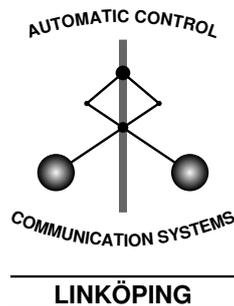


Fundamental Filtering Limitations in Linear Non-Gaussian Systems

Gustaf Hendeby, Fredrik Gustafsson

Control & Communication
Department of Electrical Engineering
Linköpings universitet, SE-581 83 Linköping, Sweden
WWW: <http://www.control.isy.liu.se>
E-mail: hendeby@isy.liu.se, fredrik@isy.liu.se

9th March 2005



Report no.: [LiTH-ISY-R-2681](#)

Submitted to 16th IFAC World Congress

Technical reports from the Control & Communication group in Linköping are available at
<http://www.control.isy.liu.se/publications>.

Abstract:

The Kalman filter is known to be the optimal *linear* filter for linear non-Gaussian systems. However, nonlinear 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 nonlinear filters, without having to resort to Monte Carlo simulations, is outlined. The procedure is derived in terms of the posterior Cramér-Rao lower bound. Results are shown for a class of standard distributions and models in practice.

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

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, nonlinear 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 nonlinear filters, without having to resort to Monte Carlo simulations, is outlined. The procedure is derived in terms of the posterior Cramér-Rao lower bound. Results are shown for a class of standard distributions and models in practice.

Keywords: Kalman filters; Linear filters; Cramér-Rao Lower Bound; Nonlinear 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 , both noises white and mutually independent:

$$\begin{aligned} x_{t+1} &= F_t x_t + G_t w_t, & w_t &\sim p_w(\cdot), \\ y_t &= H x_t + e_t, & e_t &\sim p_e(\cdot). \end{aligned}$$

The *Kalman filter* (KF) (Kalman, 1960; Kailath *et al.*, 2000) minimizes the covariance matrix among all *linear* filters. The resulting covariance matrix $P_{t+1|t} = \text{cov}(\hat{x}_{t+1|t})$ is given by a *Riccati equation*, which obeys a functional recursion:

$$P_{t+1|t}^{\text{KF}} = \kappa(P_{t|t-1}^{\text{KF}}, F_t, G_t, H_t, Q_t, R_t). \quad (1)$$

There might, however, exist *nonlinear* filters that perform better. For instance, in target tracking, the state noise models pilot maneuvers, and the *interactive multiple model* (IMM) algorithm (Blom and Bar-Shalom, 1988) 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* (PF) (Gordon *et al.*, 1993; Doucet *et al.*, 2001) has proven to yield good performance.

Using results in (Bergman, 2001), it will in Sec. 3 be shown that the posterior *Cramér-Rao lower bound* (CRLB) obeys the same functional form as the KF Riccati equation

$$P_{t+1|t}^{\text{CRLB}} = \kappa(P_{t|t-1}^{\text{CRLB}}, F_t, G_t, H_t, \mathcal{I}_{w_t}^{-1}, \mathcal{I}_{e_t}^{-1}), \quad (2)$$

where \mathcal{I}_{w_t} and \mathcal{I}_{e_t} denote the *intrinsic accuracy* (IA) (Sec. 2.1) of the noises w_t and e_t , respectively.

It will also be shown that the Gaussian distribution acts as a worst case distribution, in that $P_{t+1|t}^{\text{CRLB}} \preceq P_{t+1|t}^{\text{KF}}$ with equality if and only if both process and measurement noise are Gaussian. For all other cases, a nonlinear filter *might* perform better, depending on the implementation. For instance, the PF with sufficiently many particles will always, in theory, compute the true posterior state distribution.

Formulas are, in this paper, derived to decide how much better performance can be hoped for by resorting to nonlinear filtering instead of linear filtering. If the gain is very small, it is hard to motivate using anything but the KF. In other cases, the noise distributions may provide much more information than a Gaussian second order equivalent does, 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 nonlinear non-Gaussian systems, studying the linear non-Gaussian case simplifies the CRLB expression to something that is easy to comprehend and use in practice. Furthermore, it allows for direct comparison with the *best linear unbiased estimator* (BLUE) — the KF.

