Central Difference Formulation of Risk-Sensitive Filter

Smita Sadhu, M. Srinivasan, S. Bhaumik, and T. K. Ghoshal

***Abstract*—A numerically efficient algorithm for risk-sensitive filters (known to be robust to model uncertainties) of nonlinear plants, using central difference approximation is proposed. The proposed filter, termed central difference risk-sensitive filter (CDRSF), overcomes several disadvantages associated with the extended risk-sensitive filter (ERSF), reported earlier. The theory of formulation and the algorithm of the CDRSF are presented. With an example, it is demonstrated that the proposed new filter would give much better tracking performance compared to the ERSF for certain nonlinear systems. The CDRSF would be nearly as fast as the ERSF, thus making it more preferable for real-time applications compared to the risk-sensitive particle filter.**

***Index Terms*—Central difference filters, Gaussian, non-linear, risk-sensitive filters, robustness.**

I. Introduction

THIS PAPER addresses risk-sensitive filters (RSF) which minimize the expected value of an exponential function of the squared estimation error and are more robust [1], [2] compared to the conventional minimum mean-square error (MMSE) (risk-neutral) filters. For linear Gaussian signal models, the RSF may be formulated as Kalman filter-like recursions [1] and is closely related to the H_∞ filter. In RSF, a designer-chosen parameter, called the risk-sensitive parameter, provides a tool for design tradeoff between the filtering performance for the nominal model and the robustness to model uncertainty. With a nonlinear model and/or non-Gaussian excitation, though it is possible to establish a recursion for the RSF [1], [3], [4], a closed-form solution is generally ruled out as it involves intractable integrals. Deterministic or stochastic methods of numerical integration [5], [6] require substantial computer time and may not be suitable for real-time and onboard applications.

In this letter, we propose a new risk-sensitive numerically efficient derivative-free filter based on the central difference [9] quadrature called the central difference risk-sensitive filter (CDRSF).

The only other technique for risk-sensitive filtering of nonlinear signal model and having comparable numerical efficiency is the extended risk-sensitive filter (ERSF) [7]. ERSF employs the small signal linearization technique of the extended Kalman filter (EKF) [13] for the risk-sensitive case and captures the essence of risk-sensitive filtering for plants with mild nonlinearities. However, all shortcomings of the EKF [13] are inherited by the ERSF.

The CDRSF, in contrast, is based on large signal linearization and is inspired by a class of filters called sigma-point filters (SPF) [8] for risk-neutral nonlinear filtering problems. The SPF family accommodates large disturbances or sample-to-sample excursions in state-space, where the EKF may exhibit poor performance. All the techniques in the SPF family including the central difference filter (CDF) [9] circumvent the problem of intractable integration with approximate quadrature using a finite number of deterministic support points called sigma-points [8]. The extent (relative distance in state-space) of the sigma-points is selected based on the square root of the error covariance matrix and therefore takes care of the amount of perturbation.

The definition and expressions for general nonlinear risk-sensitive filtering problem are discussed next. This is followed by the formulation of the proposed risk-sensitive filter and the corresponding recursive equations.

II. Risk-Sensitive Filter

Consider the nonlinear plant described by the state and measurement equations as follows:

$$x_{k+1} = \phi(x_k) + w_k \tag{1}$$

$$y_k = \gamma(x_k) + v_k \tag{2}$$

where $x_k \in R^n$ denotes the state of the system, $y_k \in R^p$ is the measurement at the instance k where $k = \{0, 1, 2, 3, \ldots, N\}$, and $\phi(x_k)$ and $\gamma(x_k)$ are known nonlinear functions of x_k and k. The process noise $w_k \in R^n$ and measurement noise $v_k \in R^p$ are assumed to be uncorrelated and have known covariances.

The following notations have been used for the probability density functions:

$$f(x_{k+1} \mid x_k) \triangleq p_{X_{k+1}|X_k}(\cdot \mid x_k)$$
$$g(y_k \mid x_k) \triangleq p_{Y_k|X_k}(\cdot \mid x_k).$$

The objective is to estimate a known function $\Phi(x)$ of the state variables. The estimate is designated as $\hat{\Phi}(x)$ and its optimal value in the risk-sensitive sense is denoted as $\hat{\Phi}^*(x)$, which minimizes the cost function $C\left(\hat{\Phi}_1, \ldots, \hat{\Phi}_k\right) = E\left[\exp\left(\mu_1 \sum_{i=1}^{k-1} \rho_1\left(\Phi(x_i) - \hat{\Phi}_i^*\right)\right) + \left(\mu_2 \rho_2\left(\Phi(x_k) - \hat{\Phi}_k^*\right)\right)\right]$, where $\mu_1 \geq 0$ and $\mu_2 > 0$ are two risk parameters. Functions $\rho_1(\cdot)$ and $\rho_2(\cdot)$ are both strictly convex, continuous and bounded from below, attaining global minima at 0. Conventionally, they represent squared deviation for the mean-squared measure.

