MINIMAX FILTER FOR STATISTICALLY INDETERMINATE STOCHASTIC DIFFERENTIAL SYSTEM

Alexey Pankov [*] **Gregory Miller** [*,**,1]

[*] *Probability Theory Department, Applied Mathematics and Physics Faculty of Moscow Aviation Institute: 4, Volokolamskoe sh., GSP-3, Moscow A-80, Russia 125993*
[**] *Institute of Informatics Problems: 44/2, Vavilova st., Moscow, Russia 117333*
A. Pankov: arpankov@bk.ru
G. Miller: gregbmiller@mail.ru

Abstract: The linear stochastic differential system with uncertain intensity of noises in dynamics and observations is considered. For this system the minimax filtering procedure is proposed. The filter is optimal in terms of integral criterion. The obtained filtering equations depend on the dual optimization problem solution, which can be obtained by means of provided numerical procedure. The convergence of the numerical procedure is also considered. Some numerical results are described.

Keywords: Stochastic systems, Optimal filtering, Minimax techniques

1. INTRODUCTION

The optimal recursive filtering algorithms for linear stochastic systems are well-known and widely used in various fields (Liptser and Shiryaev, 1978; Davis, 1977; Pugachev and Sinitsyn, 1987). For practical applications of those methods the whole information about the first and second order moments of noises in dynamics as well as in observation system is necessary. Inaccurate definition of the moments can lead to serious difference between the real filtering errors and their nominal values (Sage and White, 1977).

In (Bertsekas and Rhodes, 1971; Morris, 1976) the minimax approach was applied to systems with uncertain deterministic disturbances, and the guaranteed estimate problem solution was obtained. Models with deterministic disturbances were also investigated in (Matasov, 1998), and discrete stochastic dynamic models were investigated in (Golubev *et al.*, 1989; Katz and Timofeeva, 1994; Verdu and Poor, 1984). Some new valuable results for minimax filtering of processes in discrete-time systems with uncertain dynamics and perturbations are presented in (Li *et al.*, 2002). Statistically uncertain models described by stochastic differential equations were examined in (Matasov, 1998; Bobrik *et al.*, 1997; Orlov and Basin, 1995; Borisov and Pankov, 1998). Rather general results were obtained only for stationary systems by applying the spectral methods. The theory for time-dependent differential systems implies the necessity of solving the nonsmooth variational problem when using the local optimization criterion (Matasov, 1998). It is a complicated problem unless some serious simplifications are made.

[1] This work was supported by INTAS (YSF 04-83-3623)

In this paper, a model described by a system of stochastic differential equations with piecewise continuous coefficients is considered. It is assumed that the noise intensities in dynamics and observation equations are not known exactly, but belong to some known uncertainty sets. The estimate accuracy is determined by the integral mean-square criterion.

The minimax filter is obtained by solving the dual optimization problem (Borisov and Pankov, 1998; Pankov and Miller, 2001). Using this approach the criterion saddle point existence is proved, and is shown, that the minimax estimator can be expressed analytically as a function of the dual optimization problem solution. The main objective of this paper is to provide an effective convergent numerical method, which makes possible to solve the dual optimization problem and, hence, to derive the minimax filtering algorithm.

2. THE MODEL DEFINITION

Consider the following continuous-time observation model:

$$\mathrm{d}y_t = a_t y_t \mathrm{d}t + b_t \mathrm{d}w_t,\ y_0 = 0, \tag{1}$$

$$\mathrm{d}z_t = c_t y_t \mathrm{d}t + d_t \mathrm{d}w_t,\ t \in [0, T]. \tag{2}$$

In (1),(2) $y_t \in \mathbb{R}^p$ is the system state and $z_t \in \mathbb{R}^q$ is the observation vector at time $t \in [0, T]$. The process $w_t \in \mathbb{R}^r$ is supposed to be homogeneous random process with orthogonal increments:

$$\begin{gathered} K_w(t,\tau) = \mathrm{cov}(w_t, w_\tau) = \gamma \min(t,\tau), \\ \mathrm{M}\,[w_t] = 0. \end{gathered} \tag{3}$$

The matrix-valued functions a_t, b_t, c_t and d_t are assumed to be known and piecewise continuous. The intensity γ is not known, but belongs to some set of positively definite matrices: $\gamma \in \Gamma \subset \mathbb{R}^{r\times r}$. It is assumed also that the observation model is nonsingular (Liptser and Shiryaev, 1978): $\exists C > 0$ such that $\forall t \in [0, T]$ and $\forall \mu \in \mathbb{R}^q$: $\|\mu\| = 1$

$$\inf_{\gamma\in\Gamma} \mu^* d_t \gamma d_t^* \mu \geq C. \tag{4}$$

Let $\hat{y}_t$ be a nonanticipating linear estimate for y_t given the observations $Z^t = \{z_\tau, 0 \leq \tau \leq t\}$. Then $\hat{y}_t$ can be represented as follows:

$$\hat{y}_t = F(Z^t) = \int_0^T g(t,\tau)\mathrm{d}z_\tau,\ g(t,\tau) = 0\ \forall \tau > t. \tag{5}$$

In (5) F is a filter operator and $g(t, \tau)$ is it's weighting function. Denote $\mathcal{F}$ the set of filters (5), satisfying the condition $\mathrm{M}\left[\|\hat{y}_t\|^2\right] < \infty$. This set is convex and closed but unbounded in general.

