ESTIMATION IN GENERALIZED UNCERTAIN-STOCHASTIC LINEAR REGRESSION

ANDREI V. BORISOV
Department of Industrial & Operations Engineering
University of Michigan
Ann Arbor, MI 48109-2117

Technical Report 91-39

December , 1991.

ESTIMATION IN GENERALIZED UNCERTAIN-STOCHASTIC LINEAR REGRESSION

ANDREI V. BORISOV[1]

(Revised on December, 17, 1991)

Absract. The problem under consideration is to estimate optimally some finite-dimensional vector, which is induced by linear operators, defined on the special class of stochastic and uncertain processes. Optimality of the estimate implies minimization of a minimax-stochastic criterion. We assume a linear model of observations, which contains random disturbances and has a discrete-continuous structure. The optimal estimate must be found as a linear operator of observations.
Necessary and sufficient conditions for the vector identifiability and estimate optimality are proved. The optimal filtering algorithm for uncertain-stochastic differential systems is obtained as an application of this estimation theory.

Key Words. Optimal estimation, generalized linear regression, minimax-stochastic criterion, identifiability of parameters, filtering problem for uncertain-stochastic systems.

AMS(MOS) subject classifications. 93E11 , 93E12.

1.Introduction. Parameter estimation in the linear regression arises often in stochastic control and signal processing theory. Diffrerent approaches may be used to solve this problem. The choice depends on both a priori information about regression parameters and noises and optimality criteria for the estimates.
If both parameters and noises in the model are gaussian, then the estimate, which minimizes mean square estimate error, i.e., conditional expectation is also equal to the best linear estimate (BLE) [1]. In the case parameters and noises are random but nongaussian with given moment characteristics, the BLE may by found by using the same formula as in the gaussian case. There also exists a sufficient condition for the linear estimate to be the BLE - the well-known Wiener-Hopf condition [2]. The estimation problem for pure stochastic regression is solved in [3] and [4] when parameters and noises have singular variance matrices. In [2] the RKHS approach is also considered when parameters and noises are random but contain uncertainties in their moment characteristics.
When it is assumed that parameters and/or noises are uncertain and bounded, a minimax optimality criteria and minimax [5] , [6] or ellipsoidal bounded [7] , [8] estimation methods are usually used. If the parameters are uncertain and unbounded,

[1]Department of Industrial & Operations Engineering, University of Michigan, Ann Arbor MI48109-2117

then the least square estimate (LSE) [9] is the BLE.
In practical situations it is necessary to estimate some vector of parameters in generalized uncertain-stochastic regression [10]. That means the estimated vector is a linear combination of unbounded uncertain and random parts induced by some linear operators from the space of random and uncertain processes respectively into the finite-dimensional space. We assume that observations contain random noises and have a discrete-continuous structure. Similar problems in the pure random case arise by filtering or smoothing of the state in the linear difference, differential and integral dynamic systems [11],[12],[13],[14].
In this article, the necessary and sufficient conditions of the state identifiability and linear estimate optimality for generalized uncertain-stochastic linear regression are given. Problem, which is similar to the identifiability problem, is considered in [15] for pure uncertain differential inclusions. Here, identifiability conditions are obtained for arbitrary linear bounded operators. Presented estimate optimality conditions are generalizations of the Wiener-Hopf condition from pure stochastic to uncertain-stochastic regression.
The filtering problem for linear uncertain-stochastic differential systems given complex observations of both system state and input signals is solved. This algorithm was presented in [16] briefly for the special case of separate observations of the state and inputs. In this article, optimality of this algorithm is proved for the general case as an application of the optimal estimation theory in generalized regression.

2.Preliminaries. Consider a probability space (Ω, F, P), where Ω is a sample space, F is a $\sigma-$algebra on Ω, P is a probability measure on F.
Further, we consider a finite time interval $[0, T]$.

DEFINITION 2.1. The n-dimensional stochastic process $\mu_t = (\mu_t^1, ..., \mu_t^n)^T$ belongs to the class S^n iff the following conditions hold:
i) each component of μ_t $\mu_t^i = \mu_i(\omega, t)$, $\omega \in \Omega$ is F-measurable, $i = 1, ..., n$;
ii) all paths of μ_t^i, $i = 1, ..., n$ belong to the L_2 space of square integrable functions on $[0, T]$;
iii) μ_t is a second-order zero-mean process with known covariance matrix-valued function $cov(\mu_\tau, \mu_\sigma) = R(\tau, \sigma)$, $0 \leq \tau \leq T$, $0 \leq \sigma \leq T$.

The class S^n is quite broad. For example, all n-dimensional zero-mean second-order diffusion processes belong to this class [1].
We will also denote the k-dimensional space of square integrable functions on $[0, T]$ as L_2^k; Φ_0^+ and Φ^+ are the left pseudoinverse matrix and operator respectively, i.e. the matrix and the operator, which satisfy

$$\Phi_0 \Phi_0^+ \Phi_0 = \Phi_0 \ , \ \Phi \Phi^+ \Phi = \Phi.$$

We will consider a square of the following norm of vector a, $||a||^2 = a^T\Sigma a$, where Σ is a known symmetric matrix, $\Sigma \geq 0$.

