

Improved State Estimation of Stochastic Systems

NICHOLAS A. NECHVAL, GUNDARS BERZINS, MARIS PURGAILIS

Department of Mathematical Statistics

University of Latvia

Raina Blvd 19, LV-1050, Riga

LATVIA

KONSTANTIN N. NECHVAL

Department of Applied Mathematics

Transport and Telecommunication Institute

Lomonosov Street 1, LV-1019, Riga

LATVIA

Abstract: - In the present paper, for constructing the minimum risk estimators of state of stochastic systems, a new technique of invariant embedding of sample statistics in a loss function is proposed. This technique represents a simple and computationally attractive statistical method based on the constructive use of the invariance principle in mathematical statistics. Unlike the Bayesian approach, an invariant embedding technique is independent of the choice of priors. It allows one to eliminate unknown parameters from the problem and to find the best invariant estimator, which has smaller risk than any of the well-known estimators. Also the problem of how to select the total number of the observations optimally when a constant cost is incurred for each observation taken is discussed. To illustrate the proposed technique, an example is given.

Key-Words: - Stochastic system; State; Estimation; Invariant embedding technique

1 Introduction

The state estimation of discrete-time systems in the presence of random disturbances and measurement noise is an important field in modern control theory. A significant research effort has been devoted to the problem of state estimation for stochastic systems. Since Kalman's noteworthy paper [1], the problem of state estimation in linear and nonlinear systems has been treated extensively and various aspects of the problem have been analyzed [2-8].

The problem of determining an optimal estimator of the state of stochastic system in the absence of complete information about the distributions of random disturbances and measurement noise is seen to be a standard problem of statistical estimation. Unfortunately, the classical theory of statistical estimation has little to offer in general type of situation of loss function. The bulk of the classical theory has been developed about the assumption of a quadratic, or at least symmetric and analytically simple loss structure. In some cases this assumption is made explicit, although in most it is implicit in the search for estimating procedures that have the "nice" statistical properties of unbiasedness and minimum variance. Such procedures are usually satisfactory if the estimators so generated are to be used solely for

the purpose of reporting information to another party for an unknown purpose, when the loss structure is not easily discernible, or when the number of observations is large enough to support Normal approximations and asymptotic results. Unfortunately, we seldom are fortunate enough to be in asymptotic situations. Small sample sizes are generally the rule when estimation of system states and the small sample properties of estimators do not appear to have been thoroughly investigated. Therefore, the above procedures of the state estimation have long been recognized as deficient, however, when the purpose of estimation is the making of a specific decision (or sequence of decisions) on the basis of a limited amount of information in a situation where the losses are clearly asymmetric – as they are here.

There exists a class of control systems where observations are not available at every time due to either physical impossibility and/or the costs involved in taking a measurement. For such systems it is realistic to derive the optimal policy of state estimation with some constraints imposed on the observation scheme.

It is assumed in this paper that there is a constant cost associated with each observation taken. The

optimal estimation policy is obtained for a discrete-time deterministic plant observed through noise. It is shown that there is an optimal number of observations to be taken.

The outline of the paper is as follows. A formulation of the problem is given in Section 2. Section 3 is devoted to characterization of estimators. A comparison of estimators is discussed in Section 4. A general analysis is presented in Section 5. An example is given in Section 6.

2 Problem Statement

To make the above introduction more precise, consider the discrete-time system, which in particular is described by vector difference equations of the following form:

$$\mathbf{x}(k+1) = \mathbf{A}(k+1, k)\mathbf{x}(k) + \mathbf{B}(k)\mathbf{u}(k), \quad (1)$$

$$\mathbf{z}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{w}(k), \quad k = 1, 2, 3, \dots, \quad (2)$$

where $\mathbf{x}(k+1)$ is an n vector representing the state of the system at the $(k+1)$ th time instant with initial condition $\mathbf{x}(1)$; $\mathbf{z}(k)$ is an m vector (the observed signal) which can be termed a measurement of the system at the k th instant; $\mathbf{H}(k)$ is an $m \times n$ matrix; $\mathbf{A}(k+1, k)$ is a transition matrix of dimension $n \times n$, and $\mathbf{B}(k)$ is an $n \times p$ matrix, $\mathbf{u}(k)$ is a p vector, the control vector of the system; $\mathbf{w}(k)$ is a random vector of dimension m (the measurement noise). By repeated use of (1) we find

