FUNDAMENTAL FILTERING LIMITATIONS IN LINEAR NON-GAUSSIAN SYSTEMS

Gustaf Hendeby * **Fredrik Gustafsson** *

* *Division of Automatic Control*
Department of Electrical Engineering,
Linköpings universitet, SE-581 83 Linköping, Sweden
{hendeby, fredrik}@isy.liu.se

Abstract: The Kalman filter is known to be the optimal *linear* filter for linear non-Gaussian systems. However, non-linear filters such as Kalman filter banks and more recent numerical methods such as the particle filter are sometimes superior in performance. Here a procedure to *a priori* decide how much can be gained using non-linear filters, without having to resort to Monte Carlo simulations, is outlined. The procedure is derived in terms of the Cramér-Rao lower bound. Explicit results are shown for a number of standard distributions and models in practice.

Keywords: Kalman filters; Linear filters; Cramér-Rao Lower Bound; Non-linear filters; Optimal filtering

1. INTRODUCTION

Consider a linear non-Gaussian system with state vector x_t, process noise w_t, and measurement noise e_t:

$$x_{t+1} = F_t x_t + G_t w_t, \qquad w_t \sim p_w \tag{1a}$$
$$y_t = H x_t + e_t, \qquad e_t \sim p_e. \tag{1b}$$

The Kalman filter (Kalman, 1960; Kailath *et al.*, 2000) minimizes the covariance matrix among all linear filters. The resulting covariance matrix $P_t = \mathsf{cov}(x_t)$ is given by the Riccati equation, which obeys the functional recursion:

$$P_{t+1}^{\text{KF}} = \kappa(P_t, F_t, G_t, H_t, Q_t, R_t). \tag{2}$$

There might, however, exist non-linear filters that perform better. For instance, in target tracking literature, the state noise models pilot maneuvers, and the *interactive multiple model* (IMM) algorithm (Blom and Bar-Shalom, 1988; Blackman and Popoli, 1999) has become a standard tool in this context. Other examples include a multi-modal measurement noise distribution for radar sensors used in for instance (Bergman *et al.*, 1999), in which case the *particle filter* (Gordon *et al.*, 1993; Doucet *et al.*, 2001) has proven to yield good performance.

Using results in (Bergman, 2001) it will here be shown that the *Cramér-Rao lower bound* (CRLB) obeys the same functional form as the Riccati equation

$$P_{t+1}^{\text{CRLB}} = \kappa\big(P_t, F_t, G_t, H_t, \mathcal{I}_{w_t}^{-1}, \mathcal{I}_{e_t}^{-1}\big), \tag{3}$$

where $\mathcal{I}_{w_t}$ and $\mathcal{I}_{e_t}$ are the Fisher information of the noises w_t and e_t, respectively.

The Gaussian distribution act as a worst case distribution, in that $P_t^{\text{CRLB}} \preceq P_t^{\text{KF}}$ with equality if and only if both process and measurement noise are Gaussian. For all other cases, a non-linear filter *might* perform better, depending on the implementation. For instance, the particle filter with sufficiently many particles will always, in theory, reach the CRLB, at the price of high computational complexity.

The subject of this paper is to derive formulas to decide how much better performance we can hope for by resorting to non-linear filtering. If the gain is very small, is it hard to motivate not using the Kalman filter. In other cases, the noise distribution may reveal much more information than a Gaussian second order equivalent, and the performance can be improved considerably. The results can also be used in practice for tuning,

since when the achieved filter performance has reached, or come close to, the CRLB further tuning is useless.

Though more general results for the CRLB exist (Tichavský *et al.*, 1998; Bergman, 1999) for non-linear non-Gaussian systems, studying the linear non-Gaussian case simplifies the CRLB expression to something that is easy to comprehend and use in practice. It furthermore allows a direct comparison to the best linear filter (the Kalman filter).

The paper will first discuss information of distributions before going on to find the CRLB for linear systems. Simulations are then used to exemplify the theory and finally draw some conclusions about the findings.