Let $P_w \in \mathcal{P}_w$ be the distribution law of the process w_t, where $\mathcal{P}_w$ is the set of all distributions which satisfy (3) with $\gamma \in \Gamma$.

For every $P_w \in \mathcal{P}_w$ the estimate $\hat{y}_t$ accuracy is determined by the following integral mean-square criterion

$$\mathbf{J}(F, P_w) = \mathrm{M}\left[\int_0^T \xi_t^* \Sigma_t \xi_t \mathrm{d}t\right], \tag{6}$$

where $\mathrm{M}\,[\cdot]$ is the expectation operator (with respect to the distribution P_w); $\xi_t = \hat{y}_t - y_t$ is the estimate error; Σ_t is given piecewise continuous weighting matrix function, $\Sigma_t = \Sigma_t^*$ and $\Sigma_t \geq 0$, $t \in [0, T]$ (a matrix inequality $A \geq B$ means that the matrix $A - B$ is positively semidefinite one).

Using (1)-(3), (5), (6) it can be shown by straightforward calculations that

$$\mathbf{J}(F, P_w) = J_T(F,\gamma) = \int_0^T \mathrm{tr}[\Sigma_t R_t(F,\gamma)]\mathrm{d}t, \tag{7}$$

where $R_t(F, \gamma)$ is the error ξ_t covariance matrix, which depends only on the operator F and the intensity matrix γ.

3. MINIMAX FILTER

Denote $\operatorname*{argmax}_{x\in X} f(x)$ $(\operatorname*{argmin}_{x\in X} f(x))$ the set of points of maxima (minima) of $f(x)$ on the set X.

Definition 1. The operator $\hat{F}$ is minimax one with respect to $J_T(F, \gamma)$ criterion on the set $F \times \Gamma$ if

$$\hat{F} \in \operatorname*{argmin}_{F\in\mathcal{F}} \sup_{\gamma\in\Gamma} J_T(F,\gamma). \tag{8}$$

If the set Γ contains only one point $\Gamma = \theta$, then the Kalman-Bucy filter provides the operator which is minimax with respect to $J_T(F, \gamma)$ (Morris, 1976; Verdu and Poor, 1984). For the case (1),(2) we have

$$\mathrm{d}\hat{y}_t = a_t\hat{y}_t\mathrm{d}t + K_t(\theta)(\mathrm{d}z_t - c_t\hat{y}_t\mathrm{d}t),\ \ \hat{y}_0 = 0, \tag{9}$$

$$K_t(\theta) = (R_t(\theta)c_t^* + b_t\theta d_t^*)(d_t\theta d_t^*)^{-1}, \tag{10}$$

$$\begin{cases} \dot{R}_t(\theta) = & a_tR_t(\theta) + R_t(\theta)a_t^* + \\ & +b_t\theta b_t^* - K_t(\theta)d_t\theta d_t^* K_t^*(\theta), \\ R_0(\theta) = 0. & \end{cases} \tag{11}$$

The equation (8) describes the direct minimax optimization problem, and the following equation describes the dual optimization problem:

$$\hat{\gamma} \in \operatorname*{argmax}_{\gamma\in\Gamma} J_T^0(\gamma), \tag{12}$$

where $J_T^0(\gamma) = \inf_{F\in\mathcal{F}} J_T(F,\gamma)$ is the dual criterion.

Theorem 1. Let Γ be a convex compact set of positively definite matrices γ satisfying (4). Then

1. the dual criterion $J_T^0(\gamma)$ has the following analytical representation:
$$J_T^0(\gamma) = \int\limits_0^T \mathrm{tr}[\Sigma_t R_t(\gamma)]\mathrm{d}t, \qquad (13)$$
where $R_t(\gamma)$ is determined by (10), (11) with $\theta = \gamma$;
2. the dual problem (12) solution $\hat{\gamma}$ exists;
3. the pair $(\hat{F}, \hat{\gamma})$, where $\hat{F} = F_K(\hat{\gamma})$ is the Kalman-Bucy filter (9)-(11) with $\theta = \hat{\gamma}$, forms the saddle point of $J_T(F, \gamma)$ on $\mathcal{F} \times \Gamma$;
4. the guaranteed value $\hat{J}_T$ of criterion (7) is equal to
$$\hat{J}_T = \int\limits_0^T \mathrm{tr}[\Sigma_t R_t(\hat{\gamma})]\mathrm{d}t = J_T^0(\hat{\gamma}). \qquad (14)$$

The proof is given in the Appendix.