3.Problem Formulation. Consider the following model of observations

$$\begin{cases} y_0 = \Phi_0 u_0 + \Phi_1 u + H_0\xi_0 + H_1\xi + Q_0\theta_0, \\ y = \Phi_2 u_0 + \Phi u + H_2\xi_0 + H\xi + Q\theta. \end{cases} \tag{1}$$

Here $u_0 \in R^{p_0}$ is an uncertain vector, $\xi_0 \in R^{q_0}$ and $\theta_0 \in R^{r_0}$ are zero-mean random vectors with known variance matrices $cov(\xi_0, \xi_0) = C_0$ and $cov(\theta_0, \theta_0) = P_0$ respectively. Matrices Φ_0, H_0, Q_0 of appropriate dimensions are given. $\{u_\tau\} \in L_2^p$ is an uncertain process, $\{\xi_\tau\} \in S^q$ and $\{\theta_\tau\} \in S^r$ are q- and r-dimensional stochastic processes with known covariance matrix-valued functions $cov(\xi_\tau, \xi_\sigma) = C(\tau, \sigma)$, $cov(\theta_\tau, \theta_\sigma) = P(\tau, \sigma)$. The linear bounded operators $\Phi_1 : L_2^r \to R^{m_0}$, $H_1 : L_2^q \to R^{m_0}$, $\Phi : L_2^p \to L_2^m$, $\Phi_2 : R^{p_0} \to L_2^m$, $H : L_2^q \to L_2^m$, $H_2 : R^{q_0} \to L_2^m$ and $Q : L_2^r \to L_2^m$ and given.
We suppose for simplicity that $\xi_0, \{\xi_\tau\}, \theta_0, \{\theta_\tau\}$ are not correlated.
Consider also a n-dimensional vector

$$x = A_0 u_0 + B_0\xi_0 + Au + B\xi, \tag{2}$$

where matrices A_0, B_0 are given, A and B are known linear operators $L_2^p \to R^n$ and $L_2^q \to R^n$ respectively.
The problem under consideration is to calculate a linear estimate $\hat{x} = M_0 y_0 + My$ of x, which minimizes the following criterion

$$J = \sup_{u_0 \in R^{p_0}, \{u_\tau\} \in L_2^p} \mathbf{E}\{||\hat{\Delta}||^2\}, \tag{3}$$

where $\hat{\Delta} = x - \hat{x}$ is an estimate error, M_0 and M are appropriate matrix and linear operator respectively. Such estimate will be called further *the best linear minimax estimate* (BLME).

4.Necessary and Sufficient Conditions of Vector Identifiability and Estimate Optimality.

DEFINITION 4.1.The vector x given by (2) is *identifiable* by the observations (1) iff there exists some measurable operator κ (may be nonlinear) such that
$\hat{x} = \kappa(y_0, y)$, $J(\hat{x}) < \infty$.

LEMMA 4.1. *If x is identifiable,then* $(\ker\Phi \bigcap \ker\Phi_1) \subseteq \ker\Sigma A$ *and* $(\ker\Phi_0 \bigcap \ker\Phi_2) \subseteq \ker\Sigma A_0$.

Proof of Lemma 4.1: see Appendix A.

This is a generalized necessary condition for the optimality of x.

THEOREM 4.1.*If $\hat{x} = M_0 y_0 + My$ is the BLME of x, then*

$$\begin{cases} \Sigma(A_0 - M_0\Phi_0 - M\Phi_2) = 0, \\ \Sigma(A - M_0\Phi_1 - M\Phi) = 0. \end{cases} \quad (4)$$

Proof of Theorem 4.1: see Appendix B.

Conditions (4) are equivalent to unbiasing the estimate $\hat{x}$: $\Sigma\mathbf{E}\{\hat{\Delta}\} = 0$.
This theorem defines some joint restrictions on the matrices A_0, B_0 and operators A, Φ_1, Φ_2, Φ for the existence of the optimal estimate $\hat{x}$.

THEOEM 4.2. *Let x be identifiable, then $\hat{x}$ is the BLME of x iff the following conditions hold*

$$\begin{cases} \Sigma cov(\hat{\Delta}, (I - \Phi_0\Phi_0^+)(y_0 - \Phi_1\Psi_1^+(y - \Phi_2\Phi_0^+ y_0))) = 0, \\ \Sigma cov(\hat{\Delta}, [(I - \Psi_1\Psi_1^+)(y - \Phi_2\Phi_0^+ y_0)](\tau)) \equiv 0 \quad for\ all\ \tau \in [0, T], \end{cases} \quad (5)$$

where $\Psi_1 = \Phi - \Phi_2\Phi_0^+\Phi_1$.

Proof of Theorem 4.2: see Appendix C.