2. INFORMATION IN DISTRIBUTIONS

This section first defines the *Fisher Information* (FI) and *Relative Information* (RI) of distributions, and then presents some results regarding distributions of different kinds.

2.1 Fisher Information (FI)

The Fisher Information is defined (Kay, 1993), under mild conditions on p, as

$$\mathcal{I}_x = \mathcal{I}(x) := -\mathsf{E}\left(\frac{\partial^2 \log p(\xi|x)}{\partial x^2}\right), \qquad (4)$$

and is closely tied to the CRLB through the relation

$$\mathsf{cov}(x) \succeq \mathcal{I}^{-1}(x) = P^{\text{CRLB}}, \qquad (5)$$

with equality if and only if x is Gaussian.

2.2 Relative Information (RI)

Define *Relative Information*, Ψ, as

$$\Psi_x = \Psi(x) := \mathsf{cov}^{-1}(x)\,\mathcal{I}^{-1}(x). \qquad (6)$$

The RI is then a measure of how much information is available in moments higher than the second one of the distribution. Another interpretation of RI is as a measure of how much more informative the distribution of x is compared to a Gaussian distribution with the same variance. It also follows that $\|\Psi(x)\| \leq 1$ with equality if and only if x is Gaussian.

2.3 Bi-Gaussian Distribution

One type of non-Gaussian noise that occurs naturally is bi-Gaussian noise, that can be observed in *e.g.*, radar applications (Bergman *et al.*, 1999; Bergman, 1999; Dahlgren, 1998) or as an description of outliers.

All bi-Gaussian distributions can be parameterized using the parameters $\alpha_1 > 0$, $\alpha_2 > 0$, μ_1, μ_2, $R_1 > 0$, and $R_2 > 0$, with $\alpha_i > 0$ and $\sum_i \alpha_i = 1$, yielding

$$p(x) = \alpha_1 \mathcal{N}(x; \mu_1, R_1) + \alpha_2 \mathcal{N}(x; \mu_2, R_2). \qquad (7)$$

There is no closed expression for FI or the RI for bi-Gaussian distributions, but it can nevertheless be computed using Monte Carlo integration (Robert and Casella, 1999).

Other statistical properties of a bi-Gaussian distribution is its skewness, γ_1, and kurtosis, γ_2,

$$\gamma_1 = \sum_{i=1}^{2} \alpha_i \bar{\mu}(3R_i + \bar{\mu}_i^2) \cdot R^{-\frac{3}{2}} \qquad (8a)$$

$$\gamma_2 = \sum_{i=1}^{2} \alpha_i (3R_i^2 + 6\bar{\mu}_i^2 R_i + \bar{\mu}_i^4) \cdot R^{-2} - 3, \qquad (8b)$$

with R the variance of the distribution, μ its mean, and $\bar{\mu}_i = \mu_i - \mu$. Compare this with the Gaussian distribution where $\gamma_1 = \gamma_2 = 0$.

Example: Bi-Gaussian Noise This section exemplifies the bi-Gaussian distributions by looking at a subset of them where μ_1 and R_1 are used as parameters, $\alpha = 0.9$, and μ_2 and R_2 are used to obtain zero mean and unit variance. These distributions can be expressed as

$$\begin{aligned} p(e) = \; & \tfrac{9}{10}\mathcal{N}(e; \mu, R) \\ & + \tfrac{1}{10}\mathcal{N}(e; 9\mu, 10 - 18\mu^2 - 9R), \end{aligned} \qquad (9)$$

with obvious restrictions on μ and R to obtain valid distributions.

The RI of the distributions in (9) has been calculated and a contour plot of the result is found in Fig. 1. The parameter pair $\mu = 0.2$ and $R = 0.3$ is marked with an $\times$ in the figure, and its PDF is on displayed in Fig. 2. This distribution has $\Psi = 0.37$, $\gamma_1 = -5.1$, and $\gamma_2 = 9.7$, mainly due to its one heavy tail.

