Particle filtering with dependent noise processes

Saikat Saha, *Member, IEEE,* and Fredrik Gustafsson, *Fellow, IEEE*

***Abstract*—Modeling physical systems often leads to discrete time state space models with dependent process and measurement noises. For linear Gaussian models, the Kalman filter handles this case, as is well described in literature. However, for nonlinear or non-Gaussian models, the particle filter as described in literature provides a general solution only for the case of independent noise. Here, we present an extended theory of the particle filter for dependent noises with the following key contributions: (i) The optimal proposal distribution is derived. (ii) The special case of Gaussian noise in nonlinear models is treated in detail, leading to a concrete algorithm that is as easy to implement as the corresponding Kalman filter. (iii) The marginalized (Rao-Blackwellized) particle filter, handling linear Gaussian substructures in the model in an efficient way, is extended to dependent noise. Finally, (iv) the parameters of a joint Gaussian distribution of the noise processes are estimated jointly with the state in a recursive way.**

***Index Terms*—Bayesian methods, recursive estimation, particle filters, dependent noise, Rao-Blackwellized particle filter**

I. Introduction

The particle filter (PF) provides an arbitrary good numerical approximation to the online nonlinear filtering problem. More specifically, the PF approximates the posterior distribution $p(x_k|Y_k)$ of the latent state x_k at time k, given the observations $Y_k = \{y_1, y_2, \ldots, y_k\}$, based on the discrete time state space model

$$x_{k+1} = f(x_k, v_k), \tag{1a}$$

$$y_k = h(x_k, e_k). \tag{1b}$$

Here, the model is specified by the state dynamics $f(\cdot)$, the observation dynamics $h(\cdot)$ and the initial state x_0. Note that the latent state is usually assumed to be Markovian, i.e., the conditional density of x_{k+1} given the past state $x_{0:k} \equiv (x_0, x_1, \ldots, x_k)$, depends only on x_k. The process noise v_k and measurement noise e_k, $k = 1, 2, \cdots$ are both assumed to be independent over time (white noise). The processes v_k and e_k are independent, except for either v_k, e_k (type I) or v_{k-1}, e_k (type II), which in this contribution, are assumed to be dependent. In this model, we assume that the probability density functions for x_0, v_k and e_k are known.

The theory of the PF as described in survey and tutorial articles [6]– [11]) treats the process noise and measurement noise as independent processes. This is in contrast to the Kalman filter, where the case of correlated noise is a standard modification, see for instance [4] where type I dependence is assumed throughout the whole book. It is the purpose of this contribution to fill this gap in the PF theory.

The case of dependent noise might be more common in practice than is acknowledged in the literature. More specifically, it occurs whenever a (linear or nonlinear) filter is based on a discrete time model that is derived from a continuous time model (the only exception is when a piecewise constant noise process is assumed). For instance, in a typical target tracking application, a radar is used to track an object. Even if the measurement noise in the radar is completely independent of the motion of the object, the sampling process of the motion dynamics gives an extra noise contribution to the sensor model which is dependent with the object's motion. We will explain this phenomenon in more detail in the last section. Also downsampling the dynamics in filtering problem introduces such noise dependency (see [23] for details). This dependency also arises in modeling many practical applications of interests, see e.g., [21].

In this article[1], we propose a new class of particle filter algorithms which can take care of the noise dependency. The organization of this article is as follows. In section II, we start with outlining the somewhat different structures of the dependency and treat the optimal filtering for these different structures in parallel. We then derive the optimal proposal densities to be used in combination with the particle filters in section III. The two most common approximations (prior and likelihood proposals) are also outlined. The optimal proposals are then specialized to the case of Gaussian dependent noise processes. Moreover, with affine sensor model, the optimal proposals for Gaussian dependent noise processes are obtained in closed form. We next develop the marginalized particle filter framework with dependent noise processes in section IV. In section V, we address a recursive framework of estimating the unknown noise statistics of the dependent Gaussian noises, driving a general state space model. Finally, in section VI, as illustration, we show how sampling continuous time models can lead to the noise dependency.

II. Optimal filtering with dependent noise processes

For simplicity, consider the dynamic system (1) with additive measurement noise

$$x_{k+1} = f(x_k, v_k) \tag{2a}$$

$$y_l = h(x_l) + e_l \tag{2b}$$

where x_k is the latent state at time step k while y_l is the observation at time step l. v_k and e_l are the respective process and measurement noises with their joint density assumed to

The authors are with the Department of Electrical Engineering, Division of Automatic Control, Linköping University, Linköping, SE 581-83, Sweden, e-mail: {saha, fredrik}@isy.liu.se

[1]A preliminary version was presented in FUSION 2010 [21].

be known. Furthermore, as in (1), the noise sequences are individually assumed to be independent. The joint posterior $p(X_k|Y_k)$ can be recursively obtained (up to a proportionality constant) as

$$p(X_k|Y_k) \propto p(y_k|X_k, Y_{k-1})p(x_k|X_{k-1}, Y_{k-1}) \times \times p(X_{k-1}|Y_{k-1}). \quad (3)$$

For the standard Markovian model with independent process and measurement noises (i.e. $p(v_i, e_j) = p(v_i)p(e_j)$), we have

$$p(x_k|X_{k-1}, Y_{k-1}) = p(x_k|x_{k-1}) \quad (4a)$$
$$p(y_k|X_k, Y_{k-1}) = p(y_k|x_k). \quad (4b)$$

Here the predictive density $p(x_k|x_{k-1})$ and the likelihood density $p(y_k|x_k)$ can be characterized in terms of the noise densities, $p(v_{k-1})$ and $p(e_k)$ respectively. This explains why the standard model in particle filtering literature (see e.g. [6]–[11]) is based on the prediction model and the likelihood function, respectively.