$$\mathbf{x}(k) = \mathbf{A}(k, j)\mathbf{x}(j) + \sum_{i=j}^{k-1} \mathbf{A}(k, i+1)\mathbf{B}(i)\mathbf{u}(i), \quad j \leq k, \quad (3)$$

where the discrete-time system transition matrix satisfies the matrix difference equation,

$$\mathbf{A}(k+1, j) = \mathbf{A}(k+1, k)\mathbf{A}(k, j), \quad \forall k, j, \quad (4)$$

$$\mathbf{A}(k, k) = \mathbf{I}, \quad \mathbf{A}(k, j) = \prod_{i=j}^{k-1} \mathbf{A}(i+1, i). \quad (5)$$

From these properties, it immediately follows that

$$\mathbf{A}^{-1}(k, j) = \mathbf{A}(j, k), \quad \forall k, j, \quad (6)$$

$$\mathbf{A}(\alpha, \beta)\mathbf{A}(\beta, \gamma) = \mathbf{A}(\alpha, \gamma), \quad \forall \alpha, \beta, \gamma. \quad (7)$$

Thus, for $j \leq k$,

$$\mathbf{x}(j) = \mathbf{A}(j, k)\mathbf{x}(k) - \sum_{i=j}^{k-1} \mathbf{A}(j, i+1)\mathbf{B}(i)\mathbf{u}(i). \quad (8)$$

The problem to be considered is the estimation of the state of the above discrete-time system. This problem may be stated as follows. Given the observed sequence, $\mathbf{z}(1), \dots, \mathbf{z}(k)$, it is required to obtain an estimator \mathbf{d} of $\mathbf{x}(l)$ based on all available observed data $\mathbf{Z}^k = \{\mathbf{z}(1), \dots, \mathbf{z}(k)\}$ such that the expected losses (risk function)

$$R(\boldsymbol{\theta}, \mathbf{d}) = E_{\boldsymbol{\theta}} \{r(\boldsymbol{\theta}, \mathbf{d})\} \quad (9)$$

is minimized, where $r(\boldsymbol{\theta}, \mathbf{d})$ is a specified loss function at decision point $\mathbf{d} \equiv \mathbf{d}(\mathbf{Z}^k)$, $\boldsymbol{\theta} = (\mathbf{x}(l), \boldsymbol{\omega})$, $\boldsymbol{\omega}$ is an unknown parametric vector of the probability distribution of $\mathbf{w}(k)$, $k \leq l$.

If it is assumed that a constant cost $c > 0$ is associated with each observation taken, the criterion function for the case of k observations is taken to be

$$r_k(\boldsymbol{\theta}, \mathbf{d}) = r(\boldsymbol{\theta}, \mathbf{d}) + ck. \quad (10)$$

In this case, the optimization problem is to find

$$\min_k \min_{\mathbf{d}} E_{\boldsymbol{\theta}} \{r_k(\boldsymbol{\theta}, \mathbf{d})\}, \quad (11)$$

where the inner minimization operation is with respect to $\mathbf{d} \equiv \mathbf{d}(\mathbf{Z}^k)$, when the k observations have been taken, and where the outer minimization operation is with respect to k .

3 Characterization of Estimators

For any statistical decision problem, an estimator (a decision rule) \mathbf{d}_1 is said to be equivalent to an estimator (a decision rule) \mathbf{d}_2 if $R(\boldsymbol{\theta}, \mathbf{d}_1) = R(\boldsymbol{\theta}, \mathbf{d}_2)$ for all $\boldsymbol{\theta} \in \Theta$, where $R(\cdot)$ is a risk function, Θ is a parameter space. An estimator \mathbf{d}_1 is said to be uniformly better than an estimator \mathbf{d}_2 if $R(\boldsymbol{\theta}, \mathbf{d}_1) < R(\boldsymbol{\theta}, \mathbf{d}_2)$ for all $\boldsymbol{\theta} \in \Theta$. An estimator \mathbf{d}_1 is said to be as good as an estimator \mathbf{d}_2 if $R(\boldsymbol{\theta}, \mathbf{d}_1) \leq R(\boldsymbol{\theta}, \mathbf{d}_2)$ for all $\boldsymbol{\theta} \in \Theta$. However, it is also possible that we may have “ \mathbf{d}_1 and \mathbf{d}_2 are incomparable”, that is, $R(\boldsymbol{\theta}, \mathbf{d}_1) < R(\boldsymbol{\theta}, \mathbf{d}_2)$ for at least one $\boldsymbol{\theta} \in \Theta$, and $R(\boldsymbol{\theta}, \mathbf{d}_1) > R(\boldsymbol{\theta}, \mathbf{d}_2)$ for at least one $\boldsymbol{\theta} \in \Theta$. Therefore, this ordering gives a partial ordering of the set of estimators.

