2010 American Control Conference
Marriott Waterfront, Baltimore, MD, USA
June 30-July 02, 2010
WeA17.1

Linear Minimax Estimation for Random Vectors with Parametric Uncertainty

Eilyan Bitar, Enrique Baeyens, Andrew Packard, and Kameshwar Poolla

***Abstract*— In this paper, we take a minimax approach to the problem of computing a *worst-case* linear mean squared error (MSE) estimate of X given Y, where X and Y are jointly distributed random vectors with parametric uncertainty in their distribution. We consider two uncertainty models, $\mathcal{P}_A$ and $\mathcal{P}_B$. Model $\mathcal{P}_A$ represents X and Y as jointly Gaussian whose covariance matrix Λ belongs to the convex hull of a set of m known covariance matrices. Model $\mathcal{P}_B$ characterizes X and Y as jointly distributed according to a Gaussian mixture model with m known zero-mean components, but unknown component weights. We show: (a) the linear minimax estimator computed under model $\mathcal{P}_A$ is identical to that computed under model $\mathcal{P}_B$ when the vertices of the uncertain covariance set in $\mathcal{P}_A$ are the same as the component covariances in model $\mathcal{P}_B$, and (b) the problem of computing the linear minimax estimator under either model reduces to a semidefinite program (SDP). We also consider the dynamic situation where $x(t)$ and $y(t)$ evolve according to a discrete-time LTI state space model driven by white noise, the statistics of which is modeled by $\mathcal{P}_A$ and $\mathcal{P}_B$ as before. We derive a recursive linear minimax filter for $x(t)$ given $y(t)$.**

I. INTRODUCTION

The problem of filtering noisy measurements generated by linear discrete-time dynamical systems with Gaussian additive noise has been studied extensively in the literature beginning with the seminal work of Kalman [4]. A subset of this work, *robust filtering*, explicitly accounts for uncertainty in the system model [1], [2], [3], [7], [9], [10], often assuming that the underlying noise model is accurate. However, in practice this assumption may be suspect, which can lead to poor worst-case filter performance. The work in [8] considers the problem of designing a linear minimax filter for a linear Gaussian system with uncertainty in the covariances of the process and measurement noise, where the covariance matrices are assumed to live a convex set and offers heuristics for computing the set of least favorable covariances.

In this paper we treat uncertainty in the noise model by allowing the noise covariance matrix to be unknown but confined to a convex polyhedral set with known vertices. For the static case of two jointly Gaussian random vectors, X and Y, we show that problem of computing a linear minimax estimate of X given Y under squared-l^2 loss can be reformulated as a semidefinite program (SDP). Moreover, we connect these results to the alternative static case where X and Y are modeled as being jointly distributed according to a Gaussian mixture model with known zero-mean components, but unknown component weights.

We also consider the dynamic situation where the stochastic processes, $\{x(t)\}$ and $\{y(t)\}$, evolve according to a discrete-time LTI state space model with additive noise in the state-transition and measurement maps. We extend the results in the static case to derive a recursive linear minimax filter for $x(t)$ given $y(t)$.

Supported in part by OOF991-KAUST US LIMITED under award number 025478, the UC Discovery Grant ele07-10283 under the IMPACT program, NASA Langley NRA NNH077ZEA001N, and NSF under Grant EECS-0925337.

Bitar, Packard, Poolla are with Mechanical Engineering, U.C. Berkeley {ebitar,poolla,pack}@berkeley.edu

Baeyens is with Systems Engineering and Automatic Control, Universidad de Valladolid enrbae@eis.uva.es

II. PROBLEM FORMULATION

Consider two jointly distributed random vectors $X \in \mathbb{R}^{n_x}$ and $Y \in \mathbb{R}^{n_y}$. Define $Z \doteq \begin{bmatrix} X^*, & Y^* \end{bmatrix}^*$. We assume that distribution of Z belongs to the parametric family of probability distributions $\mathcal{P}_\theta = \{P_\theta \ : \ \theta \in \Theta\}$, where Θ is some convex set.

