International Symposium on Signal Processing and its Applications, ISSPA, Gold Coast, Australia, 25-30 August, 1996.
Organized by the Signal Processing Research Centre, QUT, Brisbane, Australia.

ON FINITE-DIMENSIONAL RISK-SENSITIVE ESTIMATION

Subhrakanti Dey and John B. Moore

Department of Systems Engineering
Research School of Information Science and Engineering
Australian National University
Canberra, ACT 0200 Australia
e-mail: subhra@syseng.anu.edu.au

ABSTRACT

In this paper, we address the finite-dimensionality issues regarding discrete-time risk-sensitive estimation for stochastic nonlinear systems. We show that for a bilinear system with an unknown parameter, finite-dimensional risk-sensitive estimates can be obtained. A necessary condition is obtained for nonlinear systems with no process noise such that one can obtain finite-dimensional risk-sensitive estimates.

1. INTRODUCTION

It is well known that in optimal stochastic linear filtering theory, the Kalman filter achieves the conditional mean estimate or the minimum variance estimate for Gauss-Markov systems, and is finite-dimensional for finite-dimensional models. Minimum variance estimation for nonlinear systems can also be achieved, but via infinite-dimensional filters, in general. These filters optimize a quadratic cost criterion which is the estimation error energy, and are optimal under the assumption that the plant and noise parameters are fully known. They are insensitive to risks involved due to uncertainties in the plant or the noise process, and are termed as ***risk-neutral filters***. Risk-sensitive filters optimize the expectation of the exponential of the estimation error energy, thus penalizing all its higher order moments. This obviously makes the risk-sensitive filter very cautious and robust to system uncertainties, the degree of cautiousness being determined by a risk-sensitive parameter which weights the index of the exponential.

A contemporary question is: when risk-neutral filters are finite-dimensional, are the corresponding risk-sensitive versions also finite-dimensional (probably under modified conditions)? It appears that at this stage of our knowledge, this question can only be answered on a case by case basis. Here we consider two such important cases beyond the well understood Kalman filter and hidden Markov model filters.

Risk-sensitive filtering for discrete-time linear Gauss-Markov systems addressed in [2] results in a finite-dimensional linear filter which is an H_∞ filter. Risk-sensitive control problems have been addressed in [3] [4] [5]. A solution to the output feedback risk-sensitive control problem for nonlinear discrete-time systems using information state feedback techniques resulting in infinite-dimensional controllers has been given in [6]. These information states have been found to be not a conditional probability density functions of the states given the observations, but that of an augmented plant which includes part of the risk-sensitive cost in the state process [9]. Risk-sensitive filters for stochastic nonlinear systems have also been derived in [7] and for hidden Markov models with finite-discrete state space in [8]. These filters are obtained via similar information state techniques where the information states are given by a linear but infinite-dimensional recursion (finite-dimensional in case of linear signal models and hidden Markov models). The optimizing state estimate is then obtained as the minimizing argument of a particular integral involving the information state. There are interesting interpretations of the results obtained from the risk-sensitive filters when the risk-sensitive parameter approaches certain limits. It is shown that when it approaches zero, the known risk-neutral filters are derived. On the other hand, in the small noise limit, the risk-sensitive filters have an interpretation [9] in terms of a deterministic worst-case noise estimation problem given from a differential game.

As is known from optimal stochastic nonlinear filtering theory, the conditional density filter is infinite-dimensional in general. Special cases of finite-dimensional filters for continuous-time systems, other than linear or discrete-state systems, have been found in [11] [15] [14]. A necessary and sufficient condition for the existence of finite-dimensional optimal filters for a class of discrete-time nonlinear systems has been found in [12]. Finite-dimensional risk-sensitive information states for continuous-time partially observed risk-sensitive control problems have been obtained in [13]. Also, the risk-sensitive cost has been generalised to absorb nonlinearities to yield finite-dimensional optimal controllers in [10] and finite-dimensional risk-sensitive filters/smoothers in [16] for discrete-time nonlinear systems.

In this paper, we derive risk-sensitive estimation results for discrete-time linear stochastic systems with time-varying unknown parameters using reference probability methods. These methods use a discrete-time version of Girsanov's Theorem, Fubini's theorem and Kolmogorov's Extension theorem [1] and have been used to derive risk-

This work was supported by the Cooperative Research Centre for Robust and Adaptive Systems under the Cooperative Research Centres Program of the Australian Commonwealth Government

International Symposium on Signal Processing and its Applications, ISSPA, Gold Coast, Australia, 25-30 August, 1996. Organized by the Signal Processing Research Centre, QUT, Brisbane, Australia.

sensitive filtering and control results in [6] [7] [8] [16] [9]. The presence of unknown parameters essentially makes it a nonlinear problem, thus resulting in an infinite-dimensional risk-sensitive filter/estimator in general. Rather than using the very general nonlinear risk-sensitive filtering results of [7] by augmenting the parameters to the state, this problem is treated as a special case of such general systems and simpler results are obtained. The special case of a bilinear structure where the unknown parameters enter the system matrices linearly is also considered to obtain closed-form finite-dimensional results.