Now we consider the more general case where the process noise and the measurement noise are dependent. To further explain this dependency, consider the graphical representation of the dynamic system given by (2a)–(2b) in Figure 1. Here

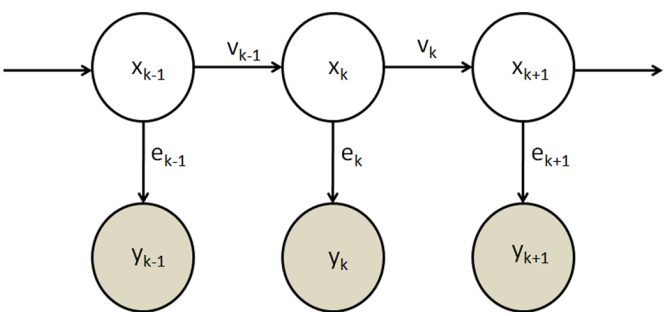

Figure 1. A graphical representation of the state space model (2a)–(2b)

the process noise v_k is driving the latent state x_k to the next time step x_{k+1}. The measurements corresponding to the adjacent states x_k and x_{k+1} are y_k and y_{k+1}, obtained through the measurement noises e_k and e_{k+1} respectively. According to the time occurrence of the dependency (adjacent in time), we treat here two dependency structures where the process noise v_k either depends on (1) the measurement noise e_k (same time step) or (2) e_{k+1} (one step apart). As we will explain here, the two cases are not quite the same. However, to proceed with both the cases, the main idea is a suitable decomposition of the joint density of the dependent noises $p(v_i, e_j),\ j = i\ or\ (i+1)$, into factors of appropriate conditionals.

A. Type I dependency

We first consider the dependency structure where v_{k-1} is dependent to $e_{k-1},\ k = 1, 2, \cdots$, as shown in Figure 2. Here the sequence of the noise vector $(v_{k-1}, e_{k-1})^T$ over different k is assumed to be independent. We call this Type I dependency.

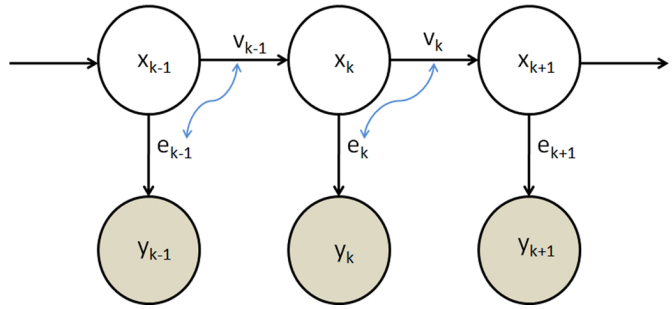

Figure 2. Type I dependency between process and measurement noise processes

This is pretty common in the engineering literature, see e.g., [5]. For this dependency, we have

$$p(x_k|X_{k-1}, Y_{k-1}) = p(x_k|x_{k-1}, y_{k-1}) \quad (5a)$$
$$p(y_k|X_k, Y_{k-1}) = p(y_k|x_k). \quad (5b)$$

Now, we use the decomposition

$$p(v_{k-1}, e_{k-1}) = p(v_{k-1}|e_{k-1})p(e_{k-1}). \quad (6)$$

Since knowing x_{k-1} and y_{k-1} together would provide the complete information on e_{k-1}, $p(x_k|x_{k-1}, y_{k-1})$ and $p(y_k|x_k)$ can be characterized in terms of $p(v_{k-1}|e_{k-1})$ and $p(e_k)$ respectively. Clearly, for this case, the hidden states and the observations form jointly a Markov chain [14]. This joint Markov chain model was considered in [13] for the particle filter with correlated Gaussian noises.

B. Type II dependency

We next consider the dependency structure where v_k and e_{k+1} are dependent to each other for $k = 0, 1, \cdots$, while the sequence of the noise vector $(v_k, e_{k+1})^T$ over different k is assumed to be independent. We call this Type II dependency. This is shown in 3. This dependency structure has been

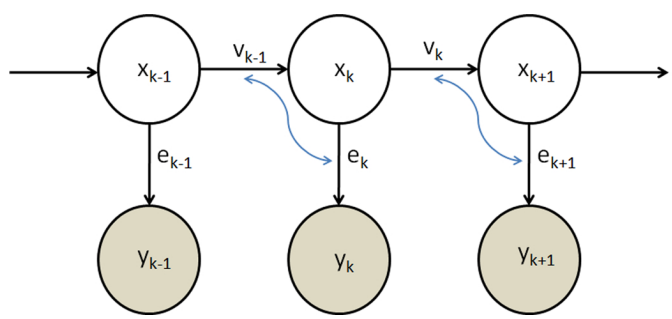

Figure 3. Type II dependency between process and measurement noise processes

considered, e.g., in [12] for the treatment of Kalman filter. It follows that

$$p(x_k|X_{k-1}, Y_{k-1}) = p(x_k|x_{k-1}) \quad (7a)$$
$$p(y_k|X_k, Y_{k-1}) = p(y_k|x_k, x_{k-1}). \quad (7b)$$

For this case, we use the decomposition

$$p(v_{k-1}, e_k) = p(e_k|v_{k-1})p(v_{k-1}). \tag{8}$$

Like the previous case, knowing x_k and x_{k-1} together would provide the complete information on v_{k-1}. As a result, $p(x_k|x_{k-1})$ and $p(y_k|x_k, x_{k-1})$ can be characterized in terms of $p(v_{k-1})$ and $p(e_k|v_{k-1})$ respectively. Note that, this dependency treatment here does not require any so called joint Markovian assumption.

