \varepsilon\} = 0,$$

then

$$\mathsf{E}\{\xi\} = \mu, \quad \mathsf{cov}\{\xi, \xi\} = V, \quad \mathsf{P}\{\varepsilon = H\xi\} = 1. \tag{18}$$

The proof of Proposition 5 is given in the Appendix.

So, in order to obtain the desired "worst-case" perturbation vector ξ, one should put $H = \hat{F}B - b$, $\mu = 0$, and assume that ε satisfies (16).

Note that the "worst-case" distribution of (15) is a discrete one, while the corresponding q-dimensional law of ξ may be absolutely continuous as it follows from (17).

VII. Appendix

Proof of Proposition 1: For the random vector $\varepsilon = Fy - x = (FA - a)\theta + (FB - b)\xi$, we have

$$\mathsf{E}\{\varepsilon\} = (FA - a)\theta, \quad \mathsf{cov}\{\varepsilon, \varepsilon\} = R_\varepsilon(V), \tag{19}$$

where $R_\varepsilon(V) = (FB - b)V(FB - b)^\top$.

Let $\mathcal{P}_\varepsilon(F)$ be the set of m-dimensional distributions with characteristics (19), where $\theta \in \mathbb{R}^p$ and $V \in \mathcal{V}$ are arbitrary. Due to Proposition 5, we can claim

$$\sup_{\mathsf{P}_\rho \in \mathcal{P}} J_\gamma(F, \mathsf{P}_\rho) = \sup_{\mathsf{P}_\varepsilon \in \mathcal{P}_\varepsilon(F)} \mathsf{P}\{|\varepsilon| \geq \gamma\}.$$

First, let us consider the case: $F \notin \mathcal{F}_0$, i.e., $FA - a \neq 0$. Then

$$\sup_{\mathsf{P}_\varepsilon \in \mathcal{P}_\varepsilon(F)} \mathsf{P}\{|\varepsilon| \geq \gamma\} = 1 \quad \forall \gamma > 0,$$

since $\sup\limits_{\theta \in \mathbb{R}^p} |\mathsf{E}\{\varepsilon\}| = \sup\limits_{\theta \in \mathbb{R}^p} |(FA - a)\theta| = \infty$.

Now assume $F \in \mathcal{F}_0$, i.e., $\mathsf{E}\{\varepsilon\} = 0$. Using Theorem 2.1 [20, chap. 13], we obtain

$$\sup_{\mathsf{P}_\varepsilon \in \mathcal{P}_\varepsilon(F)} \mathsf{P}\{|\varepsilon| \geq \gamma\} = \min\{1, \min_{C \in \Delta} \mathrm{tr}[C R_\varepsilon(V)]\},$$

where

$$\Delta = \{C \in \mathbb{R}^{m \times m} \colon C > 0, \ \langle C\lambda, \lambda\rangle \geq 1 \ \forall \lambda \colon |\lambda| \geq \gamma\}.$$

It is clear that $C \in \Delta$ if and only if $\langle C\lambda, \lambda \rangle < 1$ implies $|\lambda| < \gamma$ for any $\lambda \in \mathbb{R}^m$. The last can be represented in the form:

$$|\mu| < 1 \quad \Longrightarrow \quad |C^{-1/2}\mu| < \gamma \quad \forall \mu \in \mathbb{R}^m. \tag{20}$$

Obviously, (20) means $\|C^{-1/2}\| \leq \gamma$, whence we obtain

$$\Delta = \{C \in \mathbb{R}^{m \times m} \colon C > 0, \ \|C^{-1}\| \leq \gamma^2\}.$$

Note that for any $C \in \Delta$

$$\mathrm{tr}[C R_\varepsilon(V)] \geq \|C^{-1}\|^{-1} \mathrm{tr}[R_\varepsilon(V)] \geq \gamma^{-2} \mathrm{tr}[R_\varepsilon(V)]\,.$$