An estimator \mathbf{d} is said to be uniformly non-dominated if there is no estimator uniformly better than \mathbf{d} . The conditions that an estimator must satisfy in order that it might be uniformly non-dominated are given by the following theorem.

Theorem 1 (Uniformly non-dominated estimator).

Let $(\xi_\tau; \tau=1,2, \dots)$ be a sequence of the prior distributions on the parameter space Θ . Suppose that $(\mathbf{d}_\tau; \tau=1,2, \dots)$ and $(Q(\xi_\tau, \mathbf{d}_\tau); \tau=1,2, \dots)$ are the sequences of Bayes estimators and prior risks, respectively. If there exists an estimator \mathbf{d}^* such that its risk function $R(\theta, \mathbf{d}^*)$, $\theta \in \Theta$, satisfies the relationship

$$\lim_{\tau \rightarrow \infty} [Q(\xi_\tau, \mathbf{d}^*) - Q(\xi_\tau, \mathbf{d}_\tau)] = 0, \quad (12)$$

where

$$Q(\xi_\tau, \mathbf{d}) = \int_{\Theta} R(\theta, \mathbf{d}) \xi_\tau(d\theta), \quad (13)$$

then \mathbf{d}^* is an uniformly non-dominated estimator.

Proof. Suppose \mathbf{d}^* is uniformly dominated. Then there exists an estimator \mathbf{d}^{**} such that $R(\theta, \mathbf{d}^{**}) < R(\theta, \mathbf{d}^*)$ for all $\theta \in \Theta$. Let

$$\varepsilon = \inf_{\theta \in \Theta} [R(\theta, \mathbf{d}^*) - R(\theta, \mathbf{d}^{**})] > 0. \quad (14)$$

Then

$$Q(\xi_\tau, \mathbf{d}^*) - Q(\xi_\tau, \mathbf{d}^{**}) \geq \varepsilon. \quad (15)$$

Simultaneously,

$$Q(\xi_\tau, \mathbf{d}^{**}) - Q(\xi_\tau, \mathbf{d}_\tau) \geq 0, \quad (16)$$

$\tau=1,2, \dots$, and

$$\lim_{\tau \rightarrow \infty} [Q(\xi_\tau, \mathbf{d}^{**}) - Q(\xi_\tau, \mathbf{d}_\tau)] \geq 0. \quad (17)$$

On the other hand,

$$\begin{aligned} & Q(\xi_\tau, \mathbf{d}^{**}) - Q(\xi_\tau, \mathbf{d}_\tau) \\ &= [Q(\xi_\tau, \mathbf{d}^*) - Q(\xi_\tau, \mathbf{d}_\tau)] - [Q(\xi_\tau, \mathbf{d}^*) - Q(\xi_\tau, \mathbf{d}^{**})] \\ &\leq [Q(\xi_\tau, \mathbf{d}^*) - Q(\xi_\tau, \mathbf{d}_\tau)] - \varepsilon \end{aligned} \quad (18)$$

and

$$\lim_{\tau \rightarrow \infty} [Q(\xi_\tau, \mathbf{d}^{**}) - Q(\xi_\tau, \mathbf{d}_\tau)] < 0. \quad (19)$$

This contradiction proves that \mathbf{d}^* is an uniformly non-dominated estimator. \square

4 Comparison of Estimators

In order to judge which estimator might be preferred for a given situation, a comparison based on some ‘‘closeness to the true value’’ criteria should be made. The following approach is commonly used [9-10]. Consider two estimators, say, \mathbf{d}_1 and \mathbf{d}_2 having risk function $R(\theta, \mathbf{d}_1)$ and $R(\theta, \mathbf{d}_2)$, respectively. Then the relative efficiency of \mathbf{d}_1 relative to \mathbf{d}_2 is given by

$$\text{rel. eff.}_R \{ \mathbf{d}_1, \mathbf{d}_2; \theta \} = R(\theta, \mathbf{d}_2) / R(\theta, \mathbf{d}_1). \quad (20)$$

