Risk Sensitive Estimators for Inaccurately Modelled Systems

Shovan Bhaumik[1], Smita Sadhu and Tapan Kumar Ghoshal

***Abstract* - Robustness of risk sensitive (RSE) estimators/filters for inaccurately modelled plant are elucidated and exemplified. A theorem which allows alternative pathway for deriving RSE filter relation and derivation of different closed form relations for RS filters in linear Gaussian cases is provided. Consequently, errors in expressions in earlier publications have been detected and rectified. Properties of RS filters are briefly reviewed and the interpretation of robustness of RS filters elaborated. Using Monte Carlo simulation, it is shown that RS filters perform significantly better compared to risk-neutral filters when (i) process noise covariance is in error (ii) the true system (truth model) contains unmodelled bias (iii) the state transition matrix is inaccurately known. Design pragmatics for the choice of the risk sensitive parameter is indicated.**

***Keywords* - Kalman filter, Model uncertainty. Risk sensitive filter, Robust Estimation**

I. INTRODUCTION

After nearly a decade, interest in Risk sensitive estimators has been resurrected in [1], which has used a more general framework compared to the earlier publications[3,5,7,8] which have their origin in Risk Sensitive Control. The concept of Risk sensitive estimators (RSE) is applicable to linear as well as nonlinear problems with an exponential cost criterion, with a mean square kernel or with a more general convex forms [1] kernel. The RSE is expected to have enhanced robustness and is closely linked with H_∞ type of robust estimators [1].

While the properties and motivations for Risk Sensitive Control law are fairly well known, despite a number of worthy publications up to 2002, the same cannot be said about Risk Sensitive Estimators. From the dearth of publication (with the notable exception of [6]) it may arguably be said that the properties of RSE are not appreciated to the extent of extracting design pragmatics for practical problems. Objective of the present paper is to attempt a mitigation of the above shortcoming.

This paper extends the work done in [4] to provide an intuitive appreciation of the properties of RSE's. The dimensions of robustness of RSE's have been discussed and extended. In particular, performance of RSE's in presence of unknown bias and inaccurate knowledge of process noise covariance has been elaborated. These aspects had not been reported earlier.

Generalised numerical methods for non linear and non Gaussian cases had not been available so far (exceptions being recent contributions [11,12,13]). Accordingly, examples using Linear Gaussian problems have been presented to illustrate the properties- so that the results can be claimed to be generally reproducible.

Available literature in RSE's tend to use different conventions and symbols which could be confusing to a beginner. To add to the problem, printers' devil abounds in some important publications [1, 9]. Using the conventions of [1], we state a more general, two parameter framework for formulating the RSE estimation problem. We also state a necessary theorem with which recursive solution of general (including non linear and non Gaussian signal models) RSE may be deduced. Using this theorem, the expressions provided in [1] for Linear Gaussian signal model have been corrected and several other closed forms for RSE's have been derived.

Department of Electrical Engineering, Jadavpur University, Kolkata – 700 032, India.

[1]Corresponding author: Tel & Fax: +913324146723, E-mail: shovan.bhaumik@gmail.com

II. FORMULATION OF RISK SENSITIVE ESTIMATION PROBLEM

The following general (nonlinear) signal model (additive noise) consisting of the state and measurement equation is considered:

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

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

Dimensionally, $w_k \in R^n, v_k \in R^p, x_k \in R^n$ and $y_k \in R^p$ of which x_k denotes the state of the system and y_k is the measurement at the instance k where $k = \{0,1,2,3\ldots n\}$. The vector $f(x_k)$ and $h(x_k)$ are general nonlinear function of x_k and k. We assume that $f(x_k)$ is nominally and $h(x_k)$ is absolutely known. The variables w_k, v_k are the process and measurement noise, respectively, which are random quantities with only sufficient statistics known. The initial state x_0 is another random variable, with known distribution and probability density. It is assumed that the process and measurement noise and the random variable signifying initial values of the variable are uncorrelated to each other.