Corollary 1. The operator $\hat{F} = F_K(\hat{\gamma})$ is a minimax one on $\mathcal{F} \times \Gamma$:
$$\sup_{\gamma\in\Gamma} J_T(\hat{F}, \gamma) \le \sup_{\gamma\in\Gamma} J_T(F, \gamma), \quad \forall F \in \mathcal{F}.$$
The operator $\hat{F}$ is recurrent and is determined by (9)-(11) with $\theta = \hat{\gamma}$.

Now consider the same problem on the extended set of admissible operators. Let $\mathcal{F}_0$ be the set of all nonanticipating estimation operators, so $F \in \mathcal{F}_0$ if $\hat{y}_t = F(Z^t)$ is measurable with respect to the σ-field generated by $Z^t = \{z_\tau, 0 \le \tau \le t\}$, $t \in [0, T]$, and $\mathrm{M}\left[\|\hat{y}_t\|^2\right] < \infty$. Let $\mathcal{P}_w$ be the set of all distributions of the process $\{w_t,\ t \in [0, T]\}$, satisfying (3) while $\gamma \in \Gamma$. Let also $\mathbf{J}_T(F, P_w)$ be the integral mean-square criterion determined on $\mathcal{F}_0 \times \mathcal{P}_w$ by (6).

Theorem 2. The pair $(\hat{F}, \hat{P}_w)$, where $\hat{F} = F_K(\hat{\gamma})$, $\hat{\gamma}$ is the solution of (12), (13) and $\hat{P}_w$ is the distribution of the Wiener process $\hat{w}_t$ with covariance $\mathrm{cov}(\hat{w}_t, \hat{w}_\tau) = \hat{\gamma}\min(t, \tau)$, forms the saddle point of $\mathbf{J}_T(F, P_w)$ on $\mathcal{F}_0 \times \mathcal{P}_w$.

4. THE DUAL PROBLEM SOLUTION

In this section it is shown how the dual optimization problem (12) can be numerically solved in the general case.

Let $F_K(\theta) \in \mathcal{F}$ be the Kalman-Bucy filter for some $\theta \in \Gamma$. Then it can be shown (Lemma 1 in Appendix) that
$$J_T(F_K(\theta), \gamma) = \mathrm{tr}[H_T^*(\theta)\gamma],$$
where $H_T(\theta) = \{H_{ij}(\theta, T)\}$ and
$$H_{ij}(\theta, T) = \int\limits_0^T \mathrm{tr}[\Sigma_t R_t(\theta, L_{ij})]\mathrm{d}t, \qquad (15)$$
$$\begin{cases} \dot{R}_t(\theta, L_{ij}) = \Psi_t(\theta)R(\theta, L_{ij}) + \\ +R_t(\theta, L_{ij})\Psi_t^*(\theta) + \psi_t(\theta)L_{ij}\psi_t^*(\theta), \\ R_0(\theta, L_{ij}) = 0, \end{cases} \qquad (16)$$
$$\begin{aligned} \Psi_t(\theta) &= a_t - K_t(\theta)c_t, \\ \psi_t(\theta) &= K_t(\theta)d_t - b_t, \end{aligned} \qquad (17)$$
and $K_t(\theta)$ is determined by (10), (11).

From the dual criterion definition we derive
$$J_T^0(\gamma) = \mathrm{tr}[H_T^*(\gamma)\gamma], \text{ and } \hat{\gamma} \in \operatorname*{argmax}_{\gamma\in\Gamma} \mathrm{tr}[H_T^*(\gamma)\gamma].$$

The iterative algorithm of the dual problem solution is as follows.

Algorithm 1.
1. Choose some initial value $\gamma^{(0)} \in \Gamma$ and set $k = 0$.
2. Calculate $H_k = H_T(\gamma^{(k)})$ using (15)-(17).
3. Solve the linear programming problem
$$\tilde{\gamma}^{(k)} \in \operatorname*{argmax}_{\gamma\in\Gamma} \mathrm{tr}[H_k^*\gamma].$$
4. Compute $\delta_k = \mathrm{tr}[H_k^*\Delta\gamma^{(k)}]$, where $\Delta\gamma^{(k)} = \tilde{\gamma}^{(k)} - \gamma^{(k)}$.
   If $\delta_k \le 0$, put $\hat{\gamma} = \gamma^{(k)}$ and terminate the iterative process.
   If $\delta_k > 0$, go to step 5.
5. Solve the one-dimensional maximization problem
$$\lambda_k \in \operatorname*{argmax}_{\lambda\in[0,1]} J_T^0(\gamma^{(k)} + \lambda\Delta\gamma^{(k)}).$$
6. Set $\gamma^{(k+1)} = \gamma^{(k)} + \lambda_k\Delta\gamma^{(k)}$, increase k by 1 and go to step 2.

The convergence of the sequence $\{\gamma^{(k)}\}$ to the set of the dual problem (12),(13) solutions, i.e.
$$\Gamma_0 = \operatorname*{argmax}_{\gamma\in\Gamma} \int\limits_0^T \mathrm{tr}[\Sigma_t R_t(\gamma)]\mathrm{d}t,$$
is stated below.