As the parameter $\theta \in \Theta$ is unknown, we address the problem of computing a linear minimax estimate of X given Y with respect to the standard l^2 loss function. More precisely, we consider the optimization problem

$$J(K^\circ) = \min_{K \in \mathcal{K}} \max_{\theta \in \Theta} \mathbf{E}_\theta \, \|X - KY\|_2^2 \qquad (1)$$

where expectation is taken with respect to $p(x, y; \theta)$ and $\mathcal{K} = \mathbb{R}^{n_x \times n_y}$. In our exposition, we investigate and establish connections between two commonly employed uncertainty models.

A. Uncertainty Model $\mathcal{P}_A$

Model $\mathcal{P}_A$ represents the family of zero-mean Gaussian distributions $\mathcal{N}(0, \Lambda)$ whose covariance matrices Λ are constrained to live in the convex hull of a set of m covariance matrices. More, precisely, $\Lambda \in \mathbf{\Lambda} \doteq \mathsf{Co}\{\Lambda_1, \cdots, \Lambda_m\}$. Hence, $\mathcal{P}_A$ is defined as

$$\mathcal{P}_A \doteq \{\mathcal{N}(0, \Lambda) \ : \ \Lambda \in \mathbf{\Lambda}\} \qquad (2)$$

B. Uncertainty Model $\mathcal{P}_B$

Model $\mathcal{P}_B$ represents the the family of Gaussian mixture models where each component $i \in \{1, \cdots, m\}$ is zero-mean with covariance Λ_i. More precisely,

$$\mathcal{P}_B \doteq \left\{ \sum_{i=1}^{m} \alpha_i \mathcal{N}(0, \Lambda_i) : \ \sum_{i=1}^{m} \alpha_i = 1, \ \alpha_i \geq 0 \right\} \qquad (3)$$

III. Main Results: Static Case

Define the linear minimax estimation problems under uncertainty models $\mathcal{P}_A$ and $\mathcal{P}_B$ respectively as:

$$J_A(K_A^\circ) = \min_{K\in\mathcal{K}} \max_{\Lambda\in\mathbf{\Lambda}} \mathbf{E}_A \|X - KY\|_2^2 \quad (4)$$

$$J_B(K_B^\circ) = \min_{K\in\mathcal{K}} \max_{\alpha\in\mathcal{A}} \mathbf{E}_B \|X - KY\|_2^2 \quad (5)$$

where $\mathcal{A} = \{a \in \mathbb{R}_+^m \ : \ \sum_{i=1}^m a_i = 1\}$ is a probability simplex and $\mathbf{\Lambda} = \mathsf{Co}\{\Lambda_1, \cdots, \Lambda_m\}$ is a convex polyhedral set.

A. Equivalence

Theorem 3.1: The estimations problems under uncertainty models $\mathcal{P}_A$ and $\mathcal{P}_B$ are equivalent:

$$J_A(K_A^\circ) = J_B(K_B^\circ) \quad \text{and} \quad K_A^\circ = K_B^\circ$$

Proof: Fix K. To establish this result, it suffices to show that

$$\max_{\Lambda\in\mathbf{\Lambda}} \mathbf{E}_A \|X - KY\|_2^2 = \max_{\alpha\in\mathcal{A}} \mathbf{E}_B \|X - KY\|_2^2$$

Taking expectation with respect to a distribution in $\mathcal{P}_A$, consider the following

$$\begin{aligned}\mathbf{E}\|X - KY\|_2^2 &= \mathbf{E}\left[(X-KY)^*(X-KY)\right] \\ &= \text{Tr} \begin{bmatrix} I \\ -K^* \end{bmatrix}^* \Lambda \begin{bmatrix} I \\ -K^* \end{bmatrix}\end{aligned}$$

where $\mathbf{E}\left[ZZ^*\right] = \Lambda$. Define $\phi(K,\Lambda) \doteq \text{Tr}\begin{bmatrix} I & -K\end{bmatrix}\Lambda\begin{bmatrix} I & -K\end{bmatrix}^*$. Maximizing $\phi(K,\Lambda)$ with respect to Λ, we obtain

$$\max_{\Lambda\in\mathbf{\Lambda}} \mathbf{E}\|X - KY\|_2^2 = \max_{\Lambda\in\mathbf{\Lambda}} \phi(K,\Lambda) = \max_{i\in\mathcal{I}} \phi(K,\Lambda_i) \quad (6)$$

where $\mathcal{I} = \{1, \cdots, m\}$. The equality above follows from the fact [6] that the maximum of a convex function over a convex polyhedral set $\mathbf{\Lambda}$ is achieved at some vertex, i.e. for some $\Lambda \in \{\Lambda_1, \cdots, \Lambda_m\}$.