A two parameter, risk sensitive squared error kernel type cost function at the time instance k may be defined as:

$$J_{RS}(\varsigma,k)=E[\exp(\mu_1\sum_{i=0}^{k-1}(x_i-\hat{x}_i)^T(x_i-\hat{x}_i)+\mu_2(x_k-\varsigma)^T(x_k-\varsigma))]$$

where $_i$'s are the optimum estimated values of state variable for past steps i∈{0,1,2,3…k-1}. The current optimum estimate $_k$ is obtained by finding the optimum value of ξ, which minimises $J^{RS}(\zeta,k)$. The constant parameters μ_1 and μ_2 are called risk sensitive parameters.

The risk sensitive cost function as above, includes an accumulated error cost up to time k, with a relative weight μ_1 and the current cost of squared estimation error is weighted by μ_2. The weighting factors, i.e. the risk sensitive parameters can be normalised by replacing μ_2 by unity and replacing μ_1 by the ratio $\theta=\mu_1/\mu_2$, (as done in almost all previous publications except [1]) without affecting the optimal estimate. The present workers however advocate the use of both the parameters as it allows changing the absolute as well as the ratio. Further, from numerical considerations also, a choice of 'small' values for risk sensitive parameters often avoids numerical overflow and inaccuracies.

III. Recursive Solution Of The RSE

The recursive solution of RSE is realized generally in a two-step process. First, a recursive relation of an information state [5] is formulated, which can be updated in each time step. In each time step, the optimal estimate is then obtained by extremizing the cost of another function involving the information state.

We propose [11,13], an inductive method for deriving the expression for the information state. The induction gives a direct and intuitive derivation, without the use of Measure change and Girsanov's theorem [2,3]. We omit the details of the induction process (proof) for space constarints and simply state the result as Theorem-1 below.

Theorem 1

The solution of the RSE problem may be obtained from the following recursive relations:

$$\hat{x}_{k+1|k+1}=\arg\min_{\varsigma}\int_{-\infty}^{+\infty}\exp[\mu_2(x_{k+1}-\varsigma)^T(x_{k+1}-\varsigma)]\alpha_{k+1}dx_k \tag{3}$$

Where, $\alpha_{k+1}=p(y_{k+1}\mid x_{k+1})$

$$\times\int_{-\infty}^{+\infty}\exp[\mu_1(x_{k+1}-\hat{x}_{k+1})^T(x_{k+1}-\hat{x}_{k+1})]\alpha_k p(x_{k+1}\mid x_k)dx_k \tag{4}$$

Notes (1) The recursive relation needs to be initialised for the information state. (2) The estimation equation is specifically applicable for posterior estimation where the estimation is performed after receiving the measurement. Similar relations can be obtained for prior estimates (also known as delayed measurement [4]), as used in [1]. (3) The relations, expressed in Theorem-1 are different from those obtained by [3,5]. Compared to [3], the expressions herein may be shown to be more numerically efficient in recursive update of information state and finding the optimal estimate. (4) The integrations, may not, in general be performed for non-linear or non Gaussian case.

IV. Linear Gaussian Signal Model

For the simplified case, closed form recursive, Kalman Filter like forms are possible. We call such filters as Risk sensitive Kalman Filter (RSKF) forms. The RSKF expressions can be obtained by using Theorem-1 as the Gaussian forms are integrable. The multiplicity of forms in earlier literature [1,3,9,5] arises because one may choose (a) the posterior or prior estimation expressions and derive them using (b) posterior or prior covariance intermediates to obtain the estimates. Thus there are, at least, four possible forms. Other subtle differences like using information filter (inverse of covariances) [5] are also possible.

The outline for deriving the relations for linear Gaussian case is as follows:

The linear Gaussian signal model is described as:

$$x_{k+1}=Fx_k+w_k,\quad w_k\sim N(0,Q)$$

$$y_k=Hx_k+v_k,\quad v_k\sim N(0,R)$$

From this,

$$p(y_{k+1}\mid x_{k+1})=N(y_{k+1}-Hx_{k+1},R),\ p(x_{k+1}\mid x_k)=N(x_{k+1}-Fx_k,Q)$$

All the items in equations (3) and (4) being Gaussian, and noting that the argmin operation for Gaussian expression is the mean, the equations can be evaluated in a straight forward, even if tedious process.