Conditions (5) are analogues of the Wiener-Hopf conditions of optimal parameter estimation in the pure stochastic linear model [2].
Now let us consider the estimation problem for the pure discrete-time regression

$$y_0 = \Phi_0 u_0 + H_0\xi_0 + Q_0\theta_0, \quad (6)$$

$$x_0 = A_0 u_0 + B_0\xi_0, \quad (7)$$

$$J_0 = \sup_{u_0 \in R^{p_0}} \mathbf{E}\{||x_0 - \hat{x}_0||^2\}, \quad (8)$$

$$\hat{\Delta}_0 = x_0 - \hat{x}_0,$$

to obtain the necessary and sufficient condition for the optimality of $\hat{x}$.

COROLLARY 4.1.*Let x_0 be identifiable and $Q_0 P_0 Q_0^T > 0$,then $\hat{x}_0$ is the BLME iff*

$$\Sigma cov(\hat{\Delta}_0, y_0 - \Phi_0 u_0^*) = 0, \quad (9)$$

where $u_0^* = L_0 y_0$ is any linear estimate of u_0 with the property

$$\mathbf{E}\{\Sigma A_0(u_0 - u_0^*)\} = 0, \quad (10)$$

$$\hat{x} = \{A_0\Phi_0^+ + B_0 C_0 H_0^T R_0^{-1}(I - \Phi_0\Phi_0^+)\} y_0, \quad (11)$$

where $R_0 = H_0 C_0 H_0^T + Q_0 P_0 Q_0^T$, $\Phi_0^+ = (\Phi_0^T R_0^{-1} \Phi_0)^+ \Phi_0^T R_0^{-1}$.

Formula (9) is a special case of (5) and any estimate u_0^*, which satisfies (10) may be obtained as $u_0^* = \Phi_0^+ y_0 + (I - \Phi_0^+ \Phi_0) V_0$ with an arbitrary vector V_0 of appropriate dimension. In the case of uncertain parameters ($H_0 = 0$) we obtain
$\hat{x} = A_0 \Phi_0^+ y_0$
i.e., the MSE. In the case of stochastic parameters ($\Phi_0 = 0$), we obtain
$\hat{x} = B_0 C_0 H_0^T R_0^{-1} y_0$,
i.e., the BLE for a random vector (or corollary from the normal correlation theorem [1] in the gaussian case.)

5.Filtering Problem in Linear Uncertain-Stochastic Systems. In this section filtering algorithm is derived using a necessary and sufficient conditions of estimate optimality in generalized linear regression.
Consider the following dynamic system

$$\begin{cases} dx_t = a(t)x_t dt + b(t)u_t dt + d\xi_t, & t > 0, \\ x_0 = \nu, \end{cases} \tag{12}$$

where $x_t \in R^n$ is a state, $\nu \in R^n$ is an uncertain initial condition, $\{u_\tau\} \in L_2^q$ is a q-dimensional uncertain process, $\{\xi_\tau\}$ is a zero-mean Wiener process with the differential covariance matrix-valued function $C(t)$.
Information concerning $\{x_\tau\}, \{u_\tau\}, \{\xi_\tau\}$ and ν is given by the observations

$$\begin{cases} z_0 = \phi_0 \nu + \omega_0, \\ dy_t = \psi(t)x_t dt + \phi(t)u_t dt + \eta(t)d\xi_t + d\omega_t, \quad t > 0, \quad y_0 = 0, \end{cases} \tag{13}$$

where ω_0 is a zero-mean random vector with known covariance matrix Q_0 , $\{\omega_\tau\}$ is an m-dimensional zero-mean Wiener process with differential covariance matrix-valued function $Q(t)$.
Matrix ϕ_0 is known. The matrix-valued functions $a(t), b(t), C(t), \psi(t), \phi(t), \eta(t), Q(t)$ of appropriate dimensions are given and contain only piecewise continuous elements. The filtering problem under consideration is to calculate the linear estimate $\hat{x}_t$ of x_t, which minimizes the following criterion

$$J_t = \sup_{\nu \in R^n, \{u_\tau\}_0^t \in L_2^q} \mathbf{E}\{||x_t - \hat{x}_t||^2\}, \tag{14}$$

using all observations $\{z_0, y_\tau, 0 < \tau \leq t\}$.

THEOREM 5.1. *Let the system and the observations be given by (12)and(13). Let the following conditions also hold:*
i)$Q(\tau) > 0$ *for all* $\tau \in (0, t]$, $Q_0 > 0$;
ii)$b(\tau)\phi^+(\tau)\phi(\tau) = b(\tau)$ *for all* $\tau \in (0, t]$.