Denote $\rho(x, X) = \inf_{y\in X} \|x - y\|$, i.e. the distance between the point $x \in \mathbb{R}^n$ and the subset $X \subset \mathbb{R}^n$.

Theorem 3. Under the conditions of Theorem 1
1. if the iteration process stops after a finite number k^* of iterations, then $\gamma^{(k^*)} \in \Gamma_0$, and $\hat{J}_T = \int\limits_0^T \mathrm{tr}[\Sigma_t R_t(\gamma^{(k^*)})]\mathrm{d}t$;
2. if $k \to \infty$, then $\rho(\gamma^{(k)}, \Gamma_0) \to 0$, and $\int\limits_0^T \mathrm{tr}[\Sigma_t R_t(\gamma^{(k)})]\mathrm{d}t \to \hat{J}_T$.

The proof of this theorem is mostly the same as the proof of the similar result for discrete systems provided in (Pankov and Miller, 2001).

5. THE GUARANTEED ERROR VARIANCE

In conclusion, let us consider the problem of determination of the estimate $\hat{y}_t$ accuracy for every $t \in [0, T]$. Let $l \in \mathbb{R}^p$ and $\eta_t = (l, \xi_t)$ be a linear combination of the error $\xi_t = \hat{y}_t - y_t$ components. Lemma 1 in Appendix implies $\mathrm{D}[\eta_t] = \mathrm{tr}[g_t^* \gamma]$, where matrix g_t depends on l and $\hat{\gamma}$, and γ is the exact intensity of the process w_t in (1), (2). Since $\gamma \in \Gamma$, the guaranteed value D_t^0 of the variance $\mathrm{D}[\eta_t]$ is the solution of the following linear programming problem:

$$D_t^0 = \max_{\gamma \in \Gamma} \mathrm{tr}[g_t^* \gamma], \quad t \in [0, T].$$

Since Γ is a compact set, there exists $\tilde{\gamma}_t$ such that $D_t^0 = \mathrm{tr}[g_t^* \tilde{\gamma}_t]$. Obviously, $\tilde{\gamma}_t$ depends on $(l,\ t, \hat{\gamma})$ and does not coincide with $\hat{\gamma}$ in general. Therefore, $D_t^0 \geq \mathrm{tr}[g_t^* \hat{\gamma}] = l^* R_t(\hat{\gamma}) l$, $t \in [0, T]$.

In particular, if the set Γ has a "maximal point" γ^0, i.e. $\gamma \leq \gamma^0$, $\forall \gamma \in \Gamma$, then

$$\tilde{\gamma}_t = \hat{\gamma} = \gamma^0, \quad t \in [0, T],$$

so $D_t^0 = l^* R_t(\gamma^0) l$, and $R_t(\gamma^0)$ is the unimprovable guaranteed value of the covariance of the estimation error $\xi_t = \hat{y}_t - y_t$, $\hat{y}_t = \hat{F}(Z^t)$, $\hat{F} = F_K(\gamma^0)$.

6. EXAMPLE

Let the system (1), (2) be stationary on $t \in [0, 1]$ with the following coefficients:

$$a_t = \begin{pmatrix} 0.7 & 0.2 & -0.1 \\ 0.2 & -0.1 & 0.7 \\ 0.3 & 0.0 & 0.6 \end{pmatrix},$$
$$b_t = \begin{pmatrix} 0.7 & 0.1 & 0.2 & 0 & 0 & 0 \\ 0.2 & 0.4 & 0.3 & 0 & 0 & 0 \\ -0.1 & 0.3 & 0.6 & 0 & 0 & 0 \end{pmatrix},$$
$$c_t = \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 1.0 & -0.2 \\ 0.1 & 0.4 & 0.6 \end{pmatrix},$$
$$d_t = \begin{pmatrix} 0 & 0 & 0 & 1.0 & 0.2 & -0.2 \\ 0 & 0 & 0 & -0.2 & 0.3 & 0.3 \\ 0 & 0 & 0 & 0.5 & 0.3 & 0.6 \end{pmatrix}.$$

The initial value $y_0 = 0$.