The RSKF expressions for posterior to posterior state update and prior to prior covariance update, as worked out and used in the present work are given by:

$$\hat{x}_{k+1|k+1}=F\hat{x}_k+(H^TR^{-1}H+\Sigma_{k+1|k}^{-1})^{-1}\times H^TR^{-1}(y_{k+1}-HF\hat{x}_{k|k}) \tag{5}$$

$$\Sigma_{k+1|k}=Q+F(\Sigma_{k|k-1}^{-1}+H^TR^{-1}H-2\mu_1 I)^{-1}F^T \tag{6}$$

It may be noted that for risk neutral case µ1=0 and the formulae for state estimation and error covariance update reduce to standard Kalman filter formulae. Further, the expressions do not contain μ_2

Corresponding expressions in [1] with prior to prior state update have been obtained only for unity variance process noise and measurement noise and that too have printing errors. It may be shown that for the more general case, the covariance update equation remains the same as above and the correct formulae for state update is :

$$\begin{aligned}\hat{x}_{k+1|k} &= F(H^T R^{-1} H + \Sigma_{k+1|k}^{-1} - 2\mu_1 I)^{-1} \\ &\times (H^T y_{k+1} + \Sigma_{k+1|k}^{-1} \hat{x}_{k|k-1} - 2\mu_1 \hat{x}_{k|k-1})\end{aligned} \tag{7}$$

V. Robustness Properties And Application Pragmatics Of RSE

Essential general properties of RSE have been well compiled in [4]. We concentrate only on two distinct but interconnected aspects of the properties, namely, robustness and the pragmatics for choosing risk sensitive parameters. We draw upon the already known properties, examine and reinterpret them as also attempt to extend them. It is near impossible to comment on properties for non-linear and non Gaussian signal models. However, in the tradition of the predecessors, we use the properties of Linear Gaussian problems to gain insight and conjecture that some of the properties would also be applicable to 'mildly' non-linear non-Gaussian cases. The extended KF like linearisation, used in [6] also indicate that some properties in linear cases may be extended to non-linear cases.

A. Robustness Properties

The connection between H_∞ filter and the RSE is well known [1]. In H_∞ filter the objective is to minimize the cost ($J_{H\infty}$) such that $J_{H\infty}<\gamma^2$. In case of Kalman filter $\gamma\to\infty$. Where as in case of RSE the risk sensitive parameter functions as μ_1 is propoortional to $1/\gamma^2$ as stated in [1], with say , $\mu_2=1$.

This is the primary reason why the RSE is considered to be robust. However, the different aspects of the robustness, namely, with respect to the presence of un-modelled bias, error/inaccuracy in state matrix and the same for process would be carefully examined. Also of importance is to find the temporal aspects that is how long does it take for the RSE to recognise the modelling error and to compensate for them. The properties are enumerated as given below:

(i) As the risk sensitive cost function memorises the past errors incurred by the estimator, intuitively it is expected that the RSE would learn from past mistakes and perform better. At the same time, learning requires time and the effect of learning can only be seen over time and not instantaneously. One may therefore expect that the performance of RSE to improve over time.

(ii) The above also means that RSE would be more effective against sustained perturbations rather than quickly varying disturbances in the model. Thus, in the presence of an un-modelled bias, the RSE is expected to track with lesser errors than a risk-neutral filter. That it is indeed so, is exemplified in the next section. As a corollary, one would not expect the RSE to effectively reject un-modelled, quickly varying, disturbances.

(iii)It should be pointed out that the RS cost function assigns cost for each point in state space and therefore penalizes the regions in state space, which, in the past , more often resulted in large error covariances.

(iv)The RS cost function computes the cost associated with points in state space based on its knowledge of the nominal model and also on the particular realization of the measurement sequence. As such, RS cost function cannot directly infer perturbations and changes in the signal model, it has to depend on measurement. For a signal model with lower measurement noise (covariance), the RSE is expected to perform better.