III. Optimal Proposal for Dependent Noises

In particle filtering, the posterior is approximated in the form of (weighted) random samples, also known as particles. These particles are generated from a different distribution, called the importance distribution (or the proposal), which is ideally supposed to be as close as possible to the (unknown) posterior. Without going into further details of PF, we here provide a generic PF algorithm that will be generalized to dependent noise in our subsequent developments.

Algorithm 1: **[Particle Filter]**
Recursively over time $k = 0, 1, 2, \ldots$
For $i = 1, \ldots, N$, where N is the total number of particles,

- sample $x_k^{(i)} \sim q\left(x_k|X_{k-1}^{(i)}, Y_k\right)$ and set $X_k^{(i)} \triangleq \left(X_{k-1}^{(i)}, x_k^{(i)}\right)$
- evaluate the corresponding importance weights $w_k^{(i)}$ according to

$$w_k^{(i)} \propto \widetilde{w}_{k-1}^{(i)} \frac{p(y_k|X_k^{(i)}, Y_{k-1})p(x_k^{(i)}|X_{k-1}^{(i)}, Y_{k-1})}{q(x_k^{(i)}|X_{k-1}^{(i)}, Y_k)}.$$

- normalize the importance weights $\widetilde{w}_k^{(i)} = \frac{w_k^{(i)}}{\sum_{i=1}^N w_k^{(i)}}$.
- resample the trajectories $\{x_k^{(i)}\}_{i=1}^N$ with probabilities $\{\tilde{w}_k^{(i)}\}_{i=1}^N$ and set $\tilde{w}_k^{(i)} = 1/N$. Reampling can be done at every time (SIR-PF) or when sample depletion is indicated (SIS-PF).

The performance of PF depends critically on the selection of the proposal $q(x_k|\cdot)$. In this section, we derive the optimal proposal when the noises are dependent. Here the proposal is optimal in the sense of minimum conditional variance of the importance weights [3].

A. Optimal Proposal for Type I dependency

In engineering literature, the following state space model, and variants of it, is commonly used,

$$x_{k+1} = f_k(x_k) + G_k v_k, \tag{9a}$$
$$y_k = h_k(x_k) + e_k, \tag{9b}$$

where the process noise sequence v_k and the measurement noise sequence e_k, $k = 1, 2, \cdots$, are individually assumed to be independent, while v_k and e_k are dependent. This corresponds to Type I dependency as defined in II-A. This case of dependent noise can now be phrased as $p(y_k, x_{k+1}|x_k)$, where

$$p(y_k, x_{k+1}|x_k) \neq p(x_{k+1}|x_k)p(y_k|x_k), \tag{10}$$

that is, y_k and x_{k+1} are not independent, given x_k.

The proposal distribution has the functional form $q(x_k|X_{k-1}, Y_k)$. In the standard PF, the Markovian property and the independence of Y_{k-1} and x_k are used to get [5]

$$q(x_k|X_{k-1}, Y_k) = q(x_k|x_{k-1}, y_k). \tag{11a}$$

For the case (10), there is a dependency between y_{k-1} and x_k, so we get

$$q(x_k|X_{k-1}, Y_k) = q(x_k|x_{k-1}, y_k, y_{k-1}). \tag{11b}$$

Based on this, we derive the following theorem for the optimal proposal function.

Theorem 1: **[Optimal proposal for Type I dependent noise]** Here the optimal proposal function is given by

$$q(x_k|x_{k-1}, y_k, y_{k-1}) = \frac{p(x_k|y_{k-1}, x_{k-1})p(y_k|x_k)}{p(y_k|y_{k-1}, x_{k-1})}. \tag{12}$$

Proof: The optimal proposal is given by the posterior distribution of the same functional form, which can be rewritten using Bayes' law as

$$q(x_k|x_{k-1}, y_k, y_{k-1}) \tag{13a}$$
$$= p(x_k|x_{k-1}, y_k, y_{k-1}) \tag{13b}$$
$$= \frac{p(x_k, y_k|y_{k-1}, x_{k-1})}{p(y_k|y_{k-1}, x_{k-1})} \tag{13c}$$
$$= \frac{p(y_k|x_k)p(x_k|y_{k-1}, x_{k-1})}{p(y_k|y_{k-1}, x_{k-1})}. \tag{13d}$$

This concludes the proof. □.

The optimal proposal as described in (12) has been used as a special case for estimating the stochastic volatility in [15]. This optimal proposal should be compared to the standard one given by

$$q(x_k|x_{k-1}, y_k) = \frac{p(x_k|x_{k-1})p(y_k|x_k)}{p(y_k|x_{k-1})}. \tag{14}$$

One can, just as for the standard PF, define two extreme cases of sub-optimal proposal distributions

$$\text{Prior}: \quad q(x_k|x_{k-1}, y_{k-1}) \propto p(x_k|y_{k-1}, x_{k-1}) \tag{15a}$$

$$\text{Likelihood}: \quad q(x_k|y_k) \propto p(y_k|x_k). \tag{15b}$$

The first proposal corresponds to the model (10), while the second proposal is obtained directly from the observation model (9b).

B. Gaussian Noise Case for Type I dependency

To get instructive and explicit expressions, the Gaussian case,

$$x_0 \sim \mathcal{N}(\hat{x}_{1|0}, P_{1|0}), \tag{16a}$$

$$\begin{pmatrix} v_k \\ e_k \end{pmatrix} \in \mathcal{N}\left(0, \begin{bmatrix} Q_k & S_k \\ S_k^T & R_k \end{bmatrix}\right), \tag{16b}$$

is studied in detail. The standard PF applies for the case $S_k = 0$ only. This Gaussian noise model can also be written as

$$p\left(\begin{pmatrix} x_{k+1} \\ y_k \end{pmatrix} | x_k\right) = \mathcal{N}\left(\begin{pmatrix} f(x_k) \\ h(x_k) \end{pmatrix}, \begin{bmatrix} G_k Q_k G_k^T & G_k S_k \\ S_k^T G_k^T & R_k \end{bmatrix}\right). \tag{17}$$

Note that this noise covariance is the standard representation in the classic text book [4] on Kalman filters (KF). The main result is given in Theorem 2.