Manuscript received May 22, 2006; revised September 12, 2006. The work of M. Srinivasan was supported by the Aeronautical Research and Development Board (AR&DB), New Delhi, India. The work of S. Bhaumik was supported by the Council of Scientific and Industrial Research (CSIR), New Delhi, India. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Brian Sadler.

The authors are with the Department of Electrical Engineering, Jadavpur University, Kolkata-700 032, India (e-mail: smita@debesh.wb.nic.in).

Digital Object Identifier 10.1109/LSP.2006.887838

In particular, the minimum risk-sensitive estimate (MRSE) is defined by

$$\hat{\Phi}_k^* = \arg\min C\left(\hat{\Phi}_1^*, \ldots, \hat{\Phi}_{k-1}^*, \hat{\Phi}_k\right). \tag{3}$$

It can be shown that the MRSE satisfies the following recursion:

$$\begin{aligned}\sigma_{k+1}(x_{k+1}) = \int & f(x_{k+1} \mid x_k) g(y_k \mid x_k) \\ & \times \exp\left(\mu_1 \rho_1\left(\Phi(x_k) - \hat{\Phi}_k^*\right)\right) \\ & \times \sigma_k(x_k)\, dx_k \end{aligned} \tag{4}$$

$$\begin{aligned}\hat{\Phi}_k^* = \arg\min_{\alpha \in R} \int & \exp\left(\mu_2 \rho_2\left(\Phi(x_k) - \alpha\right)\right) \\ & \times \sigma_k(x_k)\, dx_k \end{aligned} \tag{5}$$

where $\sigma_k(x_k)$ represents an information state [1] and may be normalized and α is a parameter.

Here, the variables to be estimated are the state variables themselves ($\Phi(x) = x$), and both the convex functions $\rho_1(\cdot)$ and $\rho_2(\cdot)$ are known quadratic functions of vectors i.e., $\rho_j(\alpha) = \alpha^T\alpha$ for $j = 1, 2$. We denote the optimal estimate as $\hat{x}_{k+1|k}$ and accordingly we change the notation for the information state.

With the above assumptions

$$\begin{aligned}\sigma_{k+1|k}(x_{k+1}) = \int & f(x_{k+1} \mid x_k) g(y_k \mid x_k) \\ & \times \exp\left(\mu_1 \left(x_k - \hat{x}_{k|k-1}\right)^T \right. \\ & \left. \times \left(x_k - \hat{x}_{k|k-1}\right)\right) \\ & \times \sigma_{k|k-1}(x_k)\, dx_k \end{aligned} \tag{6}$$

$$\begin{aligned}\hat{x}_{k|k-1} = \arg\min_{\alpha \in R} \int & \exp\left(\mu_2 (x_k - \alpha)^T (x_k - \alpha)\right) \\ & \times \sigma_{k|k-1}(x_k)\ dx_k. \end{aligned} \tag{7}$$

III. Formulation of Central Difference Risk-Sensitive Filter (CDRSF)

In this section, the concept of (risk-neutral) CDF has been extended to the risk-sensitive case and appropriate recursive closed-form formulae have been derived using the same model as in the previous section. The central difference approach [9] involves use of Gaussian approximation of distributions and deterministically chosen support points to carry out the integration retaining the moments of appropriate order matched. Similar approaches have also been used in [10] and [11]. For the CDF, differentiability is not mandatory but the concerned functions $\phi(\cdot)$ and $\gamma(\cdot)$ should satisfy Lipschitz continuity condition [9] for convergence. The same restrictions have been assumed for the CDRSF, which would also require (shown later) that the risk-sensitive parameter should be sufficiently small to ensure convergence.

The CDRSF has been formulated using the normalized information states.