(v) Inspection of equation (5) (as also equation (7)) reveals that the effect of increasing tantamount to increasing the process noise covariance as well as the estimation error covariance. This enhances confidence in measurement and reduces the same in the assumed dynamics of the process- leading to higher feed back gain of the estimator- effectively increasing the filtering bandwidth.

(vi)In practical application of KF, it is customary to use a larger value of process noise covariance (Q) than the nominal to avoid the performance penalty if the true Q is larger than the nominal or to take care of modelling errors in plant dynamics. The above property shows why such artificial measures are unnecessary for RSKF as the risk sensitive parameter automatically does that. Further, increase in Q, generally makes the estimators faster, however, RS estimators, in general, takes a longer time to settle. The effect of increasing Q and increasing the risk factor, therefore does not have one to one correspondence.

B. Pragmatics Related to RS Parameters

From the discussion of previous paragraphs it may be easily inferred that enhancing the value of μ_1 increases robustness and bandwidth of the filter and therefore is desirable. However, there are limits beyond which one cannot increase the risk sensitive parameter.

The estimation in risk sensitive cases is possible over a horizon of length N $0\leq k\leq N$ where if and only if the following conditions are satisfied: $(P_{k|k}^{-1} - 2\mu_1 I) > 0 \;\; for \;\; 0 \leq k \leq N$. This is discussed in [1,4] and is also evident from inspecting equation (5) and equation (7)

In general there is no guarantee beforehand that these conditions will be satisfied for each step.

Apart from the above, larger values of risk sensitive parameters often gives rise to overflow or reduced accuracy. The present workers found that the following simple heuristic is useful to assign the first guess value for the risk sensitive parameter. This can be intuitively justified from the convergence criteria as above.

$\mu_1 \le 1/(2\,\|P\|)$ where $\|P\|$ is the largest eigenvalue of P, taken over all the values of k. The later can be estimated from a knowledge of the physics of the problem or from a run of KF or EKF.

VI. Numerical Experiment And Illustration

A. System Models For Experiment

"Truth model" [14] of the system is taken as

$$x_{k+1} = (F + \Delta F)x_k + bias + Bw_k$$

Where the filter model is $x_{k+1} = Fx_k + Bw_k$

The measurement model is $y_k = Hx_k + v_k$

Case: I: Single dimensional case

F=0.99,H=1 B=1 "Truth model" is initialised with the random number generated from the normal distribution zero mean and unity covariance.

Case II : Two-dimensional case:

In earlier publications [1], a rather contrived two dimensional model had been chosen. The system had nearly decoupled state variables but with nearly identical eigenvalues-which shows poor controllability and observability. We choose a more plausible system as discussed in [10].

$$F = \begin{bmatrix} 0 & -0.5 \\ 1 & 1 \end{bmatrix},\ B = \begin{bmatrix} -6 \\ 1 \end{bmatrix}\ H = \begin{bmatrix} -100 & 10 \end{bmatrix}\ w_k \sim N(0,1)$$

$$v_k \sim N(0,1)\ \Delta F = \begin{bmatrix} 0 & 0 \\ 0 & \delta \end{bmatrix}, |\delta| \le 0.3$$

"Truth model" is initialised with a random number generated from the zero mean and covariance $\begin{bmatrix} 1 & 0 \\ 0 & 25 \end{bmatrix}$ and filter is initialised with zero values for the both states.

B. Sensitivity to state model error

The fig 1 and 2 are the root mean square error (RMSE) of the state variables obtained from ordinary KF and risk sensitive KF (RSKF) for 100 Monte Carlo run of case II with δ=0.07 and μ_1=0.004. It should be noted from the figure that estimation by RSKF is better compared to KF. It should also be noted that RSKF estimation is much better than KF in the later time steps compared to the first few time steps. This is as expected according to our previous discussion. It is seen from the figures that even for such a small δ the RMSE calculated for the state variable from KF algorithm diverges but the RSKF does not.