We now consider uncertainty model $\mathcal{P}_B$. Taking expectation with respect to a fixed distribution in $\mathcal{P}_B$:

$$\mathbf{E}\|X - KY\|_2^2 = \phi\left(K, \sum_{i=1}^m \alpha_i\Lambda_i\right) = \sum_{i=1}^m \alpha_i\phi(K,\Lambda_i)$$

The expression above is linear in α which is confined to a convex set $\mathcal{A}$. Consequently, we have that $\mathbf{E}\|X - KY\|_2^2$ achieves its maximum at one of the vertices of the probability simplex $\mathcal{A}$, yielding

$$\max_{\alpha\in\mathcal{A}} \sum_{i=1}^m \alpha_i\phi(K,\Lambda_i) = \max_{i\in\mathcal{I}} \phi(K,\Lambda_i) \quad (7)$$

which agrees with (6), completing the proof. ■

B. Reformulation as a SDP

Thus far, we have shown that $J_A(K_A^\circ) = J_B(K_B^\circ) = J(K^\circ)$ where

$$J(K^\circ) = \min_{K\in\mathcal{K}} \max_{i\in\mathcal{I}} \phi(K,\Lambda_i) \quad (8)$$

We now show that optimization problem (8) can be reformulated as a SDP. Observe that

$$\max_{i\in\mathcal{I}} \phi(K,\Lambda_i) = \min_{\gamma} \gamma \ : \ \phi(K,\Lambda_i) \le \gamma \ \ \forall i \in \mathcal{I}$$

which yields

$$J(K^\circ) = \min_{K\in\mathcal{K},\gamma} \gamma \ : \ \phi(K,\Lambda_i) \le \gamma \ \ \forall\, i \in \mathcal{I} \quad (9)$$

Introducing the matrices $W, M \succ 0$ such that

$$M^{-1} \succeq \Lambda_i \text{ and } W \succeq \begin{bmatrix} I \\ -K^* \end{bmatrix}^* \Lambda_i \begin{bmatrix} I \\ -K^* \end{bmatrix}$$

for all i, problem (9) can be reformulated as the following SDP:

$$\begin{aligned} J(K^\circ) = \min_{K,\gamma,M,W} \gamma \ &: \ \gamma \ge \text{Tr}\, W \\ &\Lambda_i^{-1} \succeq M \ \ \forall i \in \mathcal{I} \\ &W \succ 0 \\ M = &\begin{bmatrix} M_{11} & M_{12} \\ M_{12}^* & M_{22} \end{bmatrix} \succ 0 \\ &\begin{bmatrix} W & I & -K \\ I & M_{11} & M_{12} \\ -K^* & M_{12}^* & M_{22} \end{bmatrix} \succeq 0 \end{aligned} \quad (10)$$

IV. Main Results: Dynamic Case

Consider

$$x(t) = Ax(t-1) + w(t) \quad (11)$$

$$y(t) = Cx(t) + v(t) \quad (12)$$

where $x(t) \in \mathbb{R}^{n_x}$, $y(t) \in \mathbb{R}^{n_y}$. The zero-mean white stochastic processes, $\{w(t)\}$ and $\{v(t)\}$, model process and measurement uncertainty, respectively. The distribution on the initial condition $x(0)$ is assumed to be known. We assume that the marginal probability distribution on $z(t) \doteq [w(t)^*, v(t)^*]^*$ is unknown, but constrained to live in a parametric family, $\mathcal{P}_\theta$. We constrain the minimax filter, $\hat{x}(t)$, to belong to the family of linear observers described by