The integrands (6) and (7) are rewritten as follows:

$$\sigma_{k+1|k}(x_{k+1}) = \int f(x_{k+1} \mid x_k) \sigma_{k|k}(x_k)\, dx_k \tag{8}$$

where $\sigma_{k|k}(x_k)$ is the a posteriori estimate obtainable from the previous (kth) step using the following relation:

$$\begin{aligned}&\sigma_{k+1|k+1}(x_{k+1}) \\ &\quad = g(y_{k+1} \mid x_{k+1}) \\ &\qquad \times \exp\left(\mu_1 \left(x_{k+1} - \hat{x}_{k+1|k}\right)^T \left(x_{k+1} - \hat{x}_{k+1|k}\right)\right) \\ &\qquad \times \sigma_{k+1|k}(x_{k+1}). \end{aligned} \tag{9}$$

The optimal estimation is given by the expression

$$\begin{aligned}\hat{x}_{k+1|k} = \arg\min_{\alpha \in R} \int & \exp\left(\mu_2 (x_{k+1} - \alpha)^T (x_{k+1} - \alpha)\right) \\ & \times \sigma_{k+1|k}(x_{k+1})\, dx_{k+1}. \end{aligned} \tag{10}$$

As is done in CDF [9], Gaussian approximations have been used in the proposed CDRSF to replace $\sigma_k(\cdot)$, which makes the integrations (8) and (9) tractable. Deterministic sigma-points have been used to evaluate relation (8) as a Gaussian quadrature, and another set of sigma-points have been used to evaluate the convolution (9). Further, a Gaussian approximation would allow representation of the concerned distributions simply by the respective mean and covariance. To facilitate such approximation, $\sigma_k(\cdot)$ must be normalized to qualify as a probability density function. When Gaussian approximation is used for $\sigma_{k+1|k}(\cdot)$, the optimal estimate is simply the mean value of the distribution [12].

In the development that follows, we use expressions analogous to those used in [9]. The mean and covariance of $\sigma_{k+1|k}(x_{k+1})$ are denoted by $\bar{\sigma}_{k+1|k}$ and $P_{k+1|k}$ respectively. The notations $\bar{\sigma}_{k|k}$ and $P_{k|k}$ denote the mean and covariance of $\sigma_{k|k}(x_k)$, respectively.

Filter Algorithm

(i) ***Filter initialization:*** Initialize the filter with appropriate values of $x_{0|0}$ and $P_{0|0}$

(ii) ***Predictor step:***

• Compute the factorization $P_{k|k} = S^T S$ using Cholesky decomposition as in [9].

• Choose a suitable scalar h and compute the central difference approximations of the first and second derivatives as $a_i = \left(\phi\left(x_{k|k} + hS^T e_i\right) - \phi\left(x_{k|k} - hS^T e_i\right)\right)/2h$, and $H_{i,i} = \left(\phi\left(x_{k|k} + hS^T e_i\right) - 2\phi\left(x_{k|k}\right) + \phi\left(x_{k|k} - hS^T e_i\right)\right)/h^2$ respectively, where e_i is the unit vector in the ith direction and $1 \le i \le n$. The choice of h signifies the span of the sigma-points and may be chosen to suit the nature of the distribution to be represented, e.g., for Gaussian distributions, the recommended value of $h = \sqrt{3}$ [9].

• Compute the mean and covariance as

$$\bar{\sigma}_{k+1|k} = \phi\left(x_{k|k}\right) + \sum_{i=1}^{n} \frac{1}{2} H_{i,i} \text{ and}$$

$$P_{k+1|k} = Q + \sum_{i=1}^{n} a_i a_i^T + \sum_{i=1}^{n} \frac{1}{2} H_{i,i} H_{i,i}^T$$

where Q is the process noise covariance.

(iii) ***Optimization Step:*** With Gaussian approximation, the optimal estimate is $\hat{x}_{k+1|k} = \bar{\sigma}_{k+1|k}$

(iv) ***Corrector Step:***The last two terms in (9) may be combined into an equivalent Gaussian distribution if we take the Gaussian approximation $\tilde{\sigma}_{k+1|k}(x_{k+1})$ in place of $\sigma_{k+1|k}(x_{k+1})$. The combined term may be expressed as $N(\text{mean}, \text{cov})$, where $\text{mean} = \hat{x}_{k+1|k}$ and $\text{cov} = P^{+}_{k+1|k} = \left(P^{-1}_{k+1|k} - 2\mu_2 I\right)^{-1}$. Hence, the corrector step would proceed as follows:

• Compute $P^{+}_{k+1|k} = \left(P^{-1}_{k+1|k} - 2\mu_2 I\right)^{-1}$

• Compute the factorization $P^{+}_{k+1|k} = \tilde{S}^T\tilde{S}$ using Cholesky decomposition as in [9].

• Compute the central difference approximation as $b_i = \left(\gamma\left(x_{k+1|k} + h\tilde{S}^T e_i\right) - \gamma\left(x_{k+1|k} - h\tilde{S}^T e_i\right)\right)/2h$, and $G_{i,i} = (\gamma\left(x_{k+1|k} + h\tilde{S}^T e_i\right) - 2\gamma\left(x_{k+1|k}\right) + \gamma\left(x_{k+1|k} - h\tilde{S}^T e_i\right))/h^2$, where $1 \le i \le n$.