The paper will first discuss information and accuracy before determining the CRLB for linear systems. Simulations then exemplify the presented theory. Finally conclusions are drawn.

2. INFORMATION AND ACCURACY

This section defines *Fisher information* (FI), *intrinsic accuracy* (IA), and *relative accuracy* (RA). Then results regarding distributions of different kinds are presented.

2.1 Fisher Information and Intrinsic Accuracy

The *Fisher information* (FI) is defined (Kay, 1993), under mild regularity conditions on the *probability density distribution* (PDF) of ξ , as

$$\mathcal{I}_\xi(\theta) := -E_\xi(\Delta_\theta^\theta \log p(\xi|\theta)) \quad (3)$$

evaluated for the true value of θ , where Δ denotes the *Hessian*. The FI is closely tied to the CRLB through the relation (Kay, 1993)

$$\text{cov}(\hat{\theta}) \succeq P_\theta^{\text{CRLB}} = \mathcal{I}_\xi^{-1}(\theta), \quad (4)$$

where $A \succeq 0$ denote that A is a positive semidefinite matrix. When nothing else is explicitly stated in this paper, the information is taken to be with respect to the mean of the distribution in question assumed to be 0, *i.e.*, $\theta = E(\xi) = \mu = 0$. The notation $\mathcal{I}_\xi = \mathcal{I}_\xi(\mu)$ will be used. This quantity is in (Kay and Sengupta, 1987; Kay, 1998; Cox and Hinkley, 1974) referred to as the *intrinsic accuracy* (IA) of the PDF for ξ .

2.2 Relative Accuracy

Define *relative accuracy* (RA) as the scalar constant Ψ_ξ that fulfills

$$\text{cov}(\xi) = \Psi_\xi \mathcal{I}_\xi^{-1}, \quad (5)$$

when such a scalar exists.

The RA is thus a relation between IA and covariance. RA can be interpreted as a measure of how much more informative the distribution of ξ is compared to a Gaussian distribution with the same covariance. It can be shown (Kay and Sengupta, 1987) that $\Psi_\xi \geq 1$, with equality if and only if ξ is Gaussian.

2.3 Gaussian Mixture Distribution

A class of suitable distributions for analysis is Gaussian mixtures. Gaussian mixtures can be used to approximate any distribution (Sorenson and Alspach, 1971; Anderson and Moore, 1979) and numerical methods apply well. This section will define Gaussian mixtures and study the accuracy concept using special cases.

Gaussian mixture distributions, with n_δ modes denoted δ , can be parameterized using the parameters $\omega_\delta > 0$, μ_δ , and $R_\delta \succeq 0$, representing the probability, the mean, and the variance, respectively, of the different modes, as

$$p(\xi) = \sum_\delta \omega_\delta \mathcal{N}(\xi; \mu_\delta, R_\delta), \quad (6)$$

with $\sum_\delta \omega_\delta = 1$.

There exist in general no closed expressions for IA or RA for multi-Gaussian distributions. They can nevertheless be computed using Monte Carlo integration (Robert and Casella, 1999).

Other statistical properties of multi-Gaussian distributions are mean, μ , covariance, R , skewness, γ_1 , and kurtosis, γ_2 ,

$$\mu = \sum_\delta \omega_\delta \mu_\delta \quad (7a)$$

$$R = \sum_\delta \omega_\delta (R_\delta + \bar{\mu}_\delta) \quad (7b)$$

$$\gamma_1 = \sum_\delta \omega_\delta \bar{\mu} (3R_\delta + \bar{\mu}_\delta^2) R^{-\frac{3}{2}} \quad (7c)$$

$$\gamma_2 = \sum_\delta \omega_\delta (3R_\delta^2 + 6\bar{\mu}_\delta^2 R_\delta + \bar{\mu}_\delta^4) R^{-2} - 3, \quad (7d)$$

where $\bar{\mu}_\delta = \mu_\delta - \mu$. Compare this to Gaussian distributions where $\gamma_1 = \gamma_2 = 0$.

Bi-Gaussian Noise: One type of Gaussian mixtures 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 a description of outliers (Gustafsson, 2000).