Theorem 2: **[Optimal proposal for Type I Gaussian dependent noise]** For the model specified by (17), the optimal proposal function is given by

$$\begin{aligned} &q(x_k|x_{k-1}, y_k, y_{k-1}) \propto \\ &\mathcal{N}\Big(f(x_{k-1}) + G_{k-1}S_{k-1}R_{k-1}^{-1}\big(y_{k-1} - h(x_{k-1})\big), \\ &\quad G_{k-1}\big(Q_{k-1} - S_{k-1}R_{k-1}^{-1}S_{k-1}^T\big)G_{k-1}^T\Big)\times \\ &\times \mathcal{N}(h(x_k), R_k). \end{aligned} \tag{18}$$

Proof: The result follows from (12) by studying the two factors in the numerator. The second factor $p(y_k|x_k)$ is obtained from the observation model (9b). The first factor $p(x_k|x_{k-1}, y_{k-1})$ follows by using Lemma 1 below. □

Lemma 1: **[Conditional Gaussian Distributions]** Suppose the vectors X and Y are jointly Gaussian distributed as

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{bmatrix} P_{xx} & P_{xy} \\ P_{xy}^T & P_{yy} \end{bmatrix}\right) = \mathcal{N}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, P\right). \tag{19}$$

Then, the conditional distribution for X, given the observed $Y = y$, is Gaussian distributed:

$$(X|Y = y) \sim \mathcal{N}(\mu_x + P_{xy}P_{yy}^{-1}(y - \mu_y), P_{xx} - P_{xy}P_{yy}^{-1}P_{yx}). \tag{20}$$

□

In the above Lemma, let $X = x_k|x_{k-1}$ and $Y = y_{k-1}|x_{k-1}$ and use the joint distribution of X and Y from (17), time-shifted one step, then

$$\begin{aligned} &p(x_k|x_{k-1}, y_{k-1}) = \\ &\mathcal{N}\Big(f(x_{k-1}) + G_{k-1}S_{k-1}R_{k-1}^{-1}\big(y_{k-1} - h(x_{k-1})\big), \\ &\quad G_{k-1}\big(Q_{k-1} - S_{k-1}R_{k-1}^{-1}S_{k-1}^T\big)G_{k-1}^T\Big). \end{aligned} \tag{21}$$

The prior proposal in (15a) becomes more explicit as summarized in Corollary 1.

Corollary 1: **[Prior proposal for Type I Gaussian dependent noise]** For the model specified by (17), the prior proposal function is given by

$$\begin{aligned} &q(x_k|x_{k-1}, y_{k-1}) = \\ &\mathcal{N}\Big(f(x_{k-1}) + G_{k-1}S_{k-1}R_{k-1}^{-1}\big(y_{k-1} - h(x_{k-1})\big), \\ &\quad G_{k-1}\big(Q_{k-1} - S_{k-1}R_{k-1}^{-1}S_{k-1}^T\big)G_{k-1}^T\Big). \end{aligned} \tag{22}$$

Proof: In this case, the proposal is $p(x_k|x_{k-1}, y_{k-1})$ which is directly given by Lemma 1, as shown above. □

It is thus straightforward to generate samples from this proposal using the Gaussian random number generator. The standard SIR PF is obtained by letting $S_{k-1} = 0$ in (22).

The optimal proposal in (18) cannot be analytically obtained in general. However, one important exception is for an affine sensor model, for which the optimal proposal is Gaussian. This is shown below:

Corollary 2: **[Optimal Gaussian proposal for affine sensor model with Type I dependency]** When the sensor model in (9b) is affine, the optimal proposal in (18) is Gaussian, i.e., $q(x_k|y_k, x_{k-1}, y_{k-1}) = \mathcal{N}(\bar{\mu}_k, \bar{\Sigma}_k)$ (refer to (27)).

Proof: From Corollary 1, we have $p(x_k|x_{k-1}, y_{k-1}) = N(\mu_k^*, \Sigma_k^*)$ where

$$\mu_k^* := f(x_{k-1}) + G_{k-1}S_{k-1}R_{k-1}^{-1}\big(y_{k-1} - h(x_{k-1})\big), \tag{23a}$$

$$\Sigma_k^* := G_{k-1}\big(Q_{k-1} - S_{k-1}R_{k-1}^{-1}S_{k-1}^T\big)G_{k-1}^T. \tag{23b}$$

When the sensor model is affine (i.e. $h(x_k) = A_k + C_k x_k$), we can write

$$\begin{pmatrix} x_k \\ y_k \end{pmatrix} = \begin{bmatrix} I & 0 \\ C_k & I \end{bmatrix}\begin{pmatrix} x_k \\ e_k \end{pmatrix} + \begin{pmatrix} 0 \\ A_k \end{pmatrix}. \tag{24}$$

Conditioned on (x_{k-1}, y_{k-1}), we have

$$\left(\begin{pmatrix} x_k \\ e_k \end{pmatrix} | \begin{pmatrix} x_{k-1} \\ y_{k-1} \end{pmatrix}\right) \in \mathcal{N}\left(\begin{pmatrix} \mu_k^* \\ 0 \end{pmatrix}, \begin{bmatrix} \Sigma_k^* & 0 \\ 0 & R_k \end{bmatrix}\right), \tag{25}$$

and by (24), we obtain

$$\begin{aligned} &p\left(\begin{pmatrix} x_k \\ y_k \end{pmatrix} | \begin{pmatrix} x_{k-1} \\ y_{k-1} \end{pmatrix}\right) = \\ &\mathcal{N}\left(\begin{pmatrix} \mu_k^* \\ C_k\mu_k^* + A_k \end{pmatrix}, \begin{bmatrix} \Sigma_k^* & \Sigma_k^* C_k^T \\ C_k\Sigma_k^* & C_k\Sigma_k^* C_k^T + R_k \end{bmatrix}\right) \end{aligned} \tag{26}$$