When $\text{rel. eff.}_R \{ \mathbf{d}_1, \mathbf{d}_2; \theta_0 \} < 1$ for some θ_0 , we say that \mathbf{d}_2 is more efficient than \mathbf{d}_1 at θ_0 . If $\text{rel. eff.}_R \{ \mathbf{d}_1, \mathbf{d}_2; \theta \} \leq 1$ for all θ with a strict inequality for some θ_0 , then \mathbf{d}_1 is inadmissible relative to \mathbf{d}_2 .

5 General Analysis

5.1 Inner Minimization

First consider the inner minimization, i.e., k is held fixed for the time being. Then the term ck does not affect the result of this minimization. Consider a situation of state estimation described by one of a family of density functions, indexed by the vector parameter $\theta = (\mu, \sigma)$, where $\mu = x(k)$ and $\sigma = \omega(>0)$ are respectively parameters of location and scale. For this family, invariant under the group of positive linear transformations: $z \rightarrow az + b$ with $a > 0$, we shall assume that there is obtainable from some informative experiment (a random sample of observations $\mathbf{z}^k = \{z(0), \dots, z(k)\}$) a sufficient statistic (m_k, s_k) for (μ, σ) with density function $p_k(m_k, s_k; \mu, \sigma)$ of the form

$$p_k(m_k, s_k; \mu, \sigma) = \sigma^{-2} f_k[(m_k - \mu) / \sigma, s_k / \sigma]. \quad (21)$$

We are thus assuming that for the family of density functions an induced invariance holds under the group G of transformations: $m_k \rightarrow am_k + b, s_k \rightarrow as_k$ ($a > 0$). The family of density functions satisfying the above conditions is, of course, the limited one of normal, negative exponential, Weibull and gamma (with known index) density functions.

The loss incurred by making decision d when $\mu = x(l)$ is the true parameter is given by the piecewise-linear loss function

$$r(\theta, d) = \begin{cases} \frac{c_1(d - \mu)}{\sigma} & (\mu \leq d), \\ \frac{c_2(\mu - d)}{\sigma} & (\mu > d). \end{cases} \quad (22)$$

The decision problem specified by the informative experiment density function (21) and the loss function (22) is invariant under the group G of transformations. Thus, the problem is to find the best invariant estimator of μ ,

$$d^* = \arg \min_{d \in \mathcal{D}} R(\theta, d), \tag{23}$$

where \mathcal{D} is a set of invariant estimators of μ , $R(\theta, d) = E_{\theta}\{r(\theta, d)\}$ is a risk function.

5.2 Best Invariant Estimator

It can be shown by using the invariant embedding technique [11-12] that an invariant loss function, $r(\theta, d)$, can be transformed as follows:

$$r(\theta, d) = \ddot{r}(\mathbf{v}, \eta), \tag{24}$$

where

$$\ddot{r}(\mathbf{v}, \eta) = \begin{cases} c_1(v_1 + \eta v_2) & (v_1 \geq -\eta v_2), \\ -c_2(v_1 + \eta v_2) & (v_1 < -\eta v_2), \end{cases} \tag{25}$$

$\mathbf{v}=(v_1, v_2)$, $v_1=(m_k - \mu) / \sigma$, $v_2=s_k / \sigma$, $\eta=(d-m_k)/s_k$.

It follows from (24) that the risk associated with d and θ can be expressed as

$$\begin{aligned} R(\theta, d) &= E_{\theta}\{r(\theta, d)\} = E_k\{\ddot{r}(\mathbf{v}, \eta)\} \\ &= c_1 \int_0^{\infty} dv_2 \int_{-\eta v_2}^{\infty} (v_1 + \eta v_2) f_k(v_1, v_2) dv_1 \\ &\quad - c_2 \int_0^{\infty} dv_2 \int_{-\infty}^{-\eta v_2} (v_1 + \eta v_2) f_k(v_1, v_2) dv_1, \end{aligned} \tag{26}$$

which is constant on orbits when an invariant estimator (decision rule) d is used, where $f_k(v_1, v_2)$ is defined by (21). The fact that the risk (26) is independent of θ means that a decision rule d , which minimizes (26), is uniformly best invariant. The following theorem gives the central result in this section.