Bi-Gaussian distributions ($n_\delta = 2$) will be illustrated using a zero mean unit variance subset of them, (*cf.* $\mathcal{N}(0, 1)$) where μ and R are parameters of a dominating Gaussian contribution ($\omega_\delta = \frac{9}{10}$ is used in the sequel). These distributions can be expressed as

$$p(e) = \frac{9}{10} \mathcal{N}(e; \mu, R) + \frac{1}{10} \mathcal{N}(e; 9\mu, 10 - 90\mu^2 - 9R), \quad (8)$$

for $10 - 90\mu^2 - 9R > 0$.

The RA of the distributions in (8) is presented in Fig. 1. Since $\Psi_e^{-1} = \mathcal{I}_e$ for $R = 1$, Ψ_e^{-1} tells how good a parameter in noise can be estimated using all information in a specific noise distribution. If $R \neq 1$ then Ψ_e is a relative comparison to Gaussian noise, or a linear estimator, instead. Areas with high RA provide

better conditions for accurate estimation than areas with lower RA.

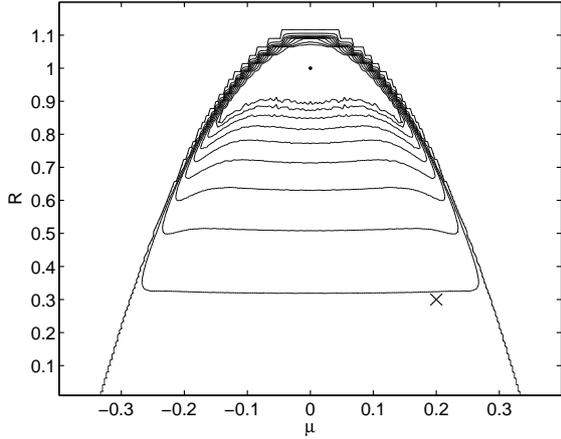


Fig. 1. RA, Ψ_e^{-1} , for the bi-Gaussian in (8). (Level curves: [0.99, 0.98, 0.97, 0.95, 0.92, 0.87, 0.78, 0.64, 0.40, 0]. \times marks $\mu = 0.2$, $R = 0.3$, and $\Psi_e = 2.7 = 0.37^{-1}$.)

The parameter pair $\mu = 0.2$ and $R = 0.3$ (see Fig. 2) yields $\Psi_e = 2.7$, $\gamma_1 = -5.1$, and $\gamma_2 = 9.7$, mainly due to its one heavy tail.

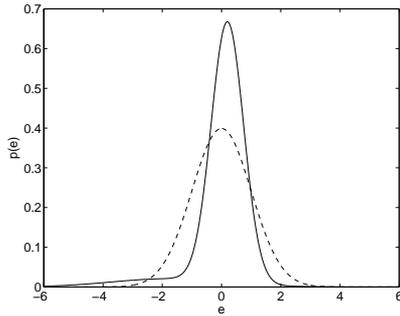


Fig. 2. Bi-Gaussian PDF (8) ($\mu = 0.2$ and $R = 0.3$) with $\Psi_e = 2.7$ and Gaussian approximation (dashed).

Tri-Gaussian Noise: Tri-Gaussian noise ($n_\delta = 3$) will be discussed using a *symmetric* zero mean unit variance subset of tri-Gaussian distributions (cf. $\mathcal{N}(0, 1)$) and can be obtained by letting ω determine the weight of the center mode, μ be a spread parameter, and enforcing the same variance on all modes. This parameterization yields

$$p(w) = \frac{1-\omega}{2}\mathcal{N}(w; -\mu, 1 - \mu^2(1 - \omega)) + \omega\mathcal{N}(w; 0, 1 - \mu^2(1 - \omega)) + \frac{1-\omega}{2}\mathcal{N}(w; +\mu, 1 - \mu^2(1 - \omega)), \quad (9)$$

for $1 - \mu^2(1 - \omega) > 0$.

This type of distributions can be used to model multiple model systems. For instance, 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 mode.