This paper also explores the finite-dimensionality issues for a given class of risk-sensitive cost indices for the same class of nonlinear discrete-time systems as in [12]. The results of [12] are extended to obtain a necessary condition for the existence of finite-dimensional risk-sensitive filters. It is found that this condition is more complicated than the corresponding result for risk-neutral conditional expectation filters in [12] and not easily verifiable. It remains an open question therefore as to whether or not it is possible to find easily verifiable conditions for the existence of finite-dimensional risk-sensitive filters for a class of nonlinear systems for which finite-dimensional risk-neutral filters exist under certain conditions.

2. RISK-SENSITIVE ESTIMATION

In this section, we introduce the discrete-time stochastic nonlinear state space signal model with unknown parameters and define the problem of risk-sensitive estimation. We derive expressions for unnormalized information states and the optimizing risk-sensitive estimate. Finally, we see how one can obtain finite-dimensional risk-sensitive estimation results when the unknown parameters enter the dynamics in a linear manner.

Consider the following state space model defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$:

$$\begin{aligned} x_{k+1} &= A(\theta_{k+1}, x_k) + w_{k+1} \\ y_k &= C(\theta_k, x_k) + v_k \end{aligned} \tag{1}$$

where $x_k \in \mathbb{R}^n, y_k \in \mathbb{R}^p$. $w_k \in \mathbb{R}^n, \{w_k\}, k \in \mathbb{N}$ is a sequence of i.i.d random variables with a density function ψ. $v_k \in \mathbb{R}^p, \{v_k\}, k \in \mathbb{N}$ is a sequence of *i.i.d* random variables with a strictly positive density function ϕ. $A(\theta_{k+1}, x_k)$ and $C(\theta_k, x_k)$ are, in general, nonlinear vector valued functions. Note that in (1), we can have the noise processes w_k, v_k scaled by matrices depending on the unknown parameter(s) θ_k. However, we do not consider that situation for the sake of simplicity. Here, x_0 or its distribution is supposed to be known.

For simplicity, let us take $\theta_k, k \in \mathbb{N}$ to be scalar valued and satisfying the following dynamics:

$$\theta_{k+1} = f(\theta_k) + \nu_{k+1} \tag{2}$$

where $f(.)$ is a real valued nonlinear function in general and $\{\nu_k\}, k \in \mathbb{N}$ is a sequence of *i.i.d* random variables with density ρ. Here, θ_0 or its distribution is supposed to be known.

Definition 2.1 *Let* $\mathcal{G}^0_{k+1} = \sigma\{\theta_0, \ldots, \theta_{k+1}, x_0, \ldots, x_{k+1}, y_0, \ldots, y_k\}, \mathcal{Y}^0_k = \sigma\{y_0, \ldots, y_k\}$ *and the corresponding complete filtrations be* $\{\mathcal{G}_{k+1}\}$ *and* $\{\mathcal{Y}_k\}$ *respectively.*

The objective of risk-sensitive estimation is to determine $\hat{z}_{k|k} = (\hat{x}_{k|k}, \hat{\theta}_{k|k})$ where

$$\begin{aligned} \hat{z}_{k|k} = \ & \underset{\xi,\zeta}{\operatorname{argmin}}\, E[\exp(\theta\{\sum_{l=0}^{k-1} L(x_l, \theta_l, \hat{x}_{l|l}, \hat{\theta}_{l|l}) \\ & + L(x_k, \theta_k, \xi, \zeta)\}) \mid \mathcal{Y}_k] \end{aligned} \tag{3}$$