Theorem 2 (Best invariant estimator of μ). Suppose that (v_1, v_2) is a random vector having density function

$$v_2 f_k(v_1, v_2) \left[\int_0^{\infty} v_2 dv_2 \int_{-\infty}^{\infty} f_k(v_1, v_2) dv_1 \right]^{-1} \quad (v_1 \text{ real}, v_2 > 0), \tag{27}$$

where f_k is defined by (21), and let G_k be the distribution function of v_1/v_2 . Then the uniformly best invariant linear-loss estimator of μ is given by

$$d^* = m_k + \eta^* s_k, \tag{28}$$

where

$$G_k(-\eta^*) = c_1 / (c_1 + c_2). \tag{29}$$

Proof. From (26)

$$\begin{aligned} &\frac{\partial E_k\{\ddot{r}(\mathbf{v}, \eta)\}}{\partial \eta} \\ &= c_1 \int_0^{\infty} v_2 dv_2 \int_{-\eta v_2}^{\infty} f_k(v_1, v_2) dv_1 - c_2 \int_0^{\infty} v_2 dv_2 \int_{-\infty}^{-\eta v_2} f_k(v_1, v_2) dv_1 \\ &= \int_0^{\infty} v_2 dv_2 \int_{-\infty}^{\infty} f_k(v_1, v_2) dv_1 [c_1 P_k\{(v_1, v_2) : v_1 + \eta v_2 > 0\} \\ &\quad - c_2 P_k\{(v_1, v_2) : v_1 + \eta v_2 < 0\}] \\ &= \int_0^{\infty} v_2 dv_2 \int_{-\infty}^{\infty} f_k(v_1, v_2) dv_1 [c_1(1 - G_k(-\eta)) - c_2 G_k(-\eta)]. \end{aligned} \tag{30}$$

Then the minimum of $E_k\{\ddot{r}(\mathbf{v}, \eta)\}$ occurs for η^* being determined by setting $\partial E_k\{\ddot{r}(\mathbf{v}, \eta)\} / \partial \eta = 0$ and this reduces to

$$c_1[1 - G_k(-\eta^*)] - c_2 G_k(-\eta^*) = 0, \tag{31}$$

which establishes (29). \square

Corollary 2.1 (Minimum risk of the best invariant estimator of μ). The minimum risk is given by

$$\begin{aligned} R(\theta, d^*) &= E_{\theta}\{r(\theta, d^*)\} = E_k\{\ddot{r}(\mathbf{v}, \eta^*)\} \\ &= c_1 \int_0^{\infty} dv_2 \int_{-\eta^* v_2}^{\infty} v_1 f_k(v_1, v_2) dv_1 - c_2 \int_0^{\infty} dv_2 \int_{-\infty}^{-\eta^* v_2} v_1 f_k(v_1, v_2) dv_1 \end{aligned} \tag{32}$$

with η^* as given by (29).

Proof. These results are immediate from (24) when use is made of $\partial E_k\{\ddot{r}(\mathbf{v}, \eta)\} / \partial \eta = 0$. \square

5.3 Outer Minimization

The results obtained above can be further extended to find the optimal number of observations. Now

$$\begin{aligned}
 E_{\theta} \{r_k(\theta, d^*)\} &= E_{\theta} \{r(\theta, d^*) + ck\} = E_k \{r(\mathbf{v}, \eta^*) + ck\} \\
 &= c_1 \int_0^{\infty} dv_2 \int_{-\eta^*v_2}^{\infty} v_1 f_k(v_1, v_2) dv_1 \\
 &\quad - c_2 \int_0^{\infty} dv_2 \int_{-\infty}^{-\eta^*v_2} v_1 f_k(v_1, v_2) dv_1 + ck \quad (33)
 \end{aligned}$$

is to be minimized with respect to k . It can be shown that this function (which is the constant risk corresponding to taking a sample of fixed sample size k and then estimating $x(l)$ by the expression (28) with k for k^*) has at most two minima (if there are two, they are for successive values of k ; moreover, there is only one minimum for all but a denumerable set of values of c). If there are two minima, at k^* and k^*+1 , one may randomize in any way between the decisions to take k^* or k^*+1 observations.