Fig. 3 shows the RA of (9) using ω and μ as parameters. With this parameterization the RA increases quickly as the modes are separated by increasing μ , this because the mode variances must decrease in order to achieve unit variance. At the border of the allowed parameter region the modes turn into point distributions explaining the infinite IA there.

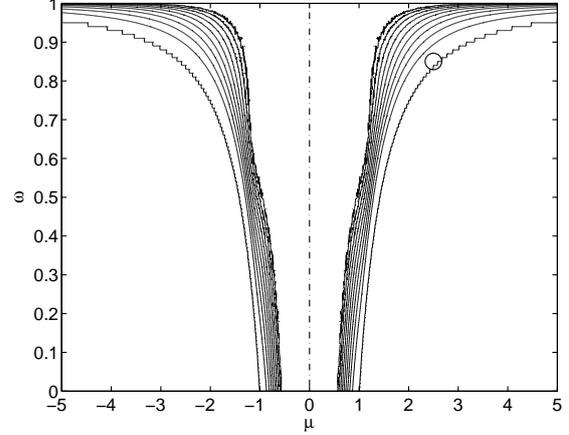


Fig. 3. RA, Ψ_w^{-1} , for the tri-Gaussian (9). (Level curves: [0.99, 0.98, 0.97, 0.95, 0.92, 0.87, 0.78, 0.64, 0.40, 0]. \circ denotes $\mu = 2.5$, $\omega = 0.58$, and $\Psi_w = 15.5 = 0.065^{-1}$.)

The parameter values $\omega = 0.85$ and $\mu = 2.5$ result in the PDF in Fig. 4. This tri-Gaussian is distinctly trimodal yielding $\Psi_w = 15.5$, $\gamma_1 = 0$, and $\gamma_2 = 3.4$.

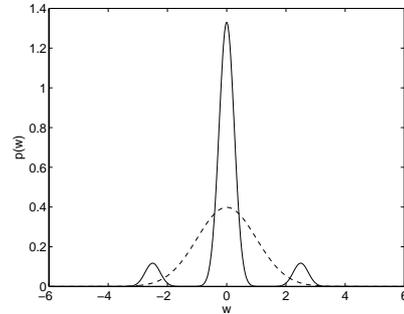


Fig. 4. Tri-Gaussian PDF (9) ($\omega = 0.85$ and $\mu = 2.5$) with $\Psi_w = 15.5$ and Gaussian approximation (dashed).

3. PREDICTION LIMITATIONS

This section first presents a general expression for the posterior *Cramér-Rao lower bound* (CRLB) for dynamic systems and derives expressions for linear systems which are then discussed further.

3.1 General System

Consider a general system described by

$$\begin{cases} x_{t+1} = f(x_t, w_t) \\ y_t = h(x_t, e_t) \end{cases} \longleftrightarrow \begin{cases} p(x_{t+1}|x_t) \\ p(y_t|x_t) \end{cases}. \quad (10)$$

The CRLB for this system, $P_{t|t-1}^{\text{CRLB}} = P_{t|t-1}$, is given in (Bergman, 1999, Theorem 4.5) by the recursion

$$P_{t+1|t}^{-1} = \tilde{Q}_t - \tilde{S}_t^T (P_{t|t-1}^{-1} + \tilde{R}_t + \tilde{V}_t)^{-1} \tilde{S}_t, \quad (11)$$

where

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

and the iteration is initiated with

$$P_0^{-1} = \mathbb{E} \left(-\Delta_{x_0}^{x_0} \log p(x_0) \right).$$

The quantities $\tilde{V}_t, \tilde{R}_t, \tilde{S}_t, \tilde{Q}_t$, and P_0^{-1} are all closely related to the IA of different aspects of the system.