Then the BLME $\hat{x}_t$ is unbiased. $\hat{x}_t$ and its error covariance matrix $k(t)$ are given by the following equations

$$\left\{\begin{array}{l} d\hat{x}_t = a(t)\hat{x}_t dt + [b(t)\phi^+(t) + M(t)(I - \phi(t)\phi^+(t))][dy_t - \psi(t)\hat{x}_t dt], \\ \dot{k}(t) = a(t)k(t) + k(t)a^T(t) + 2b(t)[\phi^T(t)P^{-1}(t)\phi(t)]^+ b^T(t) + \\ M(t)\phi(t)[\phi^T(t)P^{-1}(t)\phi(t)]^+\phi^T(t)M^T(t) + C(t) - \\ (b(t)\phi^+(t) + M(t))\ P(t)\ (b(t)\phi^+(t) + M(t))^T, \end{array}\right. \tag{15}$$

with initial conditions

$$\left\{\begin{array}{l} \hat{x}_0 = k(0)\phi_0^T y_0, \\ k(0) = [\phi_0^T Q_0^{-1}\phi_0]^{-1}, \end{array}\right. \tag{16}$$

where

$$\left\{\begin{array}{l} P(t) = \eta(t)C(t)\eta^T(t) + Q(t) > 0, \\ M(t) = [k(t)\psi^T(t) + C(t)\eta^T(t)]P^{-1}(t). \end{array}\right. \tag{17}$$

Proof of Theorem 5.1: see Appendix D.

The given algorithm is an extension of linear filtering theory from the class of pure stochastic systems to the class of uncertain-stochastic systems, where the classical Kalman filter is useless.

6.Conclusion.

i) In this article, the optimal estimation problem in generalized uncertain-stochastic linear regression has been formulated.

ii) The class of identifiable vectors has been determined and the necessary condition for vector identifiability has been proved. This is a special joint restriction on the estimated vector and observation model.

iii) The necessary and sufficient conditions for the estimate optimality have been derived. They are generalizations of the Wiener-Hopf conditions for the class of uncertain-stochastic regression. We have also obtained a formula of the parameter estimation for the discrete model as a corollary of the general case. The MSE and BLE in the pure stochastic model are special cases of this estimate.

iv) The filtering problem in the linear differential uncertain-stochastic dynamic systems has been solved under general assumptions about the linear model of observation as an illustration of the presented estimation theory in generalized linear regression.

Appendix A. Proof of Lemma 4.1. First we decompose an estimate $\hat{x} = \kappa(y_0, y)$ in the form

$$\hat{x} = \kappa(y_0, y) = \phi(v_0, v_1, v_2, v) + \epsilon(\xi_0, \xi, \theta_0, \theta),$$

where $v_0 = \Phi_0 u_0$, $v_1 = \Phi_1 u$, $v_2 = \Phi_2 u_0$, $v = \Phi u$, $\epsilon(\xi_0, \xi, \theta_0, \theta) = \hat{x} - \phi(v_0, v_1, v_2, v)$ and then fix some value $v_0^*, v_1^*, v_2^*, v^*, u_0^*, u^*$:
$v_0^* = \Phi_0 u_0^*$, $v_1^* = \Phi_1 u^*$, $v_2^* = \Phi_2 u_0^*$, $v = \Phi u^*$.
Let us also denote
$\gamma_1 = \mathbf{E}\{||B\xi + B_0\xi_0 - \epsilon(\xi_0, \xi, \theta_0, \theta)||^2\}$,
$\pi = A_0 u_0^* + Au^* - \Phi(v_0^*, v_1^*, v_2^*, v^*)$,
$U_0 = \{u_0 : v_0^* = \Phi_0 u_0 \ , \ v_2^* = \Phi_2 u_0\}$,
$U = \{u : v_1^* = \Phi_1 u \ , \ v^* = \Phi u\}$,
Then

$$J = \gamma_1 + \sup_{u_0 \in R^{p_0}, \{u_\tau\} \in L_2^p} ||A_0 u_0 + Au - \phi(v_0, v_1, v_2, v)||^2 \geq$$

$$\gamma_1 + \sup_{u_0 \in U_0, u \in U} ||A_0 u_0 + Au - \phi(v_0^*, v_1^*, v_2^*, v^*)||^2 =$$

$$\gamma_1 + ||\pi||^2 + \sup_{w_0 \in \ker \Phi_0 \bigcap \ker \Phi_2, w \in \ker \Phi \bigcap \ker \Phi_1} \{2\pi^T \Sigma(A_0 w_0 + Aw) + ||A_0 w_0 + Aw||^2\}$$

$J_1 = 2\pi^T \Sigma(A_0 w_0 + Aw)$ is a linear operator with respect to w_0 and w, hence

$$J \geq \gamma_1 + ||\pi||^2 + \sup_{w_0 \in \ker \Phi_0 \bigcap \ker \Phi_2, w \in \ker \Phi \bigcap \ker \Phi_1, J_1 \geq 0} ||A_0 w_0 + Aw||^2]$$

Supremum in this sum is finite only in the case $\Sigma(A_0 w_0 + Aw) \equiv 0$ for all $w_0 \in \ker \Sigma A_0$ and $w \in \ker \Sigma A$ i.e.

$$(\ker \Phi \bigcap \ker \Phi_1) \subseteq \ker(\Sigma A),$$

$$(\ker \Phi_0 \bigcap \ker \Phi_2) \subseteq \ker(\Sigma A_0).$$

Lemma 4.1 is proved.