Now using Lemma 1 in (26), we can show that $p(x_k|y_k, x_{k-1}, y_{k-1}) = \mathcal{N}(\bar{\mu}_k, \bar{\Sigma}_k)$, where

$$\begin{aligned} \bar{\mu}_k &= \mu_k^* + \Sigma_k^* C_k^T (C_k\Sigma_k^* C_k^T + R_k)^{-1}\times \\ &\quad \times \{y_k - C_k\mu_k^* - A_k\} \end{aligned} \tag{27a}$$

$$\begin{aligned} \bar{\Sigma}_k &= \Sigma_k^* - \Sigma_k^* C_k^T \times \\ &\quad \times (C_k\Sigma_k^* C_k^T + R_k)^{-1} C_k \Sigma_k^*. \end{aligned} \tag{27b}$$

Now defining, $K_k \triangleq \Sigma_k^* C_k^T (C_k\Sigma_k^* C_k^T + R_k)^{-1}$, we can rewrite (27a) – (27b) as

$$\bar{\mu}_k = K_k(y_k - A_k) + (I - K_k C_k)\mu_k^* \tag{27c}$$

$$\bar{\Sigma}_k = (I - K_k C_k)\Sigma_k^*. \tag{27d}$$

□

C. Optimal Proposal for Type II dependency

We consider here the following model:

$$x_k = f_k(x_{k-1}) + G_k v_{k-1}, \tag{28a}$$

$$y_k = h_k(x_k) + e_k, \tag{28b}$$

where the process noise sequence v_k and the measurement noise sequence e_k, $k = 1, 2, \cdots$, are individually assumed to be independent, while v_{k-1} and e_k are dependent for different

$k = 1, 2, \cdots$. This corresponds to Type II dependency as defined in II-B. This noise dependency implies that

$$p(x_k, y_k|x_{k-1}) = p(x_k|x_{k-1})p(y_k|x_k, x_{k-1}). \tag{29}$$

Theorem 3: **[Optimal proposal for Type II dependent noise]** For the model specified by (28a)–(28b), the optimal proposal function is given by

$$q(x_k|X_{k-1}, Y_k) = p(x_k|x_{k-1}, y_k) \tag{30}$$
$$\propto p(x_k|x_{k-1})p(y_k|x_k, x_{k-1}). \tag{31}$$

Proof:

$$\begin{aligned} p(x_k|x_{k-1}, y_k) &= \frac{p(y_k, x_k|x_{k-1})}{p(y_k|x_{k-1})} \\ &= \frac{p(y_k|x_k, x_{k-1})p(x_k|x_{k-1})}{p(y_k|x_{k-1})} \\ &\propto p(x_k|x_{k-1})p(y_k|x_k, x_{k-1}) \end{aligned}$$

□.

Just like the standard PF, one can define the following two sub-optimal proposal distributions

$$\text{Prior}: \quad q(x_k|x_{k-1}) \propto p(x_k|x_{k-1}) \tag{32a}$$
$$\text{Likelihood}: \quad q(x_k|y_k, x_{k-1}) \propto p(y_k|x_k, x_{k-1}). \tag{32b}$$

D. Gaussian Noise Case for Type II dependency

When the noises v_{k-1} and e_k are jointly Gaussian as

$$\begin{pmatrix} v_{k-1} \\ e_k \end{pmatrix} \in \mathcal{N}\left(0, \begin{bmatrix} Q_{k-1} & S_k \\ S_k^T & R_k \end{bmatrix}\right), \tag{33}$$

the equivalent probabilistic description of the state space model ((28a)–(28b)) can be given by

$$p\left(\begin{pmatrix} x_k \\ y_k \end{pmatrix}|x_{k-1}\right) = \mathcal{N}\left(\begin{pmatrix} f_k(x_{k-1}) \\ h_k(x_k) \end{pmatrix}, \begin{bmatrix} G_k Q_{k-1} G_k^T & G_k S_k \\ S_k^T G_k^T & R_k \end{bmatrix}\right). \tag{34}$$

Theorem 4: **[Optimal proposal for Type II Gaussian dependent noise]** For the model specified by (34), the optimal proposal function is given by

$$\begin{aligned} &q(x_k|x_{k-1}, y_k) \propto \mathcal{N}\Big(f(x_{k-1}), G_k Q_{k-1} G_k^T\Big) \times \\ &\times \mathcal{N}\Big(h(x_k) + S_k^T Q_{k-1}^{-1} G_k^{\dagger}(x_k - f(x_{k-1})), \\ &R_k - S_k^T Q_{k-1}^{-1} S_k\Big). \end{aligned} \tag{35}$$

Proof: The result follows from (31) by studying the two factors. The first factor is obtained directly from the process model (28a). The second factor $(y_k|x_{k-1}, x_k)$ follows by using Lemma 1 in (34). □

Corollary 3: **[Prior proposal for Type II Gaussian dependent noise]** For the model specified by (34), the prior proposal function is given by

$$q(x_k|x_{k-1}, y_k) \propto \mathcal{N}\Big(f(x_{k-1}), G_k Q_{k-1} G_k^T\Big) \tag{36}$$

Proof: In this case, the proposal is $p(x_k|x_{k-1})$ which is directly obtained from the process model (28a). □

Corollary 4: **[Optimal Gaussian proposal for affine sensor model with Type II dependency]** When the sensor model in (28b) is affine, the optimal proposal in (35) is Gaussian, given by $q(x_k|y_k, x_{k-1}) = \mathcal{N}(\widetilde{\mu}_k, \widetilde{\Sigma}_k)$ (refer to (37)).