3.2 Linear System

For a linear system,

$$x_{t+1} = F_t x_t + w_t, \quad \text{cov } w_t = Q_t \quad (12a)$$

$$y_t = H_t x_t + e_t, \quad \text{cov } e_t = R_t, \quad (12b)$$

the quantities in (11) are (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 (11) is

$$\begin{aligned} P_{t+1|t}^{-1} &= \mathcal{I}_{w_t} - \mathcal{I}_{w_t} F_t (P_{t|t-1}^{-1} + \\ &+ H_t^T \mathcal{I}_{e_t} H_t + F_t^T \mathcal{I}_{w_t} F_t)^{-1} F_t^T \mathcal{I}_{w_t}. \quad (13) \end{aligned}$$

By inverting this expression (using the matrix inversion lemma¹ repeatedly) the standard Riccati equation appears,

$$\begin{aligned} P_{t+1|t} &= F_t^T (P_{t|t-1}^{-1} + H_t \mathcal{I}_{e_t}^{-1} H_t^T)^{-1} F_t + \mathcal{I}_{w_t}^{-1} \\ &= F_t^T P_{t|t-1} F_t - F_t^T P_{t|t-1} H_t \\ &\quad \cdot (H_t^T P_{t|t-1} H_t + \mathcal{I}_{e_t}^{-1})^{-1} \\ &\quad \cdot H_t^T P_{t|t-1} F_t + \mathcal{I}_{w_t}^{-1}. \quad (14) \end{aligned}$$

If \mathcal{I}_{w_t} is singular this can be handled 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 $\mathcal{I}_{\bar{w}_t}$ is non-singular (Bergman, 1999).

Note that (14) is the standard Riccati equation for the KF with $\mathcal{I}_{w_t}^{-1} = Q_t$ and $\mathcal{I}_{e_t}^{-1} = R_t$.

Stationary Properties: In stationary state ($t \rightarrow \infty$),

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

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

Theorem 1. For matrices $Q \succeq 0$ and $R \succeq 0$, and scalar $\gamma > 0$ the following hold.

- (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) $\hat{Q} \preceq Q$ and $\hat{R} \preceq R \Rightarrow \bar{\kappa}(\hat{Q}, \hat{R}) \preceq \bar{\kappa}(Q, R)$ with equality if and only if $\hat{Q} = Q$ and $\hat{R} = R$.

The properties of Theorem 1 hold for individual iterations, $P_{t+1|t} = \kappa(P_{t|t-1}, Q, R)$, as well. This motivates the statement $P_{t+1|t}^{\text{CRLB}} \preceq P_{t+1|t}^{\text{KF}}$ for (2) and (1).

Proof: The properties (i)–(iii) are equivalent (rewrite (ii) in terms of $\hat{Q} = \frac{1}{\gamma} Q$, and use $\hat{R} = \frac{1}{\gamma} R$ in (iii)) so it suffices to show (i).

If \bar{P} is a solution to (13), then $P = \gamma \bar{P}$ solves

$$\begin{aligned} P^{-1} &= \frac{1}{\gamma} Q^{-1} - \frac{1}{\gamma} Q^{-1} F^T (P^{-1} \\ &+ H \frac{1}{\gamma} R^{-1} H^T + F \frac{1}{\gamma} Q^{-1} F^T)^{-1} F \frac{1}{\gamma} Q^{-1}, \end{aligned}$$

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

Property (iv) can be shown with induction on the KF recursion. Start by looking upon the Riccati equation as the limit $t \rightarrow \infty$ of the recursion

$$\begin{aligned} P_{t+1|t} &= F P_{t|t} F^T + Q \\ P_{t|t}^{-1} &= P_{t|t-1}^{-1} + R^{-1}, \end{aligned}$$

initiated with $P_{0|0}$. Now, using induction it is possible to show that $Q \succeq \hat{Q}$ and $R \succeq \hat{R}$ yields $P_{t|t-1} \succeq \hat{P}_{t|t-1}$ for $t \geq 0$.

- (i) This hold for $t = 0$ since it is assumed that $P_{0|0} = \hat{P}_{0|0}$.
- (ii) Assume that the statement is true for $t = \tau$, i.e., $P_{\tau|\tau-1} \succeq \hat{P}_{\tau|\tau-1}$.
- (iii) Show for $t = \tau + 1$ it holds that $P_{\tau+1} \succeq \hat{P}_{\tau+1}$ given (ii):