Assume $\gamma = \mathrm{diag}(S_1, S_2) \in \Gamma$, where $\mathrm{diag}(S_1, S_2)$ denotes a block-diagonal matrix, and the uncertainty set Γ is determined by the following elementwise constraints:

$$\begin{aligned} \Gamma = \{\ & \gamma : \gamma = \mathrm{diag}(S_1, S_2), \\ & \bar{S}_1 - d \cdot E \leq S_1 \leq \bar{S}_1 + d \cdot E, \\ & \bar{S}_2 - d \cdot E \leq S_2 \leq \bar{S}_2 + d \cdot E, \}, \end{aligned}$$

$d = 0.1$, E is a (3×3)-matrix of unities, and

$$\bar{S}_1 = \begin{pmatrix} 0.9 & -0.4 & 0.2 \\ -0.4 & 0.7 & 0.1 \\ 0.2 & 0.1 & 0.6 \end{pmatrix}, \bar{S}_2 = \begin{pmatrix} 0.5 & 0.0 & 0.0 \\ 0.0 & 0.8 & 0.0 \\ 0.0 & 0.0 & 0.1 \end{pmatrix}.$$

The accuracy of the estimate is determined by criterion (6) with $\Sigma_t = 1$, $t \in [0, 1]$. Under the assumptions of Theorem 1 the minimax filtering problem solution is determined by (9)-(11) with $\theta = \hat{\gamma}$, where $\hat{\gamma}$ is the dual problem solution. For obtaining this solution Algorithm 1 was used with the initial conditions $S_1^{(0)} = \bar{S}_1$, $S_2^{(0)} = \bar{S}_2$. The criterion value for $\gamma^{(0)} = \mathrm{diag}(S_1^{(0)}, S_2^{(0)})$ is $J_T(F_K(\gamma^{(0)}), \gamma^{(0)}) = J_T^0(\gamma^{(0)}) = 0.4690$.

The solution obtained is $\hat{\gamma} = \mathrm{diag}(\hat{S}_1, \hat{S}_2)$, where

$$\hat{S}_1 = \begin{pmatrix} 1.0 & -0.4626 & 0.2778 \\ -0.4626 & 0.8 & 0.2 \\ 0.2778 & 0.2 & 0.7 \end{pmatrix},$$
$$\hat{S}_2 = \begin{pmatrix} 0.6 & 0.0566 & 0.0516 \\ 0.0566 & 0.9 & 0.0079 \\ 0.0516 & 0.0079 & 0.2 \end{pmatrix},$$

and the criterion value is $J_T^0(\hat{\gamma}) = 0.5887$.

The robustness of the obtained estimate $\hat{y}_t$ is determined by the guaranteed value D_t^0 of the variance $D_t = \mathrm{D}[\eta_t]$, where $\eta_t = ((1, 0, 0)^*, \xi_t) = \xi_t^1$. In section 5 it is shown that D_t^0 is a solution of the following problem

$$D_t^0 = \max_{\gamma \in \Gamma} \mathrm{tr}[g_t^* \gamma], \quad t \in [0, T], \tag{18}$$

where $g_t = \{g_{ij}(\hat{\gamma}, l, t)\} \in \mathbb{R}^{r \times r}$,

$$g_{ij}(\hat{\gamma}, l, t) = l^* R_t(\hat{\gamma}, L_{ij}) l, \quad t \in [0, T], \tag{19}$$

and $R_t(\cdot)$ is determined by (16),(17).

Note, that the linear programming problem (18) with element-wise constraints on γ has the analytical solution. Let $\Gamma = \{\gamma : \underline{\gamma}_{ij} \leq \gamma_{ij} \leq \overline{\gamma}_{ij}\}$, then $D_t^0 = \mathrm{tr}[g_t^* \tilde{\gamma}_t]$, where

$$\tilde{\gamma}_{ij}(t) = \begin{cases} \overline{\gamma}_{ij}, & g_{ij}(t) \geq 0, \\ \underline{\gamma}_{ij}, & g_{ij}(t) < 0. \end{cases}$$

The evolution of the variance $\mathrm{D}[\eta_t]$ guaranteed value D_t^0 is shown on figure 1.

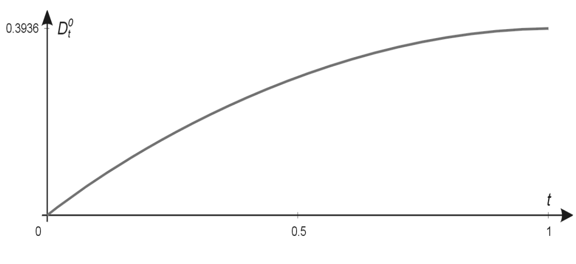

Fig. 1. Guaranteed variance D_t^0 of ξ_t^1.

The matrix function $\tilde{\gamma}_t = \underset{\gamma \in \Gamma}{\mathrm{argmax}}\, \mathrm{tr}[g_t^* \gamma]$, such that $D_t^0 = \mathrm{tr}[g_t^* \tilde{\gamma}]$, is $\tilde{\gamma}_t = \mathrm{diag}(\tilde{S}_1(t), \tilde{S}_2(t))$, where

for $t \in [0, 0.3)$

$$\tilde{S}_1(t) = \begin{pmatrix} 1.0 & -0.3 & 0.3 \\ -0.3 & 0.8 & 0.2 \\ 0.3 & 0.2 & 0.7 \end{pmatrix},$$

$$\tilde{S}_2(t) = \begin{pmatrix} 0.6 & -0.1 & 0.1 \\ -0.1 & 0.9 & -0.1 \\ 0.1 & -0.1 & 0.2 \end{pmatrix};$$