6 Example

Consider the one-dimensional discrete-time system, which is described by scalar difference equations of the form (1)-(2), and the case when the measurement noises $w(k)$, $k = 1, 2, \dots$ (see (2)) are independently and identically distributed random variables drawn from the exponential distribution with the density

$$f(w; \sigma) = (1/\sigma) \exp(-w/\sigma), \quad w \in (0, \infty), \quad (34)$$

where the parameter $\sigma > 0$ is unknown. It is required to find the best invariant estimator of $x(l)$ on the basis of the data sample $\mathbf{z}^k = (z(1), \dots, z(k))$ relative to the piecewise linear loss function

$$r(\theta, d) = \begin{cases} c_1 (d - \mu)/\sigma, & d \geq \mu, \\ c_2 (\mu - d)/\sigma, & \text{otherwise,} \end{cases} \quad (35)$$

where $\theta = (\mu, \sigma)$, $\mu = x(l)$, $c_1 > 0$, $c_2 = 1$.

The likelihood function of \mathbf{z}^k is

$$L(\mathbf{z}^k; \mu, \sigma) = \frac{1}{\sigma^k} \exp \left[- \sum_{j=1}^k (z(j) - H(j)x(j))/\sigma \right]$$

$$= \frac{1}{\sigma^k} \exp \left[- \sum_{j=1}^k a(j)(y(j) - \mu)/\sigma \right], \quad (36)$$

where

$$y(j) = [a(j)]^{-1} \left(z(j) + H(j) \sum_{i=j}^{l-1} A(j, i+1) B(i) u(i) \right), \quad j \leq l, \quad (37)$$

$$y(j) = [a(j)]^{-1} \left(z(j) - H(j) \sum_{i=1}^{j-1} A(j, i+1) B(i) u(i) \right), \quad j > l, \quad (38)$$

if $l < k$ (estimation of the past state of the system), and

$$y(j) = \frac{z(j) + b(j)}{a(j)}, \quad (39)$$

$$a(j) = H(j) A(j, k_1), \quad (40)$$

$$b(j) = H(j) \sum_{i=j}^{l-1} A(j, i+1) B(i) u(i),$$

if either $l = k$ (estimation of the current state of the system) or $l > k$ (prediction of the future state of the system).

It can be justified by using the factorization theorem that (m_k, s_k) is a sufficient statistic for $\theta = (\mu, \sigma)$, where

$$m_k = \min_{1 \leq j \leq k} y(j), \quad s_k = \sum_{j=1}^k a(j) [y(j) - m_k]. \quad (41)$$

The probability density function of (m_k, s_k) is given by

$$\begin{aligned}
 &p_k(m_k, s_k; \mu, \sigma) \\
 &= \frac{n(k)}{\sigma} e^{-\frac{n(k)[m_k - \mu]}{\sigma}} \frac{1}{\Gamma(k-1)\sigma^{k-1}} s_k^{k-2} e^{-\frac{s_k}{\sigma}}, \\
 &m_k > \mu, \quad s_k > 0, \quad (42)
 \end{aligned}$$

where

$$n(k) = \sum_{j=1}^k a(j). \quad (43)$$

Since the loss function (35) is invariant under the group G of location and scale changes, it follows (see (25)) that

$$r(\boldsymbol{\theta}, d) = \ddot{r}(\mathbf{v}, \eta) = \begin{cases} c_1(v_1 + \eta v_2)/\sigma, & v_1 \geq -\eta v_2, \\ -(v_1 + \eta v_2)/\sigma, & \text{otherwise,} \end{cases} \quad (44)$$

where $\mathbf{v}=(v_1, v_2)$,

$$v_1 = \frac{m_k - \mu}{\sigma}, \quad v_2 = \frac{s_k}{\sigma}, \quad \eta = \frac{d - m_k}{s_k}. \quad (45)$$

Thus, using (28) and (29), we find that the best invariant estimator (BIE) of μ is given by

$$d_{\text{BIE}} = m_k + \eta^* s_k, \quad (46)$$

where

$$\eta^* = [1 - (c_1 + 1)^{1/k}] / n(k) = \arg \inf_{\eta} E_k \{ \ddot{r}(\mathbf{v}, \eta) \}, \quad (47)$$

$$E_k \{ \ddot{r}(\mathbf{v}, \eta) \} = [(c_1 + 1)(1 - \eta n(k))^{-(k-1)} - 1] / n(k) - \eta(k-1). \quad (48)$$

The risk of this estimator is

$$\begin{aligned} R(\boldsymbol{\theta}, d_{\text{BIE}}) &= E_{\boldsymbol{\theta}} \{ r(\boldsymbol{\theta}, d_{\text{BIE}}) \} \\ &= E_k \{ \ddot{r}(\mathbf{v}, \eta^*) \} = k[(c_1 + 1)^{1/k} - 1] / n(k). \end{aligned} \quad (49)$$

Here the following theorem holds.