2.4 Tri-Gaussian Distribution

Tri-Gaussian distributions are made up of three Gaussian contributions,

$$p(x) = \sum_{i=1}^{3} \alpha_i \mathcal{N}(x; \mu_i, R_i), \qquad (10)$$

with $\sum_i \alpha_i = 1$, $\alpha_i > 0$, and $R_i > 0$.

Just as for the bi-Gaussian is it possible to numerically compute the relative information for this type of distributions and skewness and kurtosis follows analogously to (8) by summing to 3 instead of 2.

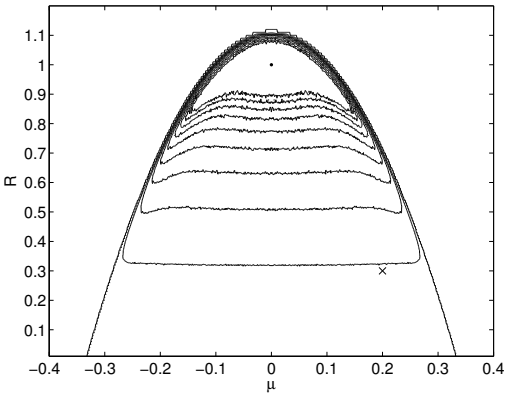

Fig. 1. The RI, $\Psi(e)$, for the bi-Gaussian (9). (Levels: [0.99, 0.98, 0.97, 0.95, 0.92, 0.87, 0.78, 0.64, 0.40, 0], 0 being the outermost level. $\times$ denotes the studied distribution.)

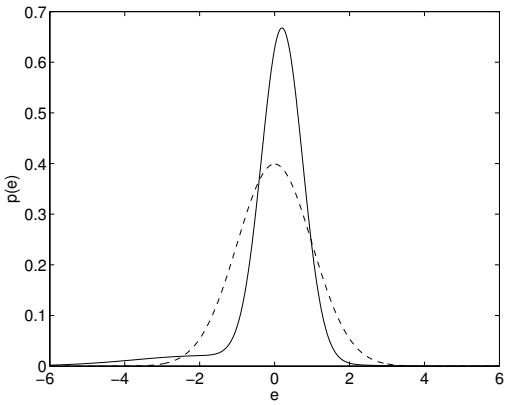

Fig. 2. PDF of (11) with $\mu = 0.2$ and $R = 0.3$. $\mathsf{var}(e) = 1$ and $\Psi(e) = 0.3$. (The dashed curve is the Gaussian approximation.)

Example: Tri-Gaussian Noise A subset of the tri-Gaussian distribution can be obtained by letting α_2 and μ_1 be parameters and further enforcing $\alpha_1 = \alpha_3$, $\mu_3 = -\mu_1$, $\mu_2 = 0$, $R = R_i$ and then use the remaining degrees of freedom to achieve unit variance. This parameterization yields

$$\begin{aligned} p(w) = \quad & \tfrac{1-\alpha}{2}\,\mathcal{N}(w; -\mu, 1-\mu^2(1-\alpha)) \\ & +\alpha\,\mathcal{N}(w; 0, R) \\ & +\tfrac{1-\alpha}{2}\,\mathcal{N}(w; +\mu, 1-\mu^2(1-\alpha)). \end{aligned} \tag{11}$$

Some restrictions apply to the values of α and μ in order to get proper distributions.

This type of distributions can be used to model certain kinds of multiple model systems. As an example, suppose w is process noise in a motion model, then the different modes represent a change in speed or acceleration with approximately the same probability as the matching Gaussian component. Other similar examples exist.

Fig. 3 shows the RI of (11) using α and μ as parameters. The plot displays an important difference between the studied bi-Gaussian noises and this tri-Gaussian noise. For the latter, most of the parameter space has RI close to 1, and then falls off very rapidly when approaching the boarder of the allowed parameter region, whereas the former has a slower behavior, especially in R. Hence, the exact parameter values in the bi-Gaussian case is less important for the information content in this case (for most of the parameter space) than in the tri-Gaussian case where a slight change may cause all extra information to disappear.