for $t \in (0.3, 0.46)$

$$\tilde{S}_1(t) = \begin{pmatrix} 1.0 & -0.3 & 0.3 \\ -0.3 & 0.8 & 0.2 \\ 0.3 & 0.2 & 0.7 \end{pmatrix},$$

$$\tilde{S}_2(t) = \begin{pmatrix} 0.6 & 0.1 & -0.1 \\ 0.1 & 0.9 & -0.1 \\ -0.1 & -0.1 & 0.2 \end{pmatrix};$$

for $t \in (0.46, 0.47)$

$$\tilde{S}_1(t) = \begin{pmatrix} 1.0 & -0.3 & 0.3 \\ -0.3 & 0.8 & 0.0 \\ 0.3 & 0.0 & 0.7 \end{pmatrix},$$

$$\tilde{S}_2(t) = \begin{pmatrix} 0.6 & 0.1 & -0.1 \\ 0.1 & 0.9 & -0.1 \\ -0.1 & -0.1 & 0.2 \end{pmatrix};$$

for $t \in (0.47, 1]$

$$\tilde{S}_1(t) = \begin{pmatrix} 1.0 & -0.5 & 0.3 \\ -0.5 & 0.8 & 0.0 \\ 0.3 & 0.0 & 0.7 \end{pmatrix},$$

$$\tilde{S}_2(t) = \begin{pmatrix} 0.6 & 0.1 & -0.1 \\ 0.1 & 0.9 & -0.1 \\ -0.1 & -0.1 & 0.2 \end{pmatrix}.$$

The function $\tilde{\gamma}_t$ is piecewise constant on $[0, 1]$, and $\tilde{\gamma}_t \neq \hat{\gamma}$ for all $t \in [0, 1]$.

7. APPENDIX

The auxiliary results considered below are necessary for the theorem 1 proof.

Let the matrix $L_{ij} \in \mathbb{R}^{r\times r}$ has all zero elements, except the element l_{ij} which is 1.

Lemma 1. Let θ, γ, $\hat{\gamma} \in \Gamma$, and

$$I_T(\theta, \gamma) = J_T(F_K(\theta), \gamma),$$

then

1. $I_T(\theta, \gamma) = \mathrm{tr}[H_T^*(\theta)\gamma]$, where $H_T(\theta) = \{H_{ij}(\theta, T)\} \in \mathbb{R}^{r\times r}$,

$$H_{ij}(\theta, T) = \int_0^T \mathrm{tr}[\Sigma_t R_t(\theta, L_{ij})]\mathrm{d}t, \qquad (20)$$

$$\begin{cases} \dot{R}_t(\theta, L_{ij}) = \Psi_t(\theta)R(\theta, L_{ij}) + \\ +R_t(\theta, L_{ij})\Psi_t^*(\theta) + \psi_t(\theta)L_{ij}\psi_t^*(\theta), \\ R_0(\theta, L_{ij}) = 0, \end{cases} \qquad (21)$$

$$\begin{aligned} \Psi_t(\theta) &= a_t - K_t(\theta)c_t, \\ \psi_t(\theta) &= K_t(\theta)d_t - b_t, \end{aligned} \qquad (22)$$

$K_t(\theta)$ is determined by (10), (11).

2. Let $\xi_t = \hat{y}_t - y_t$, $\eta_t = l^*\xi_t$. Then $\mathrm{D}[\eta_t] = \mathrm{tr}[g_t^*\gamma]$, $g_t = \{g_{ij}(\hat{\gamma}, l, t)\} \in \mathbb{R}^{r\times r}$,

$$g_{ij}(\hat{\gamma}, l, t) = l^* R_t(\hat{\gamma}, L_{ij}) l, \quad t \in [0, T]. \quad (23)$$

The proof is rather straightforward, and hence is omitted.

Lemma 2. Let the functional $J_T(F, \gamma)$, $F \in \mathcal{F}$, $\gamma \in \Gamma$, where Γ is a convex subset, satisfy the following conditions

1. $J_T(F, \gamma)$ is concave in γ on Γ for any $F \in \mathcal{F}$;
2. for any $\gamma \in \Gamma$ there exists $\tilde{F}(\gamma) \in \mathcal{F}$ such that
$$\inf_{F\in\mathcal{F}} J_T(F, \gamma) = J_T(\tilde{F}(\gamma), \gamma);$$
3. the solution of the dual problem exists:
$$\hat{\gamma} \in \operatorname*{argmax}_{\gamma\in\Gamma} J_T(\tilde{F}(\gamma), \gamma);$$
4. for any $\gamma \in \Gamma$, and $\hat{F} = \tilde{F}(\hat{\gamma})$ the following property is valid:
$$J_T(\hat{F}, \gamma) = \lim_{\alpha\to 0+} J_T(\tilde{F}(\gamma^\alpha), \gamma),$$
where $\gamma^\alpha = (1-\alpha)\hat{\gamma} + \alpha\gamma$, $\alpha \in [0, 1]$.