$$\begin{aligned} &P_{\tau|\tau}^{-1} - \hat{P}_{\tau|\tau}^{-1} \\ &= (P_{\tau|\tau-1}^{-1} + R^{-1}) - (\hat{P}_{\tau|\tau-1}^{-1} + \hat{R}^{-1}) \\ &= \underbrace{(P_{\tau|\tau-1}^{-1} - \hat{P}_{\tau|\tau-1}^{-1})}_{\succeq 0} + \underbrace{(R^{-1} - \hat{R}^{-1})}_{\succeq 0} \succeq 0, \end{aligned}$$

yielding

$$P_{\tau|\tau} \succeq \hat{P}_{\tau|\tau},$$

hence the relation that is studied is preserved through the time update. For the measurement update equation:

$$\begin{aligned} &P_{\tau+1|\tau} - \hat{P}_{\tau+1|\tau} \\ &= (F P_{\tau|\tau} F^T + Q) - (F \hat{P}_{\tau|\tau} F^T + \hat{Q}) \\ &= F \underbrace{(P_{\tau|\tau} - \hat{P}_{\tau|\tau})}_{\succeq 0} F^T + \underbrace{(Q - \hat{Q})}_{\succeq 0} \succeq 0, \end{aligned}$$

that can be reformulated as

$$P_{\tau+1|\tau} \succeq \hat{P}_{\tau+1|\tau}.$$

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. (Kailath, 1980)

Now, (i)–(iii) together with the principle of induction proves that $P_{t|t-1} \succeq \hat{P}_{t|t-1}$ for $t > 0$. Especially this holds for large values of t .

With only minor changes the proof can be altered to show that for $Q \prec \hat{Q}$ or $R \prec \hat{R}$, the inequalities becomes strict. The gap to equality is in every recursion larger than a constant, independent of t , hence the inequalities hold also as $t \rightarrow \infty$. This argumentation concludes the proof.

An intuitive explanation to property (iv) is that the KF always improves when either of the covariances decreases, *i.e.*, $\hat{Q} \prec Q$ or $\hat{R} \prec R$, and that (13) is the same Riccati equation as in the KF with just a different interpretation of the included matrices. \square

For an interpretation of Theorem 1, assume that Q and R are covariances of process and measurement noises, respectively, and view $\gamma \leq 1$ as inverse RA, then the theorem provides rules to derive optimal prediction performance. For instance, (i) shows that optimal performance is linear in Ψ^{-1} if all noise share the same RA, and (iv) shows that any non-Gaussian noise improve the optimal performance. Even more general interpretations can be obtained by letting Q and R represent inverse IA.

4. SIMULATIONS

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

4.1 System

The following system will be used in the simulations,

$$x_{t+1} = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} x_t + \begin{pmatrix} T^2 \\ T \end{pmatrix} w_t, \quad \text{cov } w_t = Q \quad (15a)$$

$$y_t = (1 \ 0) x_t + e_t, \quad \text{cov } e_t = R \quad (15b)$$

with w_t and e_t mutually independent white noises with $Q = R = 1$ and $T = 1$. The system represents a second order random walk with $x = (x, v)^T$ where x and v represent position and velocity, respectively. The system can also be viewed as a double integrator. The best stationary estimate of x (the observed state) using a linear filter (KF) is $\text{var}(\hat{x}_{t+1|t}) = \bar{\kappa}_x(1, 1) = 3.0$.

Since both w_t and e_t are scalar it is possible to illustrate how the optimal performance varies as the IA of the noises changes. In Fig. 5 the prediction performance $\bar{\kappa}_x(\Psi_{w_t}^{-1}, \Psi_{e_t}^{-1})/\bar{\kappa}_x(1, 1)$ is presented as a function of the RA involved. Hence, Fig. 5 is in much similar to Fig. 1 and Fig. 3 in that it shows the optimal prediction performance expressed in parameters. Using the figure it is possible to see if it is worthwhile to try a nonlinear filter and try to reach the lower bound or not. For instance, in the lower left corner the CRLB is much better than what is achieved with a

BLUE since there are both the noises as informative they can be. Generally, if the CRLB is much lower than the BLUE performance a nonlinear filter should be considered. However, observe that it is impossible to conclude from these results how to obtain optimal performance, how difficult it is, or even if it is possible at all.