Appendix B. Proof of Theorem 4.1. Let $\hat{x} = M_0 y_0 + My$ be the BLME of vector x, then $J(\hat{x}) < \infty$.

$$J(\hat{x}) = \mathbf{E}||B_0\xi_0 + B\xi - M_0(H_0\xi_0 + H_1\xi + Q_0\theta_0) - M(H_2\xi_0 + H\xi + Q\theta)||^2 +$$

$$\sup_{u_0 \in R^{p_0}, u \in L_2^p} ||(A_0 - M_0\Phi_0 - M\Phi_2)u_0 + (A - M_0\Phi_1 - M\Phi)u||^2]$$

Supremum in this sum is finite only in the case
$\Sigma(A_0 - M_0\Phi_0 - M\Phi_2)u_0 \equiv 0$, for all $u_0 \in R^{p_0}$, and
$\Sigma(A - M_0\Phi_1 - M\Phi)u \equiv 0$, for all $u \in L_2^p$,

which is equivalent to (4).

Theorem 4.1 is proved.

Appendix C. Proof of Theorem 4.2. We will find all M_0 and M which satisfy condition (4) and then select matrix M_0 and operator M optimally, using criterion (3).

First we consider the equation

$$\Sigma(A_0 - M_0\Phi_0 - M\Phi_2) = 0 \tag{18}$$

It is necessary to mention that the composition $\Psi_2 = M\Phi_2 : R^{p_0} \to R^n$ of operators M and Φ_2 is some matrix of appropriate dimensions. Using Lemma 4.1 and the property $\Phi_0\Phi_0^+\Phi_0 = \Phi_0$ it may be shown, that $\Sigma(A_0 - \Psi_2)\Phi_0^+\Phi_0 = \Sigma(A_0 - \Psi_2)$. Then (18) is a system of linear equations with solution [3]

$\mathrm{M}_0 = \Sigma^+\Sigma\{[A_0 - \Psi_2]\Phi_0^+ + Z_0(I - \Phi_0\Phi_0^+)\} + (I - \Sigma^+\Sigma)N_0$,

where Z_0 and N_0 are arbitrary matrices of appropriate dimension. It is easy to verify that criterion (3) does not depend on N_0, hence we can select M_0 in the form

$$M_0 = [A_0 - \Psi_2]\Phi_0^+ + Z_0(I - \Phi_0\Phi_0^+). \tag{19}$$

Substituting M_0 into the second equation (4) we obtain

$$\Sigma[A - A_0\Phi_0^+\Phi_1 - Z_0(I - \Phi_0\Phi_0^+)\Phi_1] = \Sigma M\Psi_1, \tag{20}$$

where $\Psi_1 = \Phi - \Phi_2\Phi_0^+\Phi_1$. Without loss of generality we suppose that $\Psi_1^+ : Im(\Psi_1) \to \ker(\Psi_1)^\perp$ in the one-to-one manner and $\Psi_1^+[(Im(\Psi_1))^\perp] \equiv 0$. Otherwise, we can transform Ψ_1^+ to such form.

All $u \in L_2^p$ can be decomposed uniquely as $u = u^k + u^\perp$ where $u^k \in \ker\Psi_1$, $u^\perp \in (\ker\Psi_1)^\perp$. Then using Lemma 4.1 and the property

$$\Psi_1\Psi_1^+\Psi_1 = \Psi_1, \tag{21}$$

we can show, that

$\Sigma\Psi\Psi_1^+\Psi_1 = \Sigma\Psi$

where $\Psi = A - A_0\Phi_0^+\Phi_1 - Z_0(I - \Phi_0\Phi_0^+)\Phi_1$.

Operator M may be rewritten in the form $M = \Psi\Psi_1^+ + \tilde{M}$, where $\tilde{M}$ must be found. From (20) it follows that $\tilde{M}$ satisfies the following equation

$$\Sigma\tilde{M}\Psi_1 = 0, \tag{22}$$

This is equivalent to $Im\Psi_1 \subseteq \ker(\Sigma\tilde{M})$. Using (21) and (22), we can obtain following properties of $\tilde{M}$ and Ψ_1

$\Sigma\tilde{M}v = \Sigma\tilde{M}v^\perp$,

(I-$\Psi_1\Psi_1^+)v = v^\perp$,

where $v \in L_2^m$, $v = v^I + v^\perp$, $v^I \in Im\Psi_1$, $v^\perp \in (Im\Psi_1)^\perp$.

Similarly to the common solution of (18) we obtain
$M=\Sigma^+\Sigma\{\Psi\Psi_1^+ + Z(I-\Psi_1\Psi_1^+)\} + (I-\Sigma^+\Sigma)N$,
where $Z, N : L_2^m \to R^n$ are arbitrary operators.
Again, criterion (3) does not depend on the choise of N. Hence, we can consider operator M in the form

$$M = \Psi\Psi_1^+ + Z(I - \Psi_1\Psi_1^+).$$

Then general solution of (4) is

$$\left\{\begin{array}{l} M = (A - A_0\Phi_0^+\Phi_1)\Psi_1^+ - Z_0(I-\Phi_0\Phi_0^+)\Phi_1\Psi_1^+ + Z(I-\Psi_1\Psi_1^+), \\ M_0 = A_0\Phi_0^+ - (A - A_0\Phi_0^+\Phi_1)\Psi_1^+\Phi_2\Phi_0^+ + \\ Z_0(I-\Phi_0\Phi_0^+)[I + \Phi_1\Psi_1^+\Phi_2\Phi_0^+] - Z(I-\Psi_1\Psi_1^+)\Phi_2\Phi_0^+. \end{array}\right. \quad (23)$$