Proof: When the sensor model is affine (i.e. $h(x_k) = A_k + C_k x_k$), the state space model ((28a)–(28b)) can be written as

$$\begin{pmatrix} x_k \\ y_k \end{pmatrix} = \begin{pmatrix} f(x_{k-1}) \\ A_k + C_k f(x_{k-1}) \end{pmatrix} + \begin{bmatrix} G_k & 0 \\ C_k G_K & I \end{bmatrix} \begin{pmatrix} v_{k-1} \\ e_k \end{pmatrix}. \tag{37a}$$

Now using (33) one can show

$$p\left(\begin{pmatrix} x_k \\ y_k \end{pmatrix}|x_{k-1}\right) \sim \mathcal{N}\left(\bar{\mu}_k, \bar{\Sigma}_k\right) \tag{37b}$$

where

$$\bar{\mu}_k = \begin{pmatrix} f(x_{k-1}) \\ A_k + C_k f(x_{k-1}) \end{pmatrix} \tag{37c}$$

and

$$\bar{\Sigma}_k = \begin{bmatrix} G_k Q_{k-1} G_k^T & G_k Q_{k-1} G_k^T C_k^T + G_k S_k \\ C_k G_k Q_{k-1}^T G_k^T + S_k^T G_k^T & C_k G_k (Q_{k-1} G_k^T C_k^T + S_k) + \\ & + S_k^T G_k^T C_k^T + R_k \end{bmatrix} \tag{37d}$$

$$:= \begin{bmatrix} \bar{P}_{xx,k} & \bar{P}_{xy,k} \\ \bar{P}_{xy,k}^T & \bar{P}_{yy,k} \end{bmatrix} \tag{37e}$$

Using Lemma 1 in (37b), we obtain

$$q(x_k|y_k, x_{k-1}) = p(x_k|y_k, x_{k-1}) = \mathcal{N}(\widetilde{\mu}_k, \widetilde{\Sigma}_k) \tag{37f}$$

where

$$\widetilde{\mu}_k = f(x_{k-1}) + \bar{P}_{xy,k} \bar{P}_{yy,k}^{-1}(y_k - A_k - C_k f(x_{k-1})) \tag{37g}$$
$$\widetilde{\Sigma}_k = \bar{P}_{xx,k} - \bar{P}_{xy,k} \bar{P}_{yy,k}^{-1} \bar{P}_{xy,k}^T. \tag{37h}$$

□

To conclude this section, a summary of the different proposals for dependent noise cases is presented in Table I.

IV. MPF for mixed linear /nonlinear state space models with dependent Gaussian noises

The idea behind the marginalized particle filter (MPF) is as follows. If there is an analytically tractable substructure within the general state space model, the state estimation problem can be divided into sub-parts: given any observation, the non analytical part is estimated using the PF and the tractable substructure can be estimated analytically conditioned on the PF output. This method is also referred to as Rao-Blackwellized particle filter. There are several advantages using MPF: besides obtaining an improved estimate from the Rao-Blackwellization, this helps us to keep the state dimension small enough for the PF to be feasible.

A widely used MPF for state estimation is the one containing a conditionally linear-Gaussian substructure, for which optimal estimate can be obtained analytically using the Kalman filter (see e.g., [17], [3], [19]). However, in all such available algorithms for MPF, the measurement noise vector is assumed to be independent of the process noise vector. Here we relax this assumption and extend the available results to the case

Table I
SUMMARY OF THE DIFFERENT PROPOSALS WITH DEPENDENT NOISE CASES

	Dependent noise: Type I	Dependent noise: Type II
Optimal proposal	Theorem 1	Theorem 3
Opt. proposal: Gaussian noise case	Theorem 2	Theorem 4
Prior proposal: Gaussian noise case	Corollary 1	Corollary 3
Opt. proposal: Gaussian noise case with affine sensor model	Corollary 2	Corollary 4

of dependent noise processes[2]. For the derivation, we mainly follow [19].

A. Mixed linear/nonlinear state space model

A rather general model containing a linear-Gaussian substructure is given as [5]

$$x_{k+1}^l = f_k^l(x_k^p) + A_k^l(x_k^p)x_k^l + G_k^l(x_k^p)w_k^l \tag{38a}$$
$$x_{k+1}^p = f_k^p(x_k^p) + A_k^p(x_k^p)x_k^l + G_k^p(x_k^p)w_k^p \tag{38b}$$
$$y_k = h_k(x_k^p) + C_k(x_k^p)x_k^l + e_k, \qquad k = 1, 2, \ldots \tag{38c}$$

where x_k^l denotes the state variable with conditionally linear dynamics, x_k^p denotes the nonlinear state variable and y_k is the measurement at discrete time step k. w_k^l, w_k^p are the process noises driving x_{k+1}^l and x_{k+1}^p respectively and e_k is the measurement noise. The noise vector at time step k, is assumed to be jointly Gaussian with zero mean as

$$\begin{pmatrix} w_k^l \\ w_k^p \\ e_k \end{pmatrix} \sim \mathcal{N}(0, \Sigma_k) \tag{39}$$

with covariance marix

$$\Sigma_k = \begin{bmatrix} \Sigma_k^{ll} & \Sigma_k^{lp} & \Sigma_k^{ly} \\ (\Sigma_k^{lp})^T & \Sigma_k^{pp} & \Sigma_k^{py} \\ (\Sigma_k^{ly})^T & (\Sigma_k^{py})^T & \Sigma_k^{yy} \end{bmatrix} \tag{40}$$

The sequence of this noise vector over different k is assumed to be independent. Furthermore, x_0^l is assumed to be independent of the noises and distributed according to a Gaussian as,

$$x_0^l \sim \mathcal{N}(\bar{x}_0, \bar{P}_0). \tag{41}$$

The density of x_0^p is arbitrary, but it is assumed to be known. We further assume that the dynamic model follows favorable mixing property as in [18]. For notational brevity, the dependence on x_k^p in equations (38a)–(38c) is suppressed onwards.