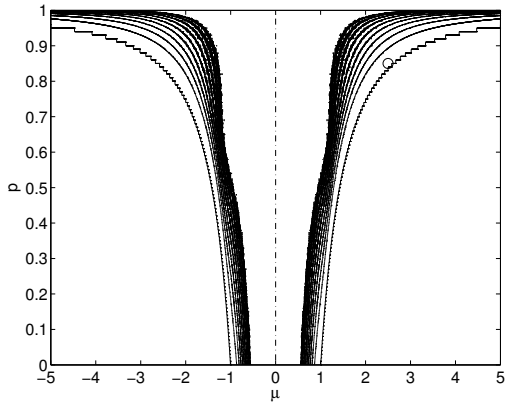

Fig. 3. The RI, $\Psi(w)$, for the tri-Gaussian (11). (Levels: [0.99, 0.98, 0.97, 0.95, 0.92, 0.87, 0.78, 0.64, 0.40, 0], 0 being the outermost level. $\circ$ denotes the studied distribution.)

The position in Fig. 3 marked with an $\circ$ represents the parameter values $\alpha = 0.85$ and $\mu = 2.5$ that result in the PDF in Fig. 4. This tri-Gaussian is distinctly tri-modal, and does this way get $\Psi = 0.065$, $\gamma_1 = 0$, and $\gamma_2 = 3.4$. Note also how steep the Relative Information curve is around these parameter values. A minor change in any parameter will change the result considerably, most likely to make the distribution less informative.

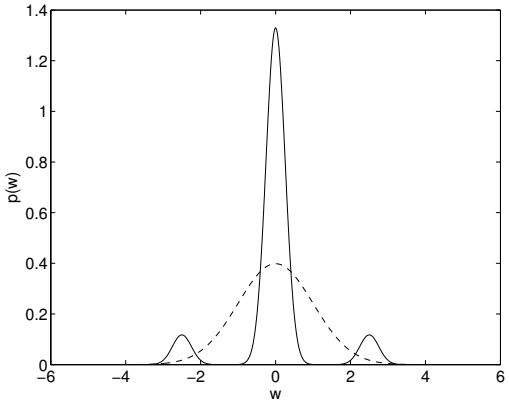

Fig. 4. PDF of (11) with $\alpha = 0.85$ and $\mu = 2.5$. $\mathsf{var}(w) = 1$ and $\Psi(w) = 0.065$. (Dashed curve, the Gaussian approximation.)

3. CRLB FOR FILTERING

This section first presents a general expression for the CRLB for dynamic systems and then derive

expressions for linear systems which are then further discussed.

3.1 General Systems

Consider a general system described by

$$\begin{cases} x_{t+1} = f(x_t, w_t) \\ y_t = h(x_t, e_t) \end{cases} \quad \longleftrightarrow \quad \begin{cases} p(x_{t+1}|x_t) \\ p(y_t|x_t) \end{cases}. \quad (12)$$

The CRLB for this system, $P^{\text{CRLB}}_{t|t} = P_{t|t}$, is given in (Bergman, 1999, Theorem 4.5) by the recursion

$$P^{-1}_{t+1} = \tilde{Q}_t - \tilde{S}^T_t (P^{-1}_t + \tilde{R}_t + \tilde{V}_t)^{-1} \tilde{S}_t, \quad (13)$$

where

$$\begin{aligned} \tilde{Q}_t &= \mathsf{E}\left(-\Delta^{x_{t+1}}_{x_{t+1}} \log p(x_{t+1}|x_t)\right), \\ \tilde{R}_t &= \mathsf{E}\left(-\Delta^{x_t}_{x_t} \log p(y_t|x_t)\right), \\ \tilde{S}_t &= \mathsf{E}\left(-\Delta^{x_{t+1}}_{x_t} \log p(x_{t+1}|x_t)\right), \\ \tilde{V}_t &= \mathsf{E}\left(-\Delta^{x_t}_{x_t} \log p(x_{t+1}|x_t)\right) \end{aligned}$$

and the iteration is initiated with

$$P^{-1}_0 = \mathsf{E}\left(-\Delta^{x_0}_{x_0} \log p(x_0)\right).$$

The quantities $\tilde{V}_t$, $\tilde{R}_t$, $\tilde{S}_t$, $\tilde{Q}_t$, and $P^{-1}_{0|0}$ are all closely related to the Fisher Information of different aspects of the the system.