Here, $L(.,.,.,.)$ is assumed to be a convex function quadratically upper bounded. Define

$$D_k = \exp(\theta\{\sum_{l=0}^{k} L(x_l, \theta_l, \hat{x}_{l|l}, \hat{\theta}_{l|l}\})$$

and $\Lambda_k = \prod_{l=0}^{k} \frac{\phi(y_k - C(\theta_k, x_k))}{\phi(y_k)}$. Following the techniques developed in [1], we define a new measure $\bar{P}$, under which $\{y_k\}$ is a sequence of *i.i.d* random variables with a density ϕ. Using a discrete-time version of Girsanov's theorem, and equating the Radon-Nikodym derivative $\frac{dP}{d\bar{P}} \mid_{\mathcal{G}_k} = \Lambda_k$, one can show that the original measure P can be derived from $\bar{P}$. Using a version of Bayes Theorem, it is evident that our optimization problem is equivalent to the following problem (where $\bar{E}$ is the expectation under the new measure):

$$\hat{z}_{k|k} = \underset{\xi,\zeta}{\operatorname{argmin}}\, \bar{E}[\Lambda_k D_{k-1} \exp\{\theta L(x_k, \theta_k, \xi, \zeta)\} \mid \mathcal{Y}_k] \tag{4}$$

Definition 2.2 *Define the unnormalized information state* $q_k(x, \theta), k \in \mathbb{N}$ *such that*

$$q_k(x,\theta) dx d\theta = \bar{E}[\Lambda_{k-1} D_{k-1} I(x_k \in dx) I(\theta_k \in d\theta) \mid \mathcal{Y}_{k-1}] \tag{5}$$

Now, we state the following Lemma and Theorem without proof.

Lemma 2.1 *The unnormalized information state* $q_k(x, \theta)$ *satisfies the following recursion*

$$\begin{aligned} q_{k+1}(x,\theta) = & \int_{\mathbb{R}^n} \int_{\mathbb{R}} \frac{\phi(y_k - C(z,\lambda))}{\phi(y_k)} \\ & \times \exp(\theta L(z, \lambda, \hat{x}_{k|k}, \hat{\theta}_{k|k})) \psi(x - A(z,\theta)) \\ & \times \rho(\theta - f(\lambda)) q_k(z,\lambda) d\lambda dz \end{aligned} \tag{6}$$

Note 2.1 Note that $q_0(x,\theta) = \pi_0(x)\rho_0(\theta)$ where $\pi_0(.), \rho_0(.)$ are the densities of x_0, θ_0 respectively.

Theorem 2.1 *The optimizing risk-sensitive estimate* $\hat{z}_{k|k}$ *is given by*

$$\begin{aligned} \hat{z}_{k|k} = \ & \underset{\xi,\zeta}{\operatorname{argmin}} \int_{\mathbb{R}^n} \int_{\mathbb{R}} \frac{\phi(y_k - C(z,\lambda))}{\phi(y_k)} \\ & \times \exp(\theta L(z, \lambda, \xi, \zeta)) q_k(z,\lambda) d\lambda dz \end{aligned} \tag{7}$$

2.1. Finite-dimensionality issues

In this section, we will briefly address some finite-dimensionality issues associated with such risk-sensitive estimation problems. Consider (1) with ϕ, ψ, ρ being Gaussian densities and the initial densities for x_0, θ_0, namely π_0, ρ_0 respectively to be Gaussian as well. Also, consider $L(x, \theta, \hat{x}_{k|k}, \hat{\theta}_{k|k})$ to be quadratic in nature, i.e.,

$$\begin{aligned} L(x, \theta, \hat{x}_{k|k}, \hat{\theta}_{k|k}) &= (x - \hat{x}_{k|k} \;\; \theta - \theta_{k|k})' \\ &\qquad Q(x - \hat{x}_{k|k} \;\; \theta - \theta_{k|k}) \end{aligned} \tag{8}$$

where $Q > 0$. Now, consider (1) again. We will concentrate on the bilinear case, where $A(\theta_{k+1}, x_k) = A\theta_{k+1}x_k$, $C(\theta_k, x_k) = C\theta_k x_k$. It is easy to see that from Lemma 2.1 that the recursion involving $q_k(x, \theta)$ is going to involve quadratic terms in the index of the exponential. Using completion of square techniques and integrating over the Gaussian densities twice (similar techniques as used in [7]), one can get a closed form expression for $q_k(x, \theta)$ which would also be a bivariate Gaussian density and hence finite-dimensional. The details of this algebraic manipulation is too long for the space available here and hence, are not provided. However, similar results can be found in [7]. Now, having obtained finite-dimensional information states, the optimizing risk-sensitive estimates can be obtained from (7) by using similar techniques. Since these estimates will be expressed as functions of the parameters of the information state, they will be finite-dimensional.

3. RISK-SENSITIVE ESTIMATION: THE CASE WHEN $W_K = 0$

In this section, we explore the possibilities of obtaining finite-dimensional risk-sensitive filters when the process noise, w_k is absent. In [12], a necessary and sufficient condition for obtaining finite-dimensional discrete-time nonlinear filters was obtained in the context of risk-neutral or minimum variance estimation for systems free of process noise. Here, we try to draw parallels of such results in the risk-sensitive context.