B. Gram-Schmidt orthogonalization for dependent noise processes

We define here two new Gaussian noise processes, $\bar{w}_k^p$ and $\bar{w}_k^l$, which are independent of each other and also individually independent of e_k, using the standard Gram-Schmidt procedure ([1], [4]) as follows:
Define

$$\bar{w}_k^p \triangleq w_k^p - \Sigma_k^{py}(\Sigma_k^{yy})^{-1}e_k \tag{42a}$$

such that

$$\bar{w}_k^p \sim \mathcal{N}(0, \Lambda_k^{\bar{p}}) \tag{42b}$$

with

$$\Lambda_k^{\bar{p}} = \mathrm{Cov}(\bar{w}_k^p) = \Sigma_k^{pp} - \Sigma_k^{py}(\Sigma_k^{yy})^{-1}(\Sigma_k^{py})^T \tag{42c}$$

Now define

$$\Lambda_k^{l\bar{p}} = E(w_k^l \bar{w}_k^p) = \Sigma_k^{lp} - \Sigma_k^{py}(\Sigma_k^{yy})^{-1}(\Sigma_k^{ly}). \tag{43}$$

Then

$$\bar{w}_k^l \triangleq w_k^l - \Sigma_k^{ly}(\Sigma_k^{yy})^{-1}e_k - \Lambda_k^{l\bar{p}}(\Lambda_k^{\bar{p}})^{-1}\bar{w}_k^p, \tag{44a}$$

leading to

$$\bar{w}_k^l \sim \mathcal{N}(0, \Lambda_k^{\bar{l}}) \tag{44b}$$

with

$$\begin{aligned} \Lambda_k^{\bar{l}} &= \mathrm{Cov}(\bar{w}_k^l) = \Sigma_k^{ll} - \Sigma_k^{ly}(\Sigma_k^{yy})^{-1}(\Sigma_k^{ly})^T - \\ &\quad -\Lambda_k^{l\bar{p}}(\Lambda_k^{\bar{p}})^{-1}(\Lambda_k^{l\bar{p}})^T. \end{aligned} \tag{44c}$$

For notational convenience, we define

$$\Gamma_k^{py} = \Sigma_k^{py}(\Sigma_k^{yy})^{-1} \tag{45a}$$
$$\Gamma_k^{ly} = \Sigma_k^{ly}(\Sigma_k^{yy})^{-1} \tag{45b}$$
$$\Gamma_k^{lp} = \Lambda_k^{l\bar{p}}(\Lambda_k^{\bar{p}})^{-1}. \tag{45c}$$

Now using (42a) and (44a), the model as described by (38a)–(38c) can be re-written as

$$x_{k+1}^l = f_k^l + A_k^l x_k^l + G_k^l[\bar{w}_k^l + \Gamma_k^{ly}e_k + \Gamma_k^{lp}\bar{w}_k^p] \tag{46a}$$
$$x_{k+1}^p = f_k^p + A_k^p x_k^l + G_k^p[\bar{w}_k^p + \Gamma_k^{py}e_k] \tag{46b}$$
$$y_k = h_k + C_k x_k^l + e_k. \tag{46c}$$

Defining the pseudo measurements obtained from the residuals as

$$Z_k^{(1)} = (x_{k+1}^p - f_k^p) \tag{47a}$$
$$Z_k^{(2)} = (y_k - h_k). \tag{47b}$$

It then follows that

$$Z_k^{(2)} = C_k x^l + e_k \tag{48a}$$
$$Z_k^{(1)} = A_k^p x_k^l + G_k^p[\bar{w}_k^p + \Gamma_k^{py}\{Z_k^{(2)} - C_k x_k^l\}]. \tag{48b}$$

Defining

$$\bar{A}_k^p = [A_k^p - G_k^p\Gamma_k^{py}C_k], \tag{49}$$

so that equation (48b) can now be written as

$$Z_k^{(1)} = \bar{A}_k^p x_k^l + G_k^p\Gamma_k^{py}Z_k^{(2)} + G_k^p\bar{w}_k^p. \tag{50}$$

[2]A multivariate Gaussian noise is completely characterized by the second order statistics. Hence dependent Gaussian noise implies correlated Gaussian noise.

Similarly, using equation (48a) and (50) in equation (46a) and assuming G_k^p to be invertible, we have

$$\begin{aligned} x_{k+1}^l &= f_k^l + A_k^l x_k^l + G_k^l[\bar{w}_k^l + \Gamma_k^{ly}\{Z_k^{(2)} - C_k x_k^l\} + \\ &\quad + \Gamma_k^{lp}(G_k^p)^{-1}\{Z_k^{(1)} - \bar{A}_k^p x_k^l - G_k^p \Gamma_k^{py} Z_k^{(2)}\}] \\ &= f_k^l + [A_k^l - G_k^l \Gamma_k^{ly} C_k - G_k^l (G_k^p)^{-1} \Gamma_k^{lp} \bar{A}_k^p] x_k^l \; + \\ &\quad + G_k^l [\Gamma_k^{ly} - \Gamma_k^{lp}(G_k^p)^{-1}(G_k^p)\Gamma_k^{py}] Z_k^{(2)} \; + \\ &\quad + [G_k^l \Gamma_k^{lp}(G_k^p)^{-1}] Z_k^{(1)} + G_k^l \bar{w}_k^l. \end{aligned} \tag{51}$$