$$\hat{x}(t) = A\hat{x}(t-1) + K(t)\left(y(t) - \hat{y}(t)\right) \quad (13)$$

$$\hat{y}(t) = CA\hat{x}(t-1) \quad (14)$$

where $K(t) \in \mathcal{K}$ is the gain on the innovation at time t. The estimation error, $e(t) \doteq x(t) - \hat{x}(t)$, evolves as

$$e(t) = F(t)e(t-1) + G(t)z(t) \quad (15)$$

where

$$F(t) \doteq (I - K(t)C)A \quad (16)$$

$$G(t) \doteq \begin{bmatrix}(I - K(t)C), & -K(t)\end{bmatrix} \quad (17)$$

For a particular $\theta \in \Theta$, the error covariance matrix, $P_\theta(t) \doteq \mathbf{E}_\theta\left[e(t)e(t)^*\right]$, evolves as

$$P_\theta(t) = F(t)P_\theta(t-1)F(t)^* + G(t)\Lambda_\theta G(t)^* \qquad (18)$$

where $\Lambda_\theta \doteq \mathbf{E}_\theta\left[z(t)z(t)^*\right]$. Iterating (18) backwards to the initial condition, P_0, gives

$$P_\theta(t) = \mathbf{F}(1:t)P_0\mathbf{F}(1:t)^* + G(t)\Lambda_\theta G(t)^* + \sum_{\tau=1}^{t-1}\mathbf{F}(\tau+1:t)G(\tau)\Lambda_\theta G(\tau)^*\mathbf{F}(\tau+1:t)^*$$

where $\mathbf{F}(s:t) \doteq \prod_{\tau=s}^{t} F(\tau)$ for $0 < s \leq t$.

A. Problem Formulation

We are interested in computing a sequence of filter gain matrices, $\{K(t)\}$, that minimize the expected l^2 norm of the estimation error $e(t)$ for the worst-case probability density in the parametric family $\mathcal{P}_\theta$ at each time t. More precisely, we are interested in solving problem (19) at each time t.

$$J_\theta(K_\theta(t)^\circ) = \min_{K(t)\in\mathcal{K}} \max_{\theta\in\Theta} \mathbf{E}_\theta \left\|e(t)\right\|_2^2 \qquad (19)$$

Continuing with the theme of the paper we consider two uncertainty models, $\mathcal{P}_A$ and $\mathcal{P}_B$, described in equations (2) and (3) respectively.

B. Equivalence

We define the linear minimax estimation problems at time t under uncertainty models $\mathcal{P}_A$ and $\mathcal{P}_B$ as

$$J_A(K_A(t)^\circ) = \min_{K\in\mathcal{K}} \max_{\Lambda\in\mathbf{\Lambda}} \mathbf{E}_A \left\|e(t)\right\|_2^2 \qquad (20)$$

$$J_B(K_B(t)^\circ) = \min_{K\in\mathcal{K}} \max_{\alpha\in\mathcal{A}} \mathbf{E}_B \left\|e(t)\right\|_2^2 \qquad (21)$$

Theorem 4.1: Assuming $P_A(0) = P_B(0) = P_0$, we have that estimations problems (20) and (21) are equivalent for all $t > 0$:

$$J_A(K_A(t)^\circ) = J_B(K_B(t)^\circ) \quad \text{and} \quad K_A(t)^\circ = K_B(t)^\circ$$

Proof: The result is proven by induction. Fix $K_A(1) = K_B(1) = K(1)$. To show that the assertion holds for the base case ($t = 1$), it is sufficient to show:

$$\min_{\Lambda\in\mathbf{\Lambda}} \text{Tr } P_A(1) = \min_{\alpha\in\mathcal{A}} \text{Tr } P_B(1) \qquad (22)$$

To further simplify, it is easy to show that equality (22) holds if

$$\min_{\Lambda\in\mathbf{\Lambda}} \text{Tr } G(1)\Lambda G(1)^* = \min_{\alpha\in\mathcal{A}} \text{Tr } G(1)\left(\sum_{i=1}^{m}\alpha_i\Lambda_i\right)G(1)^*$$