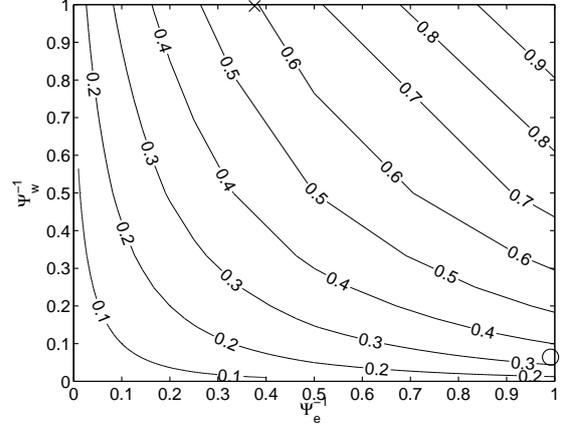


Fig. 5. Optimal filter performance, as a function of Ψ_{w_t} and Ψ_{e_t} , $\bar{\kappa}_x(\Psi_{w_t}^{-1}, \Psi_{e_t}^{-1})/\bar{\kappa}_x(1, 1)$. (\times denotes the noise in the first simulation and \circ the noise in the second simulation.)

4.2 Bi-Gaussian Measurement Noise

The first simulation features the non-Gaussian measurement noise

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

see Sec. 2.3. From Fig. 5, or by solving the appropriate Riccati equation (14), the CRLB for this system with this measurement noise can be found to be $\bar{\kappa}_x(1, 0.37) = 1.8$, *i.e.*, the optimal variance is 60% of what is obtainable with a BLUE, $\bar{\kappa}_x(1, 1) = 3.0$. Hence, this seems to be a candidate for a nonlinear filter.

The system was analyzed in a simulation study, where a KF and a PF (50 000 particles²) were applied. The *mean square error* (MSE) of these estimates were then computed for 1 000 *Monte Carlo* (MC) simulations. The MSE together with theoretical stationary limits are plotted in Fig. 7, which shows a significant improvement when using the PF (approximately 18% lower variance), but the CRLB is not reached. A reason is that the CRLB expression is asymptotic in the measurements, which leaves no guarantee that it could be reached in practice. More measurement information compared to process noise would probably improve the results, *i.e.*, more and better measurements. This could in practice be achieved with *e.g.*, more frequent measurements or a better sensor.

² The number of particles in the PF is large, intentionally throughout this paper, to get the most from the filter still without having to worry about numerical issues. Fewer particles could be sufficient but this has not been analyzed.

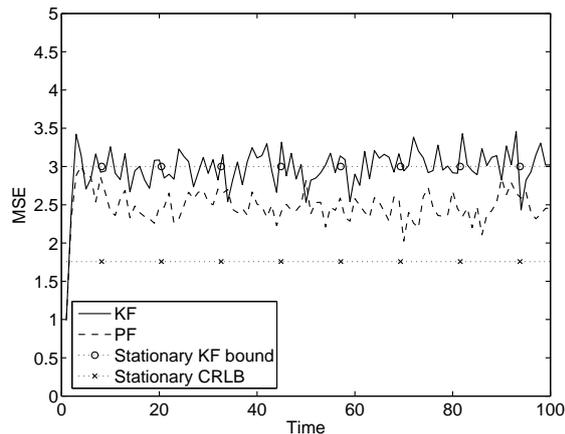


Fig. 6. MSE of 1000 MC simulations with KF and PF (50 000 particles) on the system with bi-Gaussian measurement noise. Theoretical limits are included for reference.

4.3 Tri-Gaussian Process Noise

This time, the measurements are kept Gaussian, whereas the system is driven by the tri-modal noise in Sec. 2.3,

$$w_t \sim 0.075\mathcal{N}(-2.5, 0.065) + 0.85\mathcal{N}(0, 0.065) + 0.075\mathcal{N}(+2.5, 0.065).$$

The Gaussian approximation is the same as before, with $\bar{\kappa}_x(1, 1) = 3.0$. However, the CRLB for the system is different, $\bar{\kappa}_x(0.065, 1) = 1.77$ (consult Fig. 5 or solve (14)).