Due to $C_0 = \gamma^{-2} I \in \Delta$, we have

$$\min_{C \in \Delta} \mathrm{tr}[C R_\varepsilon(V)] = \mathrm{tr}[C_0 R_\varepsilon(V)] = \gamma^{-2} \mathrm{tr}[R_\varepsilon(V)]\,.$$

This completes the proof. ■

Proof of Proposition 4: First, let us find the moment characteristics of η. By definition, we have

$$\mathsf{E}\{\eta\} = 0 \quad \text{and} \quad \mathsf{cov}\{\eta\} = \sum_{i=1}^m r_i e^{(i)} (e^{(i)})^\top = R_\varepsilon(\hat{V}),$$

since $\{\delta_i\}$ are independent and $\mathsf{D}\{\delta_i\} = 1$.

By assumption, δ_0 and η are independent. Hence, $\mathsf{E}\{\varepsilon\} = 0$ and

$$\begin{aligned} \mathsf{cov}\{\varepsilon, \varepsilon\} &= \mathsf{E}\big\{\varepsilon\varepsilon^\top\big\} \\ &= \frac{\gamma^2}{I(\hat{V})} \mathsf{E}\big\{\delta_0^2\big\}\, \mathsf{E}\big\{\eta\eta^\top\big\} = \mathsf{E}\big\{\eta\eta^\top\big\} = R_\varepsilon(\hat{V}). \end{aligned}$$

Now it remains to show

$$\mathsf{P}\{|\varepsilon| \geq \gamma\} = \gamma^{-2} I(\hat{V}). \tag{21}$$

Since $\{e^{(i)}\}$ are orthogonal, we have $|\eta|^2 = \sum\limits_{i=1}^m r_i \delta_i^2 = \mathrm{tr}\Big[R_\varepsilon(\hat{V})\Big] = I(\hat{V})$. Then $|\varepsilon|^2 = \gamma^2 \delta$ and (21) follows from $\mathsf{P}\big\{|\varepsilon|^2 \geq \gamma^2\big\} = \mathsf{P}\{\delta_0 \geq 1\} = \gamma^{-2} I(\hat{V})$. ■

Proof of Proposition 5: Let us check (18). First $\mathsf{E}\{\xi\} = GH\mu + (I - GH)\mu = \mu$. Secondly, $\mathsf{cov}\{\xi, \xi\} = \mathsf{cov}\{G\varepsilon, G\varepsilon\} + \mathsf{cov}\{(I - GH)\delta, (I - GH)\delta\}$, since ε and δ are uncorrelated. Then,

$$\begin{aligned} \mathsf{cov}\{\xi, \xi\} &= GHVH^\top G^\top + (I - GH)V(I - H^\top G^\top) \\ &= V - GHV - VH^\top G^\top + 2\,GHVH^\top G^\top = V, \end{aligned}$$

where

$$GHV = VH^\top G^\top = GHVH^\top G^\top = VH^\top (HVH^\top)^+ HV.$$

Finally, let us prove $\mathsf{E}\big\{|H\xi - \varepsilon|^2\big\} = 0$. Note that $H\xi - \varepsilon = (I - P)(H\delta - \varepsilon)$, where $P = HG = HVH^\top \times (HVH^\top)^+$ and $\mathsf{cov}\{H\delta - \varepsilon, H\delta - \varepsilon\} = 2\,HVH^\top$. Then, $\mathsf{E}\big\{|\varepsilon - H\xi|^2\big\} = 2\,\mathrm{tr}\big[(I - P)HVH^\top (I - P)\big] = 0$. ■

REFERENCES

[1] V. D. Vandelinde, "Robust properties of solutions to linear-quadratic estimation and control problems," *IEEE Trans. Automat. Control*, vol. 22, no. 1, pp. 138–139, 1977.

[2] S. Verdú and H. V. Poor, "On minimax robustness: A general approach and applications," *IEEE Trans. Inform. Theory*, vol. 30, no. 2, pp. 328–340, 1984.