3.2 Linear Systems

In the case of a linear system,

$$x_{t+1} = F_t x_t + w_t, \qquad \mathsf{cov}\, w_t = Q_t \quad (14a)$$

$$y_t = H x_t + e_t, \qquad \mathsf{cov}\, e_t = R_t, \quad (14b)$$

the parameters are given by (Bergman, 1999)

$$\begin{aligned} \tilde{Q}_t &= \mathcal{I}_{w_t} & \tilde{R}_t &= H^T_t \mathcal{I}_{e_t} H_t \\ \tilde{S}_t &= -F^T_t \mathcal{I}_{w_t} & \tilde{V}_t &= F^T_t \mathcal{I}_{w_t} F_t. \end{aligned}$$

Using these relations the equivalent to (13) becomes

$$\begin{aligned} P^{-1}_{t+1} &= \mathcal{I}_{w_t} - \mathcal{I}_{w_t} F_t \big(P^{-1}_t + \\ &\quad + H^T_t \mathcal{I}_{e_t} H_t + F^T_t \mathcal{I}_{w_t} F_t\big)^{-1} F^T_t \mathcal{I}_{w_t}. \end{aligned} \quad (15)$$

By inverting this expression (using the matrix inversion lemma [1] repeatedly) the standard Riccati equation appears,

$$\begin{aligned} P_{t+1} &= F^T_t (P^{-1}_t + H_t \mathcal{I}_{e_t} H^T_t)^{-1} F_t + \mathcal{I}^{-1}_{w_t} = \\ &= F^T_t P_t F - F^T_t P_t H_t \cdot \\ &\quad \cdot (H^T_t P_t H_t + \mathcal{I}^{-1}_{e_t})^{-1} H^T_t P_t F_t + \mathcal{I}^{-1}_{w_t}. \end{aligned} \quad (16)$$

If $\mathcal{I}_{w_t}$ is singular this can be solved by using $G_t \mathcal{I}(\bar{w}_t)^{-1} G^T_t$ instead of $\mathcal{I}(w_t)^{-1}$ where G_t and $\bar{w}_t$ are such that $w_t = G_t \bar{w}_t$ and $\mathcal{I}(\bar{w}_1)$ non-singular (Bergman, 1999).

Also, note that (16) is the standard Riccati equation for the Kalman filter for $\mathcal{I}^{-1}_{w_t} = Q_t$ and $\mathcal{I}^{-1}_{e_t} = R_t$.

[1] $(A+BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$ given that A^{-1} and C^{-1} are well defined.

Stationary Properties When $t \to \infty$ (stationary state),

$$P_t = P_{t+1} =: \bar{\kappa}\big(\mathcal{I}^{-1}(w_t), \mathcal{I}^{-1}(e_t)\big),$$

the following simple rules hold for $\bar{\kappa}$ (the system is here kept out of the notation for clarity).

For general matrices Q and R and scalar γ:

(i) $\bar{\kappa}(\gamma Q, \gamma R) = \gamma \bar{\kappa}(Q, R)$
(ii) $\bar{\kappa}(Q, \gamma R) = \gamma \bar{\kappa}(\frac{1}{\gamma} Q, R)$
(iii) $\bar{\kappa}(\gamma Q, R) = \gamma \bar{\kappa}(Q, \frac{1}{\gamma} R)$
(iv) $Q_1 \preceq Q_2 \wedge R_1 \preceq R_2 \Rightarrow \bar{\kappa}(Q_1, R_1) \preceq \bar{\kappa}(Q_2, R_2)$ with equality if and only if $Q_1 = Q_2$ and $R_1 = R_2$.

The properties (i)–(iii) are equivalent so it is enough to show one of them, *e.g.*, (i).