• Compute $\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + L^{+}_{k+1}(y_{k+1} - z_{k+1})$ and, $P_{k+1|k+1} = P^{+}_{k+1|k} - L^{+}_{k+1}P^{T}_{xz}$, where $P_{xz} = \tilde{S}^T(b_1, \ldots, b_n)^T$, $P_{zz} = \sum_{i=1}^{n} b_i b_i^T + \sum_{i=1}^{n}(1/2)G_{i,i}G_{i,i}^T$, $L^{+}_{k+1} = P_{xz}(R + P_{zz})^{-1}$, and $z_{k+1} = \gamma\left(x_{k+1|k}\right) + \sum_{i=1}^{n}(1/2)G_{i,i}$.

(v) ***Recursion:*** Estimates for the subsequent steps may be computed by repeating the predictor, optimization, and corrector steps as above for the required number of time steps.

Note 1: CDRSF and CDF have the same computational complexity, as is evident from the algorithms.

Note 2: Alternative forms of CDRSF may be derived by any combinations of: i) using the risk-sensitive term in the prediction step; ii) using posterior estimation instead of prior; and iii) calculating the cross variances directly from the function values at the sigma-points without using the coefficients b_i and $G_{i,i}$.

Note 3: The condition $2\mu_2 P_{k+1|k} < I$ should be satisfied for $P^{+}_{k+1|k}$ to be computable in the corrector step. Hence, the risk-sensitive factor μ_2 should be assumed to be low enough to ensure convergence of the CDRSF. It is to be noted, however, that the ERSF too suffers from a similar drawback.

Note 4: Simulation studies have been carried out with different nonlinear systems and representative results for a severely nonlinear system are shown in the next section. Preliminary results indicate that for the classes of nonlinearities studied, the CDRSF works satisfactorily for various combinations of risk factors, process noise, and measurement noise.

IV. Case Study

Problem statement: The CDRSF algorithm has been studied with the following severely nonlinear plant inspired by [9] having the same state equation as [9] but having a measurement equation with a weaker bimodal tendency than that given in [9]

$$\begin{aligned} x_{k+1} &= \phi(x_k) + w_k \\ &\text{where} \\ \phi(\zeta) &= 5\zeta\left(1 - \zeta^2\right),\ w_k \sim N(0, 0.05), \\ y_k &= \gamma(x_k) + v_k \\ &\text{where} \\ \gamma(\zeta) &= 0.01\zeta\left(1 - 0.5\zeta\right),\ v_k \sim N(0, 0.0001). \end{aligned}$$

The system has three equilibrium points, of which, the one at the origin is unstable and the other two at ± 1 are stable. In the absence of any bias, the system hovers around either of the two stable equilibrium points after the initial transients when the process excitation is relatively small. The filter and plant initial conditions and the initial error covariance are set respectively at 0.8, -0.2, and 2, as in [9].

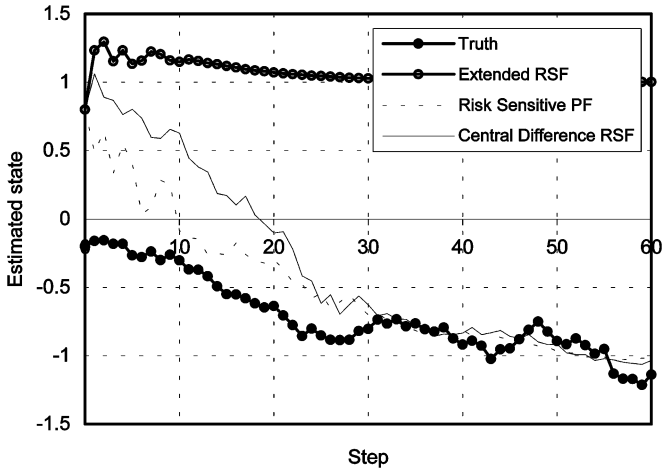

Fig. 1. Comparative performance of central difference RSF, extended RSF, and risk-sensitive particle filter.

1) Results:

- *Tracking performance* of ERSF, CDRSF, and RSPF for a typical run is shown in Fig. 1.
- *Rms error settling characteristic* over 10 000 Monte Carlo runs as given in Fig. 2 shows that the rms error is much lower compared to the ERSF.
- *Percentage fail-count* of the ERSF, indicated by the percentage of cases in which each filter fails to track the truth i.e., settles at the wrong equilibrium point, is about 26%, whereas that of the CDRSF is about 4.8%.
- *Computational cost* of the CDRSF, measured in terms of the time taken in MATLAB environment for a single run of each of the above filters, was found to be comparable to that of ERSF and the former is about 200 times faster compared to the RSPF.