Theorem 3 (Characterization of the estimator d_{BIE}). For the loss function (35), the best invariant estimator of μ , d_{BIE} , given by (46) is uniformly non-dominated.

Proof. The proof follows immediately from Theorem 1 if we use the prior distribution on the parameter space Θ ,

$$\begin{aligned} \xi_{\tau}(d\boldsymbol{\theta}) &= \frac{1}{\tau\sigma} e^{-\frac{\tau-\mu}{\tau\sigma}} \frac{1}{\Gamma(1/\tau)\sigma^{1/\tau+1}} \left(\frac{1}{\tau}\right)^{1/\tau} e^{-\frac{1}{\tau\sigma}} d\mu d\sigma, \\ \mu &\in (-\infty, \tau), \quad \sigma \in (0, \infty). \end{aligned} \quad (50)$$

This ends the proof. \square

Consider, for comparison, the following estimators of μ (state of the system):

The maximum likelihood estimator (MLE):

$$d_{\text{MLE}} = m_k; \quad (51)$$

The minimum variance unbiased estimator (MVUE):

$$d_{\text{MVUE}} = m_k - \frac{s_k}{(k-1)n(k)}; \quad (52)$$

The minimum mean square error estimator (MMSEE):

$$d_{\text{MMSEE}} = m_k - \frac{s_k}{kn(k)}; \quad (53)$$

The median unbiased estimator (MUE):

$$d_{\text{MUE}} = m_k - (2^{1/(k-1)} - 1) \frac{s_k}{n(k)}. \quad (54)$$

Each of the above estimators is readily seen to be of a member of the class

$$\mathcal{C} = \{d : d = m_k + \eta s_k\}, \quad (55)$$

where η is a real number. A risk of an estimator, which belongs to the class \mathcal{C} , is given by (48). If, say, $k=3$ and $c_1=26$, then we have that

$$\text{rel. eff.}_R \{d_{\text{MLE}}, d_{\text{BIE}}; \boldsymbol{\theta}\} = 0.231, \quad (56)$$

$$\text{rel. eff.}_R \{d_{\text{MVUE}}, d_{\text{BIE}}; \boldsymbol{\theta}\} = 0.5, \quad (57)$$

$$\text{rel. eff.}_R \{d_{\text{MMSEE}}, d_{\text{BIE}}; \boldsymbol{\theta}\} = 0.404, \quad (58)$$

$$\text{rel. eff.}_R \{d_{\text{MUE}}, d_{\text{BIE}}; \boldsymbol{\theta}\} = 0.45. \quad (59)$$

In this case (33) becomes

$$\begin{aligned} E_{\boldsymbol{\theta}} \{ r_k(\boldsymbol{\theta}, d^*) \} &= E_{\boldsymbol{\theta}} \{ r(\boldsymbol{\theta}, d_{\text{BIE}}) + ck \} = E_k \{ \ddot{r}(\mathbf{v}, \eta^*) + ck \} \\ &= k[(c_1 + 1)^{1/k} - 1] / n(k) + ck \equiv J_k. \end{aligned} \quad (60)$$

Now (60) is to be minimized with respect to k . It is easy to see that

$$\begin{aligned} J_k - J_{k-1} &= -((k-1)[(c_1 + 1)^{1/(k-1)} - 1] / n(k-1) \\ &\quad - k[(c_1 + 1)^{1/k} - 1] / n(k)) + c. \end{aligned} \quad (61)$$

Define

$$\begin{aligned} \varphi(k) &= (k-1)[(c_1+1)^{1/(k-1)} - 1]/n(k-1) \\ &\quad - k[(c_1+1)^{1/k} - 1]/n(k). \end{aligned} \quad (62)$$

Thus

$$c \underset{\leq}{\geq} \varphi(k) \Leftrightarrow J_k \underset{\geq}{\leq} J_{k-1}. \quad (63)$$

By plotting $\varphi(k)$ versus k the optimal number of observations k^* can be determined.