If $\bar{P}$ is a solution to (15), then $\gamma \bar{P}$ is a solution to

$$\begin{aligned} P^{-1} &= \tfrac{1}{\gamma} \Psi_{w_t} - \tfrac{1}{\gamma} \Psi^{-1}_{w_t} F^T_t (P^{-1} + \\ &\quad + H_t \tfrac{1}{\gamma} \Psi_{e_t} H^T_t + F_t \tfrac{1}{\gamma} \Psi_{w_t} F^T_t)^{-1} F_t \tfrac{1}{\gamma} \Psi_{w_t}, \end{aligned}$$

and hence $\bar{\kappa}(\gamma Q, \gamma R) = \gamma \bar{\kappa}(Q, R)$.

Property (iv) follows from the fact that the Kalman estimate always improves when either of the noises decreases, *i.e.*, $Q_1 \preceq Q_2$ and $R_1 \preceq R_2$, and that (15) is the same Riccati equation as in the Kalman filter with just a different interpretation of the included matrices.

4. SIMULATIONS

In this section the theory previously presented will be illustrated using the simulations.

4.1 System

The following system will be used in the simulations,

$$x_{t+1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} x_t + \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} w_t, \quad \mathsf{cov}\, w_t = Q \quad (17a)$$

$$y_t = \begin{pmatrix} 1 & 0 \end{pmatrix} x_t + e_t, \qquad \mathsf{cov}\, e_t = R \quad (17b)$$

with w_t and e_t mutually independent, white noise with unit variance. The system represents a second order random walk, or a double integrator. The best stationary estimation variance of $x^{(1)}$ (the observed state) using a linear (Kalman) filter is $\bar{\kappa}_{(1,1)}(1, 1) = 3.0$.

Since in this case both w_t and e_t are scalar it is possible to, in a regular contour plot (looking at $x^{(1)}$ only), show how the optimal performance varies as the information in the noise changes, *i.e.*, how non-Gaussian noise affects the filter performance. In Fig. 5 the filtering performance $\bar{\kappa}_{(1,1)}(\Psi_{w_t} Q, \Psi_{e_t} R)/\bar{\kappa}_{(1,1)}(Q, R)$ is presented as a function of the RI involved ($Q = 1$, $R = 1$). Note,

however that it is impossible to in this plot see how to obtain this optimal performance or how difficult it is.

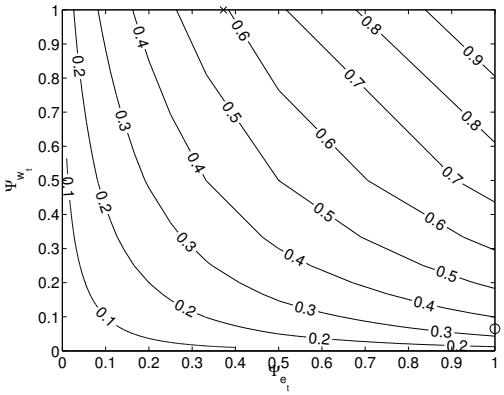

Fig. 5. Optimal filter performance, $\bar{\kappa}_{(1,1)}(\Psi_{w_t}, \Psi_{e_t})/\bar{\kappa}_{(1,1)}(1,1)$ as a function of Ψ_{w_t} and Ψ_{e_t}. ($\times$ denotes the noise in the first example and $\circ$ the noise in the second example.)

4.2 Bi-Gaussian Measurements

The first example uses non-Gaussian measurement noise

$$e_t \sim \tfrac{9}{10}\mathcal{N}(0.2, 0.3) + \tfrac{1}{10}\mathcal{N}(-1.8, 3.7).$$

This specific noise was discussed in Sec. 2.3 and has one heavy tail. From the Fig. 5, or by solving the appropriate Riccati equation, (16), the CRLB for this system with this specific measurement noise can be found to be $\bar{\kappa}(1, 0.37) = 1.8$, *i.e.*, the optimal variance is 60% of what is obtainable with any linear filters. Hence, this seems to be a candidate for a non-linear filter.