V. Conclusions

A numerically efficient and derivative free central difference approximation of the risk-sensitive filter for nonlinear systems has been proposed. Being a risk-sensitive filter, the CDRSF has a built in mechanism to trade-off between nominal tracking

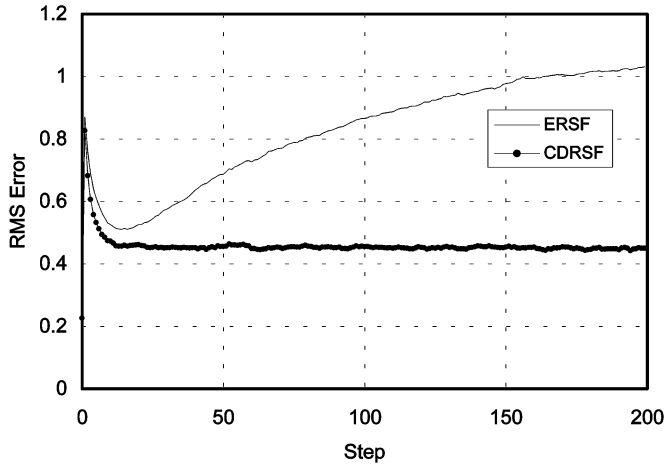

Fig. 2. RMS error of CDRSF and ERSF for 10 000 MC runs.

performance and robustness. Representative simulation results using a severely nonlinear filtering problem have been used to demonstrate the robustness of CDRSF whereby the CDRSF will have less track-loss compared to the ERSF. Using appropriate value of risk-sensitive parameter, the rms error performance of the CDRSF was shown to be better than that of the ERSF and close to that of the highly computing intensive and near optimum RSPF. Better error settling performance, lesser occurrence of track-loss, and good numerical efficiency makes it a candidate for onboard implementation.

REFERENCES

[1] R. K. Boel, M. R. James, and I. R. Peterson, "Robustness and risk-sensitive filtering," *IEEE Trans. Automat. Contr.*, vol. 47, no. 3, pp. 451–461, Mar. 2002.

[2] H. Zhang, L. Xie, and Y. C. Soh, "Risk-sensitive filtering, prediction and smoothing for discrete-time singular systems," *Automatica*, vol. 39, pp. 57–66, 2003.

[3] S. Dey and J. B. Moore, "Risk-sensitive filtering and smoothing via reference probability methods," *IEEE Trans. Automat. Contr.*, vol. 42, no. 11, pp. 1587–1591, Nov. 1997.

[4] ——, "Finite-dimensional risk-sensitive filters and smoothers for discrete-time nonlinear systems," *IEEE Trans. Automat. Contr.*, vol. 44, no. 6, pp. 1234–1239, Jun. 1999.

[5] S. Sadhu and A. Doucet, "Particle methods for risk-sensitive filtering," in *Proc. IEEE Indicon*, 2005, pp. 423–426.

[6] S. Bhaumik, M. Srinivasan, S. Sadhu, and T. K. Ghoshal, "Adaptive grid solution of risk sensitive estimator problems," in *Proc. IEEE Indicon*, 2005, pp. 356–360.

[7] M. Jayakumar and R. N. Banavar, "Risk-sensitive filters for recursive estimation of motion from images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 6, pp. 659–666, Jun. 1998.

[8] R. V. D. Merwe and E. Wan, "Sigma-point Kalman filters for probabilistic inference in dynamic state-space models," in *Proc. Workshop on Advances in Machine Learning*, Montreal, QC, Canada, Jun. 2003.

[9] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Trans. Automat. Cont.*, vol. 45, no. 5, pp. 910–927, May 2000.

[10] T. S. Schei, "A finite-difference method for linearization in nonlinear estimation algorithms," *Automatica*, vol. 33, pp. 2053–2058, 1997.

[11] M. Norgaard, N. K. Poulsen, and O. Ravn, "New developments in state estimation for nonlinear systems," *Automatica*, vol. 36, pp. 1627–1638, 2000.

[12] S. Bhaumik, S. Sadhu, and T. K. Ghoshal, "Risk-sensitive estimators for inaccurately modelled system," in *Proc. IEEE Indicon*, 2005, pp. 86–91.

[13] G. B. Robert and Y. C. H. Patrick, *Introduction to Random Signals and Applied Kalman Filtering with Matlab Exercises and Solutions*, 3rd Edition ed. New York: Wiley, 1997.