The fig 3 illustrates the variation of the RMSE of first state variable at 100th step for both KF and RSKF as the uncertainty parameter δ ranges in value from −2 to 2 for a fixed value of μ_1=0.004 for case II. The graph shows that for very low value of δ, the Kalman filter achieves a lower terminal RMSE value than RSKF. For the larger value of δ the RSKF has the lower terminal RMSE value than KF. It is also seen from the graph that in this particular case the variation with δ is not symmetric in the positive and negative regions. The nature of variation of terminal RMSE with ä of state 2 shows similar trend and is not shown.

C. Sensitivity To Unmodeled Bias

If any unmodelled bias is present in the system, the integrator like effect is observed on the true-value of the state variable. The ordinary Kalman filter fails to track the true value of the states whereas RS filter acts comparatively better. It is illustrated in fig 4 using case I when ΔF is zero and bias is taken as 0.2. The standard deviation of the process and the measurement noise is 0.1 and 2.5 respectively. $\mu_1$is taken as 0.085.

D. Sensitivity To Noise Modelling Error

From the closed form equation of the risk sensitive filter for linear case it is clear that incorporating the risk sensitive parameter increases the effective prior covariance of the filter. So if the system is modelled wrongly by using a smaller process noise covariance, the risk sensitive filter estimates better than the risk neutral filter. For this experiment we have taken the case I where filter process noise covariance and measurement noise covariance are taken as 0.01 and unity respectively whereas the true process noise covariance is taken as 4 times that of filter. The RMSE obtained from the RSKF with μ_1=0.3, and KF is shown in the fig 5.

VII. Discussion And Conclusion

Using a general theorem for recursive solution the Recursive State estimation problem, RSKF relations have been obtained and errors in previously published results have been rectified. The known and often overlooked aspects of Robustness properties of risk sensitive filters are discussed and their implications brought out. Sensitivity of RSF to state model error, un-modelled bias and noise modelling errors are discussed with examples. The role of risk sensitive parameter and the considerations for its choice are discussed. The contribution is expected to encourage new application for this estimator.

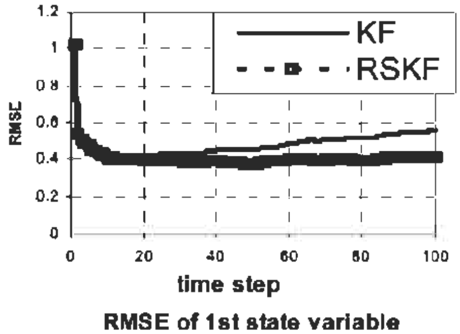

Fig.1.RMSE of first state variable for KF and RSKF of case II

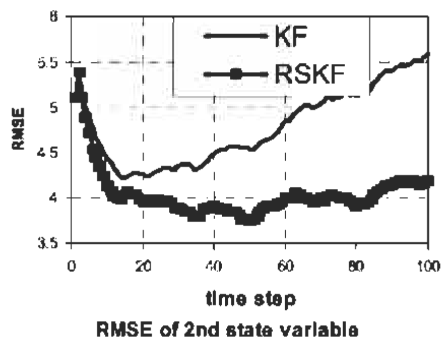

Fig.2.RMSE of 2nd state variable for KF and RSKF of case II

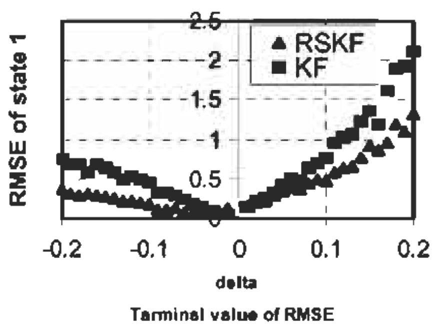

Fig.3. Terminal RMSE of 1st state variable for KF and RSKF for different δ of case II

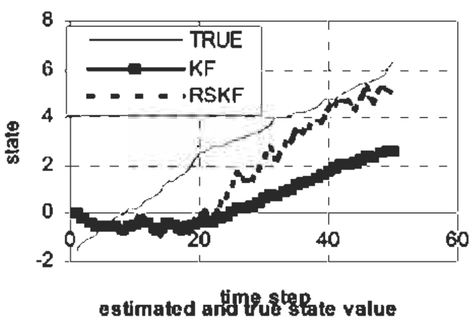

Fig 4. The true and estimated value of the state of case I with bias

VIII. ACKNOWLEDGEMENT

The first author is grateful to the Aeronautical Research and Development Board and Council of Scientific & Industrial Research (CSIR), India for supporting research.