Consider (1) and to make it simple, let us ignore the unknown parameter θ_k, so that $A(\theta_{k+1}, x_k) = A(x_k)$, $C(\theta_{k+1}, x_k) = C(x_k)$, where $A(.), C(.)$ are still nonlinear vector valued functions. For more generality (as in [12]), let us a state-dependent noise variance for v_k, such that our discrete-time nonlinear system is now given by

$$\begin{aligned} x_{k+1} &= A(x_k) \\ y_k &= C(x_k) + \eta(x_k)v_k \end{aligned} \tag{9}$$

Here, assume that $v_k \sim N(0,1)$. Also, for obvious reasons, the index of the exponential in the risk-sensitive cost kernel is given by $L(x_k, \hat{x}_{k|k})$, where $L(.,.)$ is quadratically upper bounded.

Now, let us make the following assumptions.

1. A is a C^1 diffeomorphism
2. $\eta(x)$ is invertible for every x.

Also, define

$$\begin{aligned} h_{i,j}(x) &= (\eta(x)\eta(x)')^{-1}_{i,j}, i, j = 1, \ldots, p \\ h_i(x) &= \sum_{j=1}^{p} h_{i,j}(x)C_j(x), i = 1, \ldots, p \\ h_0(x) &= \sum_{j=1}^{p} h_j(x)C_j(x) \end{aligned} \tag{10}$$

It is well known that in the context of risk-neutral or minimum variance estimation, the conditional density filter plays an important role. In [12], it is shown that the associated unnormalized conditional measure can be expressed in terms of finite number of parameters if and only if the following functional space $H = \mathbb{R} - Span\{h_{i,j} \circ A^k, h_j \circ A^k \mid i, j = 1, \ldots, p, k \geq 0\}$ is finite-dimensional. For proof, see [12].

We now state the following lemma and theorem without proof.

Lemma 3.1 *The unnormalized conditional measure $q_k(x)$ defined by $q_k(x)dx = \bar{E}[\Lambda_{k-1}D_{k-1}I(x_k \in dx) \mid \mathcal{Y}_{k-1}]$ satisfies the following recursion*

$$\begin{aligned} q_{k+1}(x) &= \frac{\phi(y_k - C(A^{-1}(x))}{\phi(y_k)} \exp(\theta L(A^{-1}(x), \hat{x}_{k|k}) \\ &\qquad \times J_A(A^{-1}(x))^{-1} q_{k-1}(A^{-1}(x)) \end{aligned} \tag{11}$$

where $J_f(x)$ is the Jacobian of f at x.

Theorem 3.1 *The optimizing risk-sensitive estimate $\hat{x}_{k|k}$ is given by*

$$\begin{aligned} \hat{x}_{k|k} &= \underset{\xi}{\text{argmin}} \int_{\mathbb{R}^n} \frac{\phi(y_k - C(z))}{\phi(y_k)} \\ &\qquad \times \exp(\theta L(z, \xi)) q_k(z) dz \end{aligned} \tag{12}$$

Note that the similarity between (11) and the corresponding recursion for the unnormalized conditional measure in [12] indicates that similar proof techniques can be used to derive finite-dimensionality conditions for the $q_k(x)$ as defined in Lemma 3.1. However, things turn out to be a bit more complicated because of the presence of the quantity $\exp(\theta L(A^{-1}(x), \hat{x}_{k|k})$ because $\hat{x}_{k|k}$ is a function of y_k (as can be seen from (12)). Nevertheless, a necessary condition can be derived using the following assumptions. Assume that $q_k(x)$ can be expressed in terms of a finite number of parameters $\beta_k = (\beta_k(1), \beta_k(2), \ldots, \beta_k(r))$. Now, we assume the following,

$$\begin{aligned} \frac{\partial L(x, \hat{x}_{j|j})}{\partial y_j^i} &= \sum_{l=1}^{n} \frac{\partial L}{\partial \hat{x}_{j|j}^l} \frac{\partial \hat{x}_{j|j}^l}{\partial y_j^i} \\ &= \sum_{m=1}^{p} M_{m,i}(x, \beta_j) y_j^m + N_i(x, \beta_j) \end{aligned} \tag{13}$$

Note here, that $j = 0, \ldots, k$ and y_j^l denotes the l-th element of the vector y_j, $l = 1, \ldots, p$.

Remark 3.1 An example where this assumption holds is the case of linear systems, with $L(x, y) = (x - y)'(x - y)$

where $\hat{x}_{k|k}$ can be expressed as a linear function of y_k. Details can be found in [7]. But for a given nonlinearity, note that it is not very easy to verify whether this assumption holds or not.