Then $(\hat{F}, \hat{\gamma})$ is a saddle point of $J_T(F, \gamma)$ on $\mathcal{F}\times\Gamma$.

The proof of Lemma 2 can be found in (Pankov and Siemenikhin, 2003). Note, that the result of Lemma 2 extends the similar result of (Verdu and Poor, 1984) to the infinite-dimentional case.

Proof of Theorem 1: Let $F_K(\theta)$ be the Kalman-Bucy filtering operator given by (9)-(11) for the case $\theta \in \Gamma$. If $\hat{\gamma} \in \operatorname*{argmax}_{\gamma\in\Gamma} \inf_{F\in\mathcal{F}} J_T(F, \gamma)$, then $\hat{F} \in \operatorname*{argmin}_{F\in\mathcal{F}} J_T(F, \hat{\gamma})$. In this case $\hat{F} = F_K(\hat{\gamma}) \in \mathcal{F}_K$, where $\mathcal{F}_K$ is a set of all filters $F_K(\theta)$, $\theta \in \Gamma$. Hence, if $(\hat{F}, \hat{\gamma})$ is a saddle point of $J_T(F, \gamma)$ on $\mathcal{F}_K \times \Gamma$, then it is a saddle point also on $\mathcal{F} \times \Gamma$. So the case $\mathcal{F} = \mathcal{F}_K$ could be considered without loss of generality. In this case by virtue of Lemma 1 one can obtain

$$J_T(F_K(\theta), \gamma) = I_T(\theta, \gamma) = \mathrm{tr}[H_T^*(\theta)\gamma],$$

where $H_T(\theta)$ is determined by (20)-(22).

If the matrix function $K_t(\theta)$ is continuous with respect to (t, θ) on $[0, T] \times \Gamma$, then from (22) it follows that $\Psi_t(\theta)$ and $\psi_t(\theta)$ are piecewise continuous with respect to $t \in [0, T]$ and continuous with respect to $\theta \in \Gamma$. Then from (21) it follows, that $R_t(\theta, L_{ij})$ are continuous on $[0, T] \times \Gamma$, and, hence, $H_T(\theta)$ is continuous with respect to $\theta \in \Gamma$ by (20) and the definition of Σ_t, $t \in [0, T]$.

For any $\theta \in \Gamma$ the function $I_T(\theta, \gamma) = \mathrm{tr}[H_T^*(\theta)\gamma]$ is linear with respect to $\gamma \in \Gamma$ and, hence, is concave on Γ. Let us show, that the dual optimization problem has a solution.

$$\begin{aligned} J_T^0(\gamma) &= \inf_{F\in\mathcal{F}} J_T(F, \gamma) = \inf_{F\in\mathcal{F}_K} J_T(F, \gamma) = \\ &= J_T(F_K(\gamma), \gamma) = I_T(\gamma, \gamma) = \mathrm{tr}[H_T^*(\gamma)\gamma]. \end{aligned}$$

Hence, J_T^0 is continuous with respect to γ on Γ, as $H_T(\gamma)$ is continuous. The last means that $\hat{\gamma} \in \underset{\gamma\in\Gamma}{\operatorname{argmax}} J_T^0(\gamma)$ exists, since Γ is compact. Note, that form Lemma 1 it follows that $\hat{\gamma} \in \underset{\gamma\in\Gamma}{\operatorname{argmax}} \int_0^T \operatorname{tr}[\Sigma_t R_t(\gamma)]\mathrm{d}t$, where $R_t(\theta)$ is defined by (11).

Now, let $\gamma^\alpha = (1-\alpha)\hat{\gamma} + \alpha\gamma$, $\forall\gamma \in \Gamma$, $\alpha \in (0,1]$.

$$\lim_{\alpha\to 0+} J_T(F_K(\gamma^\alpha),\gamma) = \lim_{\alpha\to 0+} \operatorname{tr}[H_T^*(\gamma^\alpha)\gamma] = \\ = \operatorname{tr}[\lim_{\alpha\to 0+} H_T^*(\gamma^\alpha)\gamma] = \operatorname{tr}[H_T^*(\hat{\gamma})\gamma],$$

since $H_T(\gamma)$ is continuous at $\hat{\gamma} \in \Gamma$. Hence, all conditions of Lemma 2 are fulfilled, and consequently the pair $(\hat{F},\hat{\gamma})$, where $\hat{F} = F_K(\hat{\gamma})$, is a saddle point of $J_T(F,\gamma)$ on $\mathcal{F}\times\Gamma$.