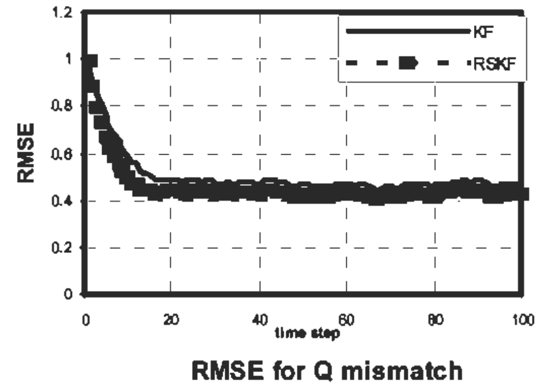

Fig.5. RMSE for RSKF and KF for Q mismatch of case I

REFERENCE

[1] R.K.Boel, M.R.James and I.R.Petersen "Robustness and risk sensitive filtering" IEEE Transaction on Automatic Control, vol 47, No 3, pp 451-460, March 2002

[2] Subhrakanti Dey and Charalambos D Charalambous " Discrete time risk sensitive filters with non gaussian initial conditions and their ergodic properties" Asian Journal of Control, vol. 3, no.4, pp. 262-271, December 2001

[3] Subrakanti Dey and John B Moore "Risk sensitive filtering and smoothing via reference probability methods" IEEE Transaction on Automatic Control, Vol 42, No 11, November 1997.

[4] R.N.Banavar and J.L.Speyer " Properties of risk sensitive filters/estimators" IEE Proceedings on Control, vol 145, no 1, January 1998

[5] I.B.Collings, Matthew R James and J B Moore "An information state approach to linear /risk-sensitive/Quadratic/Gaussian Control" 33rd IEEE Conference on Decision and Control, Vol 4, pp 3802-3807, December 1994

[6] M. Jayakumar and R. N. Banavar "Risk-sensitive Filters for Recursive Estimation of Motion from Images", Vol. 20, No. 6, pp. 659 - 666 , the IEEE Trans. on Pattern Analysis and Machine Intelligence, June 1998.

[7] J.B. Moorey R.J. Elliotty S. Dey "Risk-sensitive Generalizations of Minimum Variance Estimation and Control", Journal of Mathematical Systems, Estimation and Control, Vol. 7, No. 1, pp. 1-15 1997.

[8] James T. Lo and Thomas Wanner, "Existence and Uniqueness of Risk-Sensitive Estimates" IEEE Transaction on Automatic Control, Vol 47, No 11, pp 1945-1948, Feb 1991.

[9] Jason L. Speyer, Chih-hai Fan, and Ravi N. Banavar "Optimal Stochastic Estimation with Exponential Cost Criteria", Proceedings of the 31 Conference on Decision and Control, December 1992.

[10] Lihua Xie, Yeng Chai Soh and Carlos E. de Souza, "Robust Kalman Filtering for Uncertain Discrete Time Systems," IEEE Transaction on Automatic Control, Vol 39, No 6, pp 1310-1314, June 1994.

[11] Shovan Bhaumik, Smita Sadhu, Tapan Kumar Ghoshal, "Recursive Formulation and Properties of RSF", CKBS report no CKBSJU/TR/0705/2, Aug, 2005

[12] Smita Sadhu, A Doucet, "Particle Methods for Risk Sensitive Filtering" December 2005, Communicated to INDICON 2005.

[13] Shovan Bhaumik, Smita Sadhu, Tapan Kumar Ghoshal, "Adaptive Grid Solution of Risk Sensitive Estimator Problems", Communicated to INDICON 2005.

[14] Peter D. Hanlon, Peter S. Maybeck "Characterization of Kalman Filter Residuals in the Presence of Mismodeling", Proceedings of the 37th IEEE Conference on Decision & Control, Tampa, Florida USA December 1998