Suppose that $\hat{Z}_0$ and $\hat{Z}$ are the optimal matrix and operator respectively, i.e.
$\hat{x} = M_0(\hat{Z}_0, \hat{Z})y_0 + M(\hat{Z}_0, \hat{Z})y$ is the BLME. For arbitrary
$Z_0 = \hat{Z}_0 + \delta Z_0$ and $Z = \hat{Z} + \delta Z$ we have the corresponding estimate and criterion value,
$\tilde{x}=\hat{x} + \delta Z_0(I-\Phi_0\Phi_0^+)[y_0 - \Phi_1\Psi_1^+(y - \Phi_2\Phi_0^+y_0)] + \delta Z(I-\Psi_1\Psi_1^+)(y - \Phi_2\Phi_0^+y_0)$,
$J(\tilde{x})=J_1(\hat{Z}_0,\hat{Z}) + Z_2(\delta Z_0, \delta Z) - Z_3(\hat{Z}_0, \delta Z_0) - J_4(\hat{Z}, \delta Z)$,
where
$J_1(\hat{Z}_0, \hat{Z}) = \mathbf{E}\{||\hat{\Delta}||^2\} = J(\hat{x})$,

$$\begin{array}{l} J_2(\delta Z_0, \delta Z) = \mathbf{E}\{||\delta Z_0(I-\Phi_0\Phi_0^+)[y_0 - \Phi_1\Psi_1^+(y-\Phi_2\Phi_0^+y_0)] + \\ \delta Z(I-\Psi_1\Psi_1^+)(y - \Phi_2\Phi_0^+y_0)||^2\} \geq 0, \end{array} \quad (24)$$

$$J_3(\hat{Z}_0, \delta Z_0) = 2tr\{\Sigma cov(\hat{\Delta}, \delta Z_0(I-\Phi_0\Phi_0^+)[y_0 - \Phi_1\Psi_1^+(y - \Phi_2\Phi_0^+y_0)])\}, \quad (25)$$

$$J_4(\hat{Z}, \delta Z) = 2tr\{\Sigma cov(\hat{\Delta}, \delta Z(I-\Psi_1\Psi_1^+)(y - \Phi_2\Phi_0^+y_0)\}. \quad (26)$$

If conditions

$$\left\{\begin{array}{l} J_3(\hat{Z}_0, \delta Z_0) \equiv 0, \\ J_4(\hat{Z}, \delta Z) \equiv 0, \end{array}\right. \quad (27)$$

do not hold for some δZ_0 and δZ then from (24)-(26) it follows that there exists the estimate
$x^* = M_0(Z_0^*, Z^*)y_0 + M(Z_0^*, Z^*)y$:

$$\left\{\begin{array}{l} Z_0^* = 0.5[(J_3(\hat{Z}_0, \delta Z_0) + J_4(\hat{Z}, \delta Z))/J_2(\delta Z_0, \delta Z)]\delta Z_0 + \hat{Z}_0, \\ Z^* = 0.5[(J_3(\hat{Z}_0, \delta Z_0) + J_4(\hat{Z}, \delta Z))/J_2(\delta Z_0, \delta Z)]\delta Z + \hat{Z}, \end{array}\right.$$

and $J(x^*) - J(\hat{x}) = -0.25(J_3(\hat{Z}_0, \delta Z_0) + J_4(\hat{Z}, \delta Z))^2 < 0$.
Hence $J(x^*) < J(\hat{x})$, or $\hat{x}$ is not optimal. So (27) is necessary for optimality of $\hat{x}$.
Now let us suppose that (27) holds, then for all $\tilde{x}$
$J(\tilde{x}) = J_1(\hat{Z}_0, \hat{Z}) + Z_2(\delta Z_0, \delta Z) \geq J(\hat{x})$. Hence, $\hat{x}$ is the BLME. Furthermore, it is

easy to verify that (27) is equivalent to (5).

Theorem 4.2 is proved.

Appendix D. Proof of Theorem 5.1. First we prove that $\hat{x}_t$ is unbiased. Denote $m_t = \mathbf{E}\{x_t\}$, $\hat{m}_t = \mathbf{E}\{\hat{x}_t\}$, $\bar{\Delta}_t = m_t - \hat{m}_t = \mathbf{E}\{\hat{\Delta}_t\}$. $\hat{x}_0$ in (16) is the MSE, hence it is unbiased $\bar{\Delta}_0 = 0$. Furthermore,
$dy_t - \psi(t)\hat{x}_t dt = \psi(t)\hat{\Delta}_t dt + \phi(t)u_t dt + \eta(t)d\xi_t + d\omega_t.$
Using the first equation (15) and condition ii) of Theorem 5.1 we can obtain the differential equation for $\hat{m}_t$:
$d\hat{m}_t = a(t)\hat{m}_t dt + b(t)u_t dt + [b(t)\phi^+(t) + M(t)(I - \phi(t)\phi^+(t))]\psi(t)\bar{\Delta}_t dt.$
Using (12) we can also obtain differential equation for m_t
$dm_t = a(t)m_t dt + b(t)u_t dt.$
Then

$$\begin{cases} d\bar{\Delta}_t = \{a(t) - [b(t)\phi^+(t) + M(t)(I - \phi(t)\phi^+(t))]\psi(t)\}\bar{\Delta}_t dt, \\ \bar{\Delta}_0 = 0. \end{cases} \tag{28}$$