Definition 3.1 *Define the functional space*

$$H_1 = \mathbb{R} - Span\{(h_{i,j} + M_{i,j}) \circ A^{k+1}, (h_j + N_j) \circ A^{k+1} \mid i, j = 1, \ldots, p, k \geq 0\} \quad (14)$$

Theorem 3.2 *A necessary condition for the finite-dimensionality of $q_k(x)$ as defined in Lemma 3.1, for the discrete-time nonlinear system (9) with the risk-sensitive cost defined as in (3), is H_1 is finite-dimensional.*

Proof Proof techniques are similar to those of [12]. However, it is fairly long and cannot be given here. □

4. CONCLUSIONS

This paper deals with discrete-time finite-dimensional risk-sensitive estimation. Risk-sensitive filters, for nonlinear systems, are, in general, infinite-dimensional. However, we show that in the case of a discrete-time bilinear system with an unknown parameter where the parameter enters the dynamics in a linear fashion, one can obtain finite-dimensional risk-sensitive estimates. Also, in the case of a system where there is no process noise, a necessary condition for obtaining finite-dimensional risk-sensitive estimates is given. However, this condition is not easily verifiable as its risk-neutral counterpart. Extensions of such results for risk-sensitive filters can also be made for discrete-time nonlinear systems with more generalized measurement noise and also for nonlinear systems with delay.

References

[1] R. J. Elliott, L. Aggoun and J. B. Moore, *Hidden Markov Models: Estimation and Control*, Springer-Verlag, Application of Mathematics Series, 1994.

[2] J. L. Speyer, C. Fan and R. N. Banavar, "Optimal Stochastic Estimation with Exponential Cost Criteria," *Proceedings of the 31st Conference on Decision and Control*, Vol. 2, pp. 2293-2298, Dec. 1992.

[3] P.Whittle, "Risk-Sensitive Linear/Quadratic/Gaussian Control," *Adv. Applied Probability 13*, pp. 746-777, 1981, Applied Probability Trust.

[4] A. Bensoussan and J. H. van Schuppen, "Optimal Control of Partially Observable Stochastic Systems with an Exponential-of-Integral Performance Index," *SIAM J. Control Opt.*, vol. 23, pp. 599-613, 1985.

[5] D. H. Jacobson, "Optimal Stochastic Linear Systems with Exponential Performance Criteria and Their Relation to Deterministic Games," *IEEE Trans. on Automatic Control*, vol. AC-18, No. 2, April 1973.

[6] M. R. James, J. S. Baras and R. J. Elliott, "Risk-sensitive Control and Dynamic Games for Partially Observed Discrete-time Systems," *IEEE Trans. on Automatic Control*, 1994, to appear.

[7] S. Dey and J. B. Moore, "Risk-sensitive filtering and smoothing via Reference Probability Methods", *Proc. of the American Control Conference*, Vol. 1, pp. 129-133, Seattle, June 1995.

[8] S. Dey and J. B. Moore, "Risk-sensitive filtering and smoothing for Hidden Markov Models," *Systems and Control Letters*, Vol. 25, No. 5, pp. 361-366, August 1995.

[9] J. B. Moore, R.J. Elliott and S. Dey, "Risk-sensitive Generalizations of Minimum Variance Estimation and Control," *Journal of Mathematical Systems, Estimation and Control*, accepted for publication, 1995.

[10] J. B. Moore and J. Baras, "Finite-dimensional Optimal Controllers for Nonlinear Plants," *Systems and Control Letters*, accepted for publication, 1995.

[11] V. E. Benes, "Exact finite dimensional filters for certain diffusions with nonlinear drifts," *Stochastics*, Vol.5, pp. 65-92, 1981.

[12] J. Levine and G. Pignie, "Exact finite-dimensional filters for a class of nonlinear discrete-time systems," *Stochastics*, Vol. 18, pp. 97-132, 1986.

[13] V. E. Benes and R. J. Elliott, "Finite-dimensional risk-sensitive information states," *Proc. of the IFAC Symposium on Nonlinear Control System Design (NOLCOS)*, Vol. 2, pp. 471-476, June 1995.

[14] O. Zeitouni, "An extension of the Benes filter and some identification problems solved by nonlinear filtering methods," *System and Control Letters*, No. 5, pp. 9-17, 1984.

[15] W. S. Wong, "New classes of finite-dimensional nonlinear filters," *System and Control Letters*, No. 3, pp. 155-164, 1983.

[16] S. Dey and J. B. Moore, "Finite-dimensional risk-sensitive filters and smoothers for discrete-time nonlinear systems," submitted.