Begining with the left hand side, we have that

$$\min_{\Lambda\in\mathbf{\Lambda}} \text{Tr } G(1)\Lambda G(1)^* = \min_{i\in\mathcal{I}} \text{Tr } \left(G(1)\Lambda_i G(1)^*\right)$$

Expanding the right hand side, we have

$$\min_{\alpha\in\mathcal{A}} \text{Tr } G(1)\left(\sum_{i=1}^{m}\alpha_i\Lambda_i\right)G(1)^* = \min_{\alpha\in\mathcal{A}} \sum_{i=1}^{m}\alpha_i \text{Tr } \left(G(1)\Lambda_i G(1)^*\right) = \min_{i\in\mathcal{I}} \text{Tr } \left(G(1)\Lambda_i G(1)^*\right)$$

This completes the base case. Now assume that the assertion holds for $t = 1, \cdots, T-1$. Fix $K_A(T) = K_B(T) = K(T)$. To show that the assertion holds for $t = T$, it is once again sufficient to show that

$$\min_{\Lambda\in\mathbf{\Lambda}} \text{Tr } P_A(T) = \min_{\alpha\in\mathcal{A}} \text{Tr } P_B(T) \qquad (23)$$

which can be further simplified to

$$\min_{\Lambda\in\mathbf{\Lambda}} \text{Tr } G(t)\Lambda G(t)^* + \sum_{\tau=1}^{t-1}\mathbf{F}(\tau+1:t)G(\tau)\Lambda G(\tau)^*\mathbf{F}(\tau+1:t)^* =$$

$$\min_{\alpha\in\mathcal{A}} \text{Tr } G(t)\left(\sum_{i=1}^{m}\alpha_i\Lambda_i\right)G(t)^* + \sum_{\tau=1}^{t-1}\mathbf{F}(\tau+1:t)G(\tau)\left(\sum_{i=1}^{m}\alpha_i\Lambda_i\right)G(\tau)^*\mathbf{F}(\tau+1:t)^*$$

Using arguments identical to those in the base case, it is easy to show that the above identity holds true. ■

Using the same line of reasoning in section III-B, it is routine to show that problems (20) and (21) can be reformulated as a SDP.

V. Conclusions

An interesting open question is to consider *nonlinear* estimators of X given observations of Y. In this case, we believe that uncertainty models $\mathcal{P}_A$ and $\mathcal{P}_B$ are *not* equivalent. We intend to explore this problem shortly.

Another important problem that requires exploration is that of identifying uncertainty models for noise signals. While we have shown that uncertainty models $\mathcal{P}_A$ and $\mathcal{P}_B$ are equivalent for linear estimation, it is unclear which of these models best describes the noise signals given sample paths of observations $y(t)$. Efficient representations of uncertainty for noise signals in the context of state estimation remains open.

References

[1] Y.L. Chen and Chen B.-S. IEEE Transactions on Signal Processing, 42(1):32- 45, 1994.

[2] I. R. Petersen and A. V. Savkin. Robust Kalman Filtering for Signals and Systems with Large Uncertainties. Birkhauser, Boston, 1999.

[3] Y.K. Foo and Y.C. Soh. Systems and Control Letters, 57(6):482-488, 2008.

[4] R.E. Kalman. Transactions of the ASME-Journal of Basic Engineering, 82(Series D):35-45, 1960.

[5] S. Nakamori, R. Caballero, A. Hermoso, J. Jimenez, and J. Linares. Signal Processing, 87(5):970-982, 2007.

[6] R.T. Rockafeller. *Convex Analysis*. Princeton University Press, 1970.

[7] U. Shaked, L. Xie, and Y.C. Soh. IEEE Transactions on Signal Processing, 49(11):2620-2629, 2001.

[8] S. Verdu and H.V. Poor. IEEE Transactions on Automatic Control, 29(6):499-511, 1984.

[9] F. Wang and V. Balakrishnan. IEEE Transactions on Signal Processing, 50(4):803-814, 2002.

[10] L. Xie, Y.C. Soh, and C.E. de Souza. IEEE Transactions on Automatic Control, AC-39(6):1310-1314, 1994.