Therefore the system was simulated, and both a Kalman filter and a particle filters (using the SIR algorithm and 50 000 particles) was applied. The mean square error (MSE) of these estimates was then computed for 1000 Monte Carlo simulations. The MSE together with the theoretical stationary limits are plotted in Fig. 7.

The figure shows a significant gain from using the particle filter (approximately 18% lower variance), but the CRLB is not reached. This is a combination of at least two factors.

One reason is that the actual e_t and the second order Gaussian equivalent differs first in the third moment. The Gauss has $\gamma_1 = 0$ whereas e_t has $\gamma_1 = -5.1$. This means that the noise distribution must be kept accurate to at least third order moments in order to be able to extract the extra information about the state. This requires a large amount of particles in the particle filter to reach the CRLB. Due to computational complexity and numerical issues, especially in the resampling step,

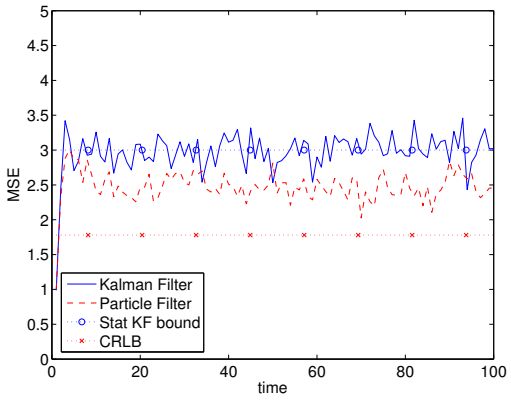

Fig. 6. MSE of 1000 Monte Carlo simulations with a Kalman filter and a particle filter (SIR, 50 000 particles) on the system with bi-Gaussian measurement noise. (Theoretical limits are include as reference.)

it is not surprising that the optimum is not obtained.

Furthermore, since all extra information is carried by the noise, many noise instances must be evaluated in order for the improved variance to show up in Monte Carlo simulations. With few simulations the simulated noise could as well come from a Gaussian distribution and the improved performance does not show to its full extent.

4.3 Tri-Gaussian Process Noise

In this example the measurements will be kept Gaussian whereas the system is driven by tri-modal noise,

$$\begin{aligned} w_t \sim \quad & 0.075\mathcal{N}(-2.5, 0.065) \\ & +0.85\mathcal{N}(0, 0.065) \\ & +0.075\mathcal{N}(+2.5, 0.065), \end{aligned}$$

with $\mathsf{var}(w_t) = 1$, $\Psi_{w_t} = 0.065$, $\gamma_1 = 0$, and $\gamma_2 = 3.4$. This tri-Gaussian was discussed in Sec. 2.4.

Since the Gaussian approximation here is the same as above, $\bar{\kappa}(1, 1) = 3.0$. However, the CRLB for this system is different, $\bar{\kappa}(0.065, 1) = 1.77$. Once again, consult Fig. 5 or solve (16).

Simulating this system and applying a Kalman filter and a particle filter (SIR algorithm with 50 000 particles) yields for 1000 Monte Carlo simulations the result in Fig. 7.

Here the particle filter is not significantly better than the Kalman filter (time mean shows a 3% improvement for the particle filter).

The reason for this seems to be the same as for the last example, just with more serious effects. The Gauss approximation and w_t differs first in the kurtosis, and even there $\gamma_2 = 3.4$ for w_t is not very much. This indicate that even more particles

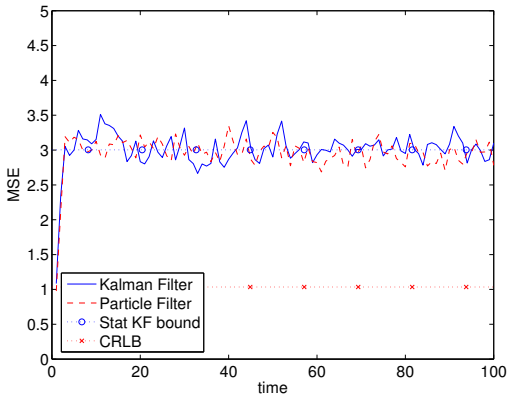

Fig. 7. MSE of 1000 Monte Carlo simulations with a Kalman filter and a particle filter (SIR, 50 000 particles) on the system with tri-Gaussian process noise. (Theoretical limits are include as reference.)

are necessary for good results. Approximately the same argumentation calls for many Monte Carlo simulations as well.