[3] A. B. Kurzhanski, "Identification — a theory of guaranteed estimates," in *From Data to Model*, J. C. Willems, Ed. N.Y.: Springer, 1989, pp. 135–214.

[4] A. I. Matasov, *Estimators for Uncertain Dynamic Systems*. Dordrecht: Kluwer Acad. Publ., 1998.

[5] A. Garulli, A. Tesi, and V. Vicino, Eds., *Robustness in identification and control*. N.Y.: Springer, 1999.

[6] A. R. Pankov, E. N. Platonov, and K. V. Siemenikhin, "On minimax identification: Method of dual optimization," in *Proceedings of the 39th IEEE Conference on Decision and Control (CDC'2000)*, Sydney, Australia, Dec. 12–15, 2000, pp. 4759–4764.

[7] A. R. Pankov and K. V. Siemenikhin, "The parametric identification methods for many-dimensional linear models in the presence of *a priori* uncertainty," *Automation and Remote Control*, vol. 61, no. 5. Part 1, pp. 785–800, 2000, (translated).

[8] G. Galafiore and L. El Ghaoui, "Worst-case maximum likelihood estimation in the linear model," *Automatica*, vol. 37, no. 4, 2001.

[9] L. El Ghaoui and G. Galafiore, "Robust filtering for discrete-time systems with structured uncertainty," *IEEE Trans. Automat. Control*, vol. 46, no. 7, 2001.

[10] L. Li, Z.-Q. Luo, T. Davidson, M. Wong, and E. Bosse, "Robust filtering via semidefinite programming with applications to target tracking," *SIAM J. Optimization*, vol. 12, pp. 740–755, 2002.

[11] A. I. Kibzun and Yu. S. Kan, *Stochastic Programming Problems (with Probability and Quantile Functions)*. N.Y.: Wiley, 1996.

[12] J. Smith, "Generalized Chebyshev inequalities: theory and applications in decision analysis," *Operations Research*, vol. 43, pp. 807–825, 1995.

[13] A. Shapiro and A. Kleywegt, "Minimax analysis of stochastic problems," *Optimization Methods and Software*, vol. 17, 2002.

[14] L. El Ghaoui, M. Oks, and F. Oustry, "Worst-case value-at-risk and robust optimization: a conic programming approach," *Operations Research*, vol. 51, pp. 543–556, 2003.

[15] A. R. Pankov, E. N. Platonov, A. S. Popov, and K. V. Siemenikhin, "Linear stochastic programming with minimax quantile and probability criterions," in *Proceedings of the 43rd IEEE Conference on Decision and Control (CDC'2004)*, Atlantis, Paradise Island, Bahamas, Dec. 14–17, 2004, pp. 3179–3182.

[16] F. Topsoe, "Maximum entropy versus minimum risk and applications to some classical discrete distributions," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2368–2376, 2002.

[17] J. C. Spall, "The Kantorovich inequality for error analysis of the Kalman filter with unknown noise distributions," *Automatica*, vol. 31, pp. 1513–1517, 1995.

[18] J. L. Maryak, J. C. Spall, and B. D. Heydon, "Use of the Kalman filter for inference in state-space models with unknown noise distributions," *IEEE Trans. Automat. Control*, vol. 49, pp. 87–90, 2004.

[19] R. T. Rockafellar, S. Uryasev, and M. Zabarankin, "Deviation measures in generalized linear regression," University of Florida, Research Rep. 2002-9. ISE Dept., 2002.

[20] C. Karlin and W. J. Studden, *Tchebycheff systems: with applications in analysis and statistics*. N.Y.: Wiley, 1966.

[21] L. D. Tong, *Probability inequalities in multivariate distributions*. N.Y.: Academic Press, 1980.

[22] A. Albert, *Regression and the Moore–Penrose Pseudoinverse*. N.Y.: Academic Press, 1972.