Equation (28) has a unique solution $\bar{\Delta}_t \equiv 0$, hence $\hat{x}_t$ is an unbiased estimate of x_t. Equations (12),(13) may be rewritten in the equivalent integral form [1]

$$x_t = \Phi(t,0)\nu + \int_0^t \Phi(t,\tau)b(\tau)u_\tau d\tau + \int_0^t \Phi(t,\tau)d\xi_\tau,$$

$$y_t = \int_0^t [\psi(\tau)x_\tau + \phi(\tau)u_\tau]d\tau + \int_0^t \eta(\tau)d\xi_\tau + \int_0^t d\omega_\tau,$$

$$z_0 = \phi_0\nu_0 + \omega_0,$$

where $\Phi(t,\tau)$ is a solution of the following differential equation
$\Phi_t'(t,\tau) = a(t)\Phi(t,\tau)$, $t > \tau$, $\Phi(\tau,\tau) = I$.
Processes $\{x_\tau\}$, $\{y_\tau\}$ have a.s. continuous paths, so these paths are square integrable in the time interval $[0,t]$. It is necessary also to mention that all operators, which transform ν , $\{u_\tau\}$, $\{\xi_\tau\}$, $\{\omega_\tau\}$ and ω_0 are linear and bounded. Let us state conformity between generalized regression (1) - (2) and system (12) - (13):

$$\Phi_0 u_0 = \phi_0\nu,$$

$$(\Phi_1 u)(\mu) \equiv 0,$$

$$(\Phi_2 u_0)(\mu) = \int_0^\mu \psi(\tau)\Phi(\tau,0)\nu\, d\tau,$$

$$(\Phi u)(\mu) = \int_0^\mu \psi(\tau)\int_0^\tau \Phi(\tau,\sigma)b(\sigma)u_\sigma d\sigma d\tau + \int_0^\mu \phi(\tau)u_\tau d\tau$$

$$H_0\xi_0 = 0,$$

$$(H_1\xi)(\mu) \equiv 0,$$

$$(H_2\xi_0)(\mu) \equiv 0,$$

$$(H\xi)(\mu) = \int_0^\mu \psi(\tau) \int_0^\tau \Phi(\tau,\sigma)d\xi_\sigma d\tau + \int_0^\mu \eta(\tau)d\xi_\tau,$$

$$Q_0\theta_0 = \omega_0,$$

$$(Q\theta)(\mu) = \int_0^\mu d\omega_\tau,$$

$$A_0u_0 = \Phi(t,0)\nu,$$

$$Au = \int_0^t \Phi(t,\tau)b(\tau)u_\tau d\tau,$$

$$B_0\xi_0 = 0,$$

$$B\xi = \int_0^t \Phi(t,\tau)d\xi_\tau.$$

So filtering problem (12) - (14) is a special case of the estimation problem (1) - (3) and conditions (5) are

$$\begin{cases} \Sigma cov(\hat{\Delta}_t, (I - \Phi_0\Phi_0^+)y_0) = 0, \\ \Sigma cov(\hat{\Delta}_t, [(I - \Phi\Phi^+)(y - \Phi_2\Phi_0^+ y_0)](\tau)) \equiv 0, \quad \tau \in [0,t] \end{cases} \tag{29}$$

Using (15) and (16), we obtain

$$\hat{\Delta}_t = \Lambda(t,0)\hat{\Delta}_0 + \int_0^t \Lambda(t,\tau)\{d\xi_\tau -$$

$$[b(\tau)\phi^+(\tau) + M(\tau)(I - \phi(\tau)\phi^+(\tau))][\eta(\tau)d\xi_\tau + d\omega_\tau]\},$$

$$\hat{\Delta}_0 = \phi_0^+\omega_0,$$

$$\big(I - \Phi_0\Phi_0^+)y_0 = (I - \phi_0\phi_0^+)\omega_0,$$

where $\Lambda(t,\tau)$ is a solution of the following equation
$\Lambda_t'(t,\tau) = \{a(t) - [b(t)\phi^+(t) + M(t)(I - \phi(t)\phi^+(t)]\psi(t)\}\Lambda(t,\tau), \ t > \tau$
$\Lambda(\tau,\tau) = I$.
Then
$\Sigma cov(\hat{\Delta}_t, (I - \phi_0\phi_0^+)\omega_0) = \Lambda(t,0)\phi_0^+ Q_0(I - \phi_0\phi_0^+)^T = 0,$
i.e., the first condition (29) holds. Furthermore,