Another difficulty with this example shows in Fig. 3. It was noted already in Sec. 2.4, that due to the steep RI curve, distributions close to the chosen one could be much less informative. With few samples in the particle filter or noise instances in the Monte Carlo simulations the actual distributions are hard to distinguish between, and this should result in higher estimation variance.

All these explanations motivate the poor performance of the particle filter in these examples. This not to say that the CRLB is not reachable, but other filtering methods than the particle filter, and probably in combination with other evaluation methods than Monte Carlo simulations would probably be preferable.

5. CONCLUSIONS

In this paper, starting from general CRLB expressions, an expression for the CRLB in linear systems is derived in terms of Relative Information (RI) of the included noises. RI is defined as a measure of the information available in higher moments than the second one of a stochastic variable. This results in a method to, given a system and its RI, calculate the CRLB by solving a few Riccati equations. The CRLB can then *e.g.*, be used to decide if it is worthwhile to try non-linear filtering or decide when no more tuning is needed, all this without resorting to time consuming Monte Carlo simulations.

Simulations are also presented to support the theoretical results. The examples indicate improved performance using a particle filter on linear systems with non-Gaussian noise, but do also point out the difficulty of reaching the CRLB. Basically, it is a combination of properties of the noise, the system, and the evaluation method used.

ACKNOWLEDGEMENTS

This work is supported by the competence center ISIS (Information Systems for Industrial Control and Supervision) at Linköpings universitet.

REFERENCES

Bergman, N. (1999). Recursive Bayesian Estimation: Navigation and Tracking Applications. Ph.D. thesis no 579. Linköping Studies in Science and Technology. SE-581 83 Linköping, Sweden.

Bergman, N. (2001). *Posterior Cramér-Rao Bounds for Sequential Estimation*. Chap. 15, pp. 321–338. in Doucet *et al.* (2001).

Bergman, N., L. Ljung and F. Gustafsson (1999). Terrain navigation using Bayesian statistics. *IEEE Control Systems Magazine* **19**(3), 33–40.

Blackman, S. S. and R. Popoli (1999). *Design and analysis of modern tracking systems*. Artech House radar library. Artech House, Inc.

Blom, H. A. P. and Y. Bar-Shalom (1988). The interacting multiple model algorithm for systems with Markovian switching coefficients. *IEEE Transactions on Automatic Control* **33**(8), 780–783.

Dahlgren, C. (1998). Nonlinear black box modelling of JAS 39 Gripen's radar altimeter. Master's thesis n:o LiTH-ISY-EX-1958. Department of Electrical Engineering, Linköpings universitet, Sweden.

Doucet, A., de Freitas, N. and Gordon, N. (Eds.) (2001). *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. Springer Verlag. New York.

Gordon, N. J., D. J. Salmond and A. F. M. Smith (1993). Novel approach to nonlinear/non-Gausian Bayesian state estimation. In: *IEE Proceedings-F*. Vol. 140(2). pp. 107–113.

Kailath, T., A. H. Sayed and B. Hassibi (2000). *Linear Estimation*. Prentice Hall Inc.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering* **82**(Series D), 35–45.

Kay, S. M. (1993). *Fundamentals of Statistical Signal Processing: Estimation Theory*. Vol. 1. Prentice Hall, Inc.

Robert, C. P. and G. Casella (1999). *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer-Verlag. New York.

Tichavský, P., C. H. Muravchik and A. Nehorai (1998). Posterior Cramér-Rao discrete-time nonlinear filtering. *IEEE Transactions on Signal Processing* **46**(5), 1386–1396.