To complete the proof, it is necessary to show that $K_t(\theta)$ is continuous with respect to (t,θ) on $[0,T]\times\Gamma$. Since $K_t(\theta) = (R_t(\theta)c_t^* + b_t\theta d_t^*)(d_t\theta d_t^*)^{-1}$, then $K_t(\theta)$ is piecewise continuous with respect to t and continuous with respect to θ, if the same is valid for $R_t(\theta)$ (note, that $(d_t\theta d_t^*)^{-1}$ is continuous with respect to θ on Γ as follows from the regularity condition (4)). The matrix function $R_t(\theta)$ is a solution of the Riccati equation (11), which can be expressed in the ordinary form

$$\begin{cases} \dot{R}_t(\theta) = (a_t - b_t\theta d_t^*(d_t\theta d_t^*)^{-1}c_t)R_t(\theta) + \\ \quad +R_t(\theta)(a_t - b_t\theta d_t^*(d_t\theta d_t^*)^{-1}c_t)^* + \\ \quad +b_t(\theta - \theta d_t^*(d_t\theta d_t^*)^{-1}d_t\theta)b_t^* - \\ \quad -R_t(\theta)c_t^*(d_t\theta d_t^*)^{-1}c_tR_t(\theta), \\ R_0(\theta) = 0. \end{cases} \quad (24)$$

The solution of (24) can be expressed as a function of the matrix of fundamental solutions of the following system of linear ordinary differential equations:

$$\begin{cases} \dot{\pi}(t) = -(a_t - b_t\theta d_t^*(d_t\theta d_t^*)^{-1}c_t)^*\pi(t) + \\ \quad +c_t^*(d_t\theta d_t^*)^{-1}c_t\mu(t), \\ \dot{\mu}(t) = b_t(\theta - \theta d_t^*(d_t\theta d_t^*)^{-1}d_t\theta)b_t^*\pi(t) + \\ \quad +(a_t - b_t\theta d_t^*(d_t\theta d_t^*)^{-1}c_t)\mu(t). \end{cases} \quad (25)$$

Let

$$\Pi(t,\theta) = \begin{bmatrix} \Pi_{11}(t,\theta) & \Pi_{12}(t,\theta) \\ \Pi_{21}(t,\theta) & \Pi_{22}(t,\theta) \end{bmatrix}$$

be a matrix of system (25) fundamental solutions. The elements of the matrix functions $\Pi_{ij}(t,\theta)$ are continuous with respect to (t,θ) on $[0,T]\times\Gamma$, since the coefficients in the right-hand side of (25) are continuous. It is known that $\Pi_{11}(t,\theta)$ is invertible for any $t \in [0,T]$ and $\theta \in \Gamma$, and $R_t(\theta)$ for the case $R_0(\theta) = 0$ can be expressed as

$$R_t(\theta) = \Pi_{21}(t,\theta)\Pi_{11}^{-1}(t,\theta).$$

From the last formulae one can see that $R_t(\theta)$ is continuous with respect to (t,θ) on $[0,T]\times\Gamma$. This completes the proof of Theorem 1.

REFERENCES

Bertsekas D. and Rhodes I.B. (1971). Recursive state estimation for a set of membership description of uncertainty. *IEEE Trans. Autom. Control* **16**(2), 117–128.

Bobrik G.I., Golovan A.A. and Matasov A.I. (1997). Kalman filter in guaranteeing approach to the topographic binding problem. *Autom. and Rem. Control* (10).

Borisov A.V. and Pankov A.R. (1998). Minimax filtering in dynamic systems described by stochastic differential equations with a measure. *Autom. and Rem. Control* (6).

Davis M.H.A. (1977). *Linear Estimation and Stochastic Control.* Chapman and Hall.

Golubev G.A., Muravlev O.V. and Pisarev V.F. (1989). Linear recursive filtering of discret dynamic processes with incomplete information about disturbing processes. *Autom. and Rem. Control* (12).

Katz I. Ya. and Timofeeva G.A. (1994). The modified method of discrepancy in statistically uncertain estimate problem. *Autom. and Rem. Control* (2).

Li L., Luo Z., Davidson T., Wong K. and Bosse E. (2002). Robust filtering via semidefinite programming with applications to target tracking. *SIAM J. Optim.* **12**(3), 740–755.

Liptser R.Sh. and Shiryaev A.N. (1978). *Statistics of Random Processes.* Springer-Verlag. Berlin.

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

Morris J.M. (1976). The kalman filter: a robust estimator for some classes of linear quadratic problems. *IEEE Trans. Inform. Theory* **22**, 526–534.

Orlov Yu. and Basin M. (1995). On minimax filtering over discrete-continuous observations. *IEEE Trans. Autom. Control* **40**, 1623–1626.

Pankov A.R. and Miller G.B. (2001). Minimax linear recurrent filtering of uncertain-stochastic sequences with integral criterion (in russian only). *Information processes* **1**(2), 150–156.

Pankov A.R. and Siemenikhin K.V. (2003). Minimax estimation of random elements with application to infinite-dimensional statistical linearization. *Proceedings of the II-nd International Conference "Identification and Control Problems"* pp. 1277–1291.

Pugachev V.S. and Sinitsyn I.N. (1987). *Stochastic Differential Systems: Analysis and Filtering.* Wiley. New York.

Sage A.P. and White C.C. (1977). *Optimum systems control.* Prentice Hall.

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