$$\hat{\Delta}_t = \Lambda(t,\tau)\hat{\Delta}_\tau + \int_\tau^t \Lambda(t,\sigma)d\xi_\sigma -$$

$$\int_\tau^t \Lambda(t,\sigma)[b(\sigma)\phi^+(\sigma) + M(\sigma)(I - \phi(\sigma)\phi^+(\sigma))][\eta(\sigma)d\xi_\sigma + d\omega_\sigma],$$

and

$$\rho_T = [(I - \Phi\Phi^+)(y - \Phi_2\Phi_0^+ y_0))](\tau) =$$

$$\int_0^T (I - \phi(\sigma)\phi^+(\sigma))[\psi(\sigma)(\hat{x}_\sigma - x_\sigma)d\sigma + \eta(\sigma)d\xi_\sigma + d\omega_\sigma].$$

Note, that $cov(\hat{\Delta}_t, \hat{\Delta}_s) = \Lambda(t,s)cov(\hat{\Delta}_s, \hat{\Delta}_s)$ for all $t, s \geq 0$. Then

$$cov(\hat{\Delta}_t, \rho_T) = \int_0^T cov(\hat{\Delta}_t, d\rho_\sigma) =$$

$$\int_0^T cov\{\{\Lambda(t,\tau)\hat{\Delta}_\tau + \int_\tau^t \Lambda(t,\sigma)d\xi_\sigma -$$

$$\int_\tau^t \Lambda(t,\mu)[b(\mu)\phi^+(\mu) + M(\mu)(I - \phi(\mu)\phi^+(\mu))][\eta(\mu)d\xi_\mu + d\omega_\mu]\},$$

$$\{(I - \phi(\sigma)\phi^+(\sigma))[\psi(\sigma)(\hat{x}_\sigma - x_\sigma)d\sigma + \eta(\sigma)d\xi_\sigma + d\omega_\sigma]\}\} =$$

$$\int_0^t \Lambda(t,\sigma)\{k(\sigma)\psi^T(\sigma) + C(\sigma)\eta^T(\sigma) - M(\sigma)R(\sigma)\}\{I - \phi(\sigma)\phi^+(\sigma)\}^T d\sigma \equiv 0$$

for all $\tau \in [0,t]$. The second condition (29) holds.

Theorem 5.1 is proved.

REFERENCES

[1] **Liptser,R.S.,and Shiryayev,A.N.,***Statistics of Random Processes,* Springer-Verlag,New York, 1977.

[2] **Barton, R.J., and Poor,H.V.,** *An RKHS Approach to Robust L^2 Estimation and Signal Detection,* IEEE Transactions on Information Theory, 36, (1990), pp.485-501.

[3] **Albert,A.E.,** *Regression and the Moore-Penrose Pseudoinverse,* Academic Press, New York, 1972.

[4] **Catlin,D.E.,***Estimation of Random States in General Linear Models,* IEEE Transactions on Automatic Control,36, (1991), pp.248-252.

[5] **Fomin,V.N.,***Recursive Estimation and Adaptive Filtering,* Nauka, Moscow, 1984.

[6] **Kurzhanskii, A.B.,***Control and Observation under Uncertainty,* Nauka, Moscow, 1977.

[7] **Chernous'ko,F.L.,***Estimation of the Phase State of Dynamic Systems. Method of Ellipsoids,* Nauka, Moscow,1988.

[8] **Rao,A.K.,and Huang,Y.-F.,***Recent Developments in Optimal Bounding Ellipsoidal Parameter Estimation,* Mathematics and Computers in Simulation,32, (1990), pp.515-526. [9] **Anderson,T.W.,** *Statistical Analysis of Time Series,* Wiley, New York, 1971.

[10] **Katz,I.Ya.,** *Minimax-Stochastic Problems of Estimation in Multistage Systems,* Estimation under Uncertainty, Ural Scientific Center, Sverdlovsk, 1982, pp.43-59.

[11] **Kalman,R.,and Bucy,R.S.,***New Results in Linear Filtering and Prediction theory,* Transactions of ASME, Journal of Basic Engineering, 82, (1960), pp.35-45.

[12] **Mayne,D.Q.,***A Solution of the Smoothing Problem for the Linear Dynamic Systems,* Automatica, 6,(1966), pp.73-92.

[13] **Astrom,K.J.,** *Introduction to Stochastic Control Theory,* Academic Press, New York, 1970.

[14] **Kleptsina,M.L.,and Veretennikov,A.Yu.,***On Filtering and Properties of Conditional Laws of Ito-Volterra Processes,* Statistics and Control of Stochastic Processes, Steklov Seminar, N.V.Krylov et al, eds., Optimization Software Inc., 1985.

[15] **Aubin,J.-P., and Frankowska, H.,** *Observability of System Under Uncertainty,* SIAM Journal of Control and Optimization, 27, (1989), pp.949-975.

[16] **Borisov,A.V.,and Pankov,A.R.,***Optimal Signal Processing for Uncertain-Stochastic Systems,* Proceedings of the 30th Conference on Detection and Control, Brighton, 3,(1991), pp.3082-3083.