For each value of c , we can find an equilibrium point of k , i.e., $c = \varphi(k^*)$. The following two cases must be considered:

1) k^* is not an integer. We have $k^{(1)} < k^* < k^{(1)} + 1 = k^{(2)}$, where $k^{(1)}$ and $k^{(2)}$ are neighboring integers. Since $\varphi(k)$ is monotonically decreasing, we know that $\varphi(k^{(1)}) > c$ and $\varphi(k^{(2)}) < c$. Then, by using these properties, (61) becomes

$$J_{k^{(1)}} - J_{k^{(1)}-1} = -\varphi(k^{(1)}) + c < 0, \quad (64)$$

$$J_{k^{(2)}} - J_{k^{(1)}} = -\varphi(k^{(2)}) + c > 0, \quad (65)$$

Thus

$$J_{k^{(2)}} > J_{k^{(1)}} < J_{k^{(1)}-1}. \quad (66)$$

Therefore, $k^{(1)}$ is the optimal number of observations. We conclude that the optimal number k^* is equal to the largest integer below the equilibrium point.

2) k^* is an integer. By the same sort of argument, we know that k^* is as good as k^*-1 . Consequently, both k^* and k^*-1 are the optimal number of observations. Notice that in this case, J_{k^*} can be computed directly and precisely from (60).

7 Conclusions

In this paper we construct the minimum risk estimators of state of stochastic systems. The method used is that of the invariant embedding of sample statistics in a loss function in order to form pivotal quantities, which make it possible to eliminate unknown parameters from the problem. This method is a special case of more general considerations applicable whenever the statistical problem is invariant under a group of transformations, which acts transitively on the parameter space.

For a class of state estimation problems where

observations on system state vectors are constrained, i.e., when it is not feasible to make observations at every moment, the question of how many observations to take must be answered. This paper models such a class of problems by assigning a fixed cost to each observation taken. The total number of observations is determined as a function of the observation cost.

Extension to the case where the observation cost is an explicit function of the number of observations taken is straightforward. A different way to model the observation constraints should be investigated.

References

- [1] R. Kalman, A new approach to linear filtering and prediction problems, *Trans. ASME, J. Basic Eng.*, Vol. 82, 1960, pp. 34-45,
- [2] T. P. McGarty, *Stochastic Systems and State Estimation*, New York: John Wiley and Sons, Inc., 1974.
- [3] A. Savkin and I. Petersen, Robust state estimation and model validation for discrete-time uncertain system with a deterministic description of noise and uncertainty. *Automatica*, Vol. 34, 1998, pp. 271-274.
- [4] M. Norgaard, N. K. Poulsen and O. Ravn, New developments in state estimation for nonlinear systems, *Automatica*, Vol. 36, 2000, pp. 1627-1638.
- [5] J. Yan and R. R. Bitmead, Incorporating state estimation into model predictive control and its application to network traffic control, *Automatica*, Vol. 41, 2005, pp. 595 - 604.
- [6] T. Alamo, J. Bravo and E. Camacho, Guaranteed state estimation by zonotopes, *Automatica*, Vol. 41, 2005, pp. 1035-1043.
- [7] S. Gillijns and B. De Moor, Unbiased minimum-variance input and state estimation for linear discrete-time systems, *Automatica*, Vol. 43, 2007, pp. 111 - 116.
- [8] S. Ko and R. R. Bitmead, State estimation for linear systems with state equality constraints, *Automatica*, Vol. 43, 2007, pp. 1363 - 1368.
- [9] N.A. Nechval, *Modern Statistical Methods of Operations Research*, Riga: RCAEI, 1982.
- [10] N.A. Nechval, *Theory and Methods of Adaptive Control of Stochastic Processes*, Riga: RCAEI, 1984.
- [11] N.A. Nechval, K.N. Nechval and E.K. Vasermanis, Optimization of interval estimators via invariant embedding technique, *IJCAS (An International Journal of Computing Anticipatory Systems)*, Vol. 9, 2001, pp. 241-255.
- [12] E.K. Vasermanis and N.A. Nechval, *Improved Decisions in Statistics*, Riga: SIA, 2004.