Simulating this system and applying a KF and a PF (50 000 particles) yields for 1000 MC simulations the result in Fig. 7. Here the PF is not significantly better

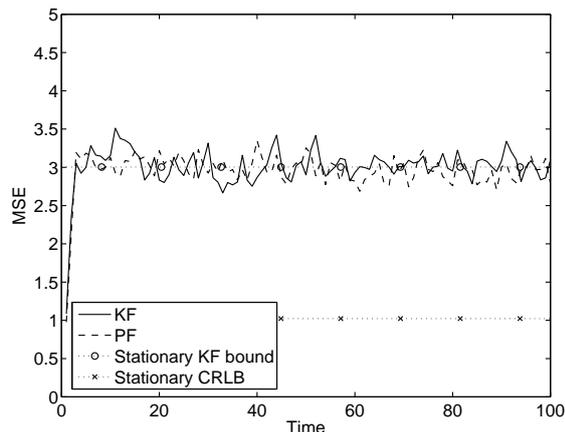


Fig. 7. MSE for x of 1000 MC simulations with KF and PF (50 000 particles) on the system with tri-Gaussian process noise. Theoretical limits are included for reference.

than the KF (mean over time shows a 3% improvement for the PF). The same argumentation as for the previous simulation apply.

This case is also complicated by the fact that the non-Gaussian noise is only measured indirectly. As a result

it is harder to extract all available information, especially for prediction since there is no measurement available to give information about the latest process noise affecting the system.

5. CONCLUSIONS

In this paper, starting from general posterior *Cramér-Rao lower bound* (CRLB) expressions, an expression for the CRLB in linear systems is derived in terms of *intrinsic accuracy* (IA) or *relative accuracy* (RA) and covariance of the included noises. This results in a method to, given a system and its IA or RA, 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 nonlinear filtering or decide when no further tuning is needed, all this without resorting to time consuming MC simulations. Simulations, presented to support the theoretical results, indicate improved performance using a particle filter on linear systems with non-Gaussian noise, but do also point out that it is sometimes difficult to reach the CRLB.

ACKNOWLEDGEMENTS

This work is supported by VINNOVA's Center of Excellence ISIS (Information Systems for Industrial Control and Supervision) at Linköpings universitet, Linköping, Sweden.

REFERENCES

- Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering*. Prentice-Hall, Inc. Englewood Cliffs, NJ.
- Bergman, N. (1999). Recursive Bayesian Estimation: Navigation and Tracking Applications. Dissertation no 579. Linköpings universitet.
- 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 Syst. Mag.* **19**(3), 33–40.
- Blom, H. A. P. and Y. Bar-Shalom (1988). The interacting multiple model algorithm for systems with Markovian switching coefficients. *IEEE Trans. Automat. Contr.* **33**(8), 780–783.
- Cox, D. R. and D. V. Hinkley (1974). *Theoretical Statistics*. Chapman and Hall. New York.
- Dahlgren, C. (1998). Nonlinear black box modelling of JAS 39 Gripen's radar altimeter. Master's thesis no LiTH-ISY-EX-1958. Dept. E.E., 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-Gaussian Bayesian state estimation. *IEE Proceedings-F* **140**(2), 107–113.
- Gustafsson, F. (2000). *Adaptive Filtering and Change Detection*. John Wiley & Sons, Ltd. Chichester, West Sussex, England.
- Kailath, T. (1980). *Linear Systems*. Prentice-Hall, Inc.
- 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.
- Kay, S. M. (1998). *Fundamentals of Statistical Signal Processing: Detection Theory*. Vol. 2. Prentice-Hall, Inc.
- Kay, S. M. and D. Sengupta (1987). Optimal detection in colored non-Gaussian noise with unknown parameter. In: *IEEE ICASSP '87*. Vol. 12. pp. 1087–1089.
- Robert, C. P. and G. Casella (1999). *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer-Verlag. New York.
- Sorenson, H. W. and D. L. Alspach (1971). Recursive Bayesian estimation using Gaussian sums. *Automatica* **7**(4), 465–479.
- Tichavský, P., C. H. Muravchik and A. Nehorai (1998). Posterior Cramér-Rao discrete-time non-linear filtering. *IEEE Trans. Signal Processing* **46**(5), 1386–1396.