Define

$$\bar{A}_k^l = [A_k^l - G_k^l \Gamma_k^{ly} C_k - G_k^l (G_k^p)^{-1} \Gamma_k^{lp} \bar{A}_k^p] \tag{52}$$

and

$$\bar{f}_k^l = f_k^l + G_k^l [\Gamma_k^{ly} - \Gamma_k^{lp}\Gamma_k^{py}] Z_k^{(2)} + [G_k^l \Gamma_k^{lp}(G_k^p)^{-1}] Z_k^{(1)}, \tag{53}$$

so that, equation (51) can be written as

$$x_{k+1}^l = \bar{f}_k^l + \bar{A}_k^l x_k^l + G_k^l \bar{w}_k^l. \tag{54}$$

C. Revised Model Definition

The state space model obtained using equations (54), (50) and (48a), is linear, driven by independent zero mean Gaussian noises as

$$x_{k+1}^l = \bar{f}_k^l + \bar{A}_k^l x_k^l + G_k^l \bar{w}_k^l \tag{55a}$$

$$Z_k^{(1)} = \bar{A}_k^p x_k^l + G_k^p \Gamma_k^{py} Z_k^{(2)} + G_k^p \bar{w}_k^p \tag{55b}$$

$$Z_k^{(2)} = C_k x_k^l + e_k \tag{55c}$$

with

$$Z_k^{(1)} = (x_{k+1}^p - f_k^p) \tag{55d}$$

$$Z_k^{(2)} = (y_k - h_k), \tag{55e}$$

where

$$\mathrm{Cov}\begin{pmatrix} \bar{w}_k^l \\ \bar{w}_k^p \\ e_k \end{pmatrix} = \begin{bmatrix} \Lambda_k^{\bar{l}} & 0 & 0 \\ 0 & \Lambda_k^{\bar{p}} & 0 \\ 0 & 0 & \Sigma_k^{yy} \end{bmatrix}. \tag{56}$$

Here $\bar{f}_k^l$, $\bar{A}_k^l$ and $\bar{A}_k^p$ are obtained using equations (53), (52) and (49) respectively. Now one can apply the standard results (e.g. in [19]) for MPF utilizing a linear-Gaussian substructure on this revised model. The summary of the main steps are given in Table II and the details are outlined in appendix. An application of this framework to the terrain navigational problem is presented in [24].

V. Estimating the unknown noise statistics of dependent Gaussian noises

Most of the estimation algorithms involving a state space model assume a prior knowledge of the noise distributions, whereas the properties of the noise processes are often unknown for many practical problems. Moreover, the noise distributions may be non-stationary or state dependent, which further prevents the so called off-line tuning approach. For linear Gaussian model, the adaptive Kalman filters can estimate the unknown noise parameters jointly with the state. However, the same problem for a general state space problem is less studied. In this section, we address a joint state and noise parameter estimation problem involving dependent Gaussian noise processes using PF.

To proceed with, consider the following state-space model with additive process and measurement noises:

$$x_k = f(x_{k-1}) + v_k, \tag{57a}$$

$$y_k = h(x_k) + e_k, \qquad k = 1, 2, \ldots \tag{57b}$$

where $x_k \in \mathbb{R}^{n_v}$ is the hidden state and $y_k \in \mathbb{R}^{n_e}$ is the measurement, at time step k. v_k and e_k are the corresponding process and measurement Gaussian noises, which are dependent to each other [3]. Define $w_k \in \mathbb{R}^d$ (here $d = n_v + n_e$) as

$$w_k = \begin{pmatrix} v_k \\ e_k \end{pmatrix}, \tag{58}$$

where the sequence of w_k is independent Gaussian conditioned on an unknown mean μ_k and covariance matrix Σ_k. Here we assume the parameters (μ_k, Σ_k) to be slowly varying in time. This slowly varying nature can arise e.g., due to model mis-specification [26]. Now, the conditional distribution of w_k is given as

$$w_k | (\mu_k, \Sigma_k) \sim \mathcal{N}(\mu_k, \Sigma_k) \tag{59}$$

with

$$\mu_k = [\mu_{v,k}^T \; \mu_{e,k}^T]^T; \quad \Sigma_k = \begin{bmatrix} \Sigma_{vv,k} & \Sigma_{ve,k} \\ \Sigma_{ve,k}^T & \Sigma_{ee,k} \end{bmatrix}. \tag{60}$$

One key objective here is to learn the noise parameters $\theta_k \triangleq (\mu_k, \Sigma_k)$ adaptively as the new measurement arrives. The problem of learning those parameters (using a MPF approach) when the noises are independent, has recently been addressed in [22]. Here we consider a more general case with dependent noises.

A. Conjugate prior for unknown Gaussian noise parameters

Following [2], a suitable conjugate prior for (μ_k, Σ_k) is known to be a Normal-inverse-Wishart distribution of the form $(\mu_k, \Sigma_k) \sim \mathrm{NiW}(\nu_k, V_k)$, where

$$\mu_k | \Sigma_k \sim \mathcal{N}(\hat{\mu}_k, \hat{\Sigma}_k) \tag{61a}$$

$$\Sigma_k \sim \mathrm{iW}(\nu_k - d - 1, \Lambda_k). \tag{61b}$$

Here $\mathrm{iW}(\cdot)$ denotes Inverse Wishart distribution. The parameters ν_k and V_k hold the sufficient statistics and can be updated recursively. The relevant quantities are defined as,

$$V_k = \begin{pmatrix} V_{w_k w_k,k} & V_{1 w_k,k} \\ V_{w_k 1,k} & V_{11,k} \end{pmatrix} \tag{62a}$$

$$\hat{\mu}_k = V_{11,k}^{-1} V_{1 w_k,k} \tag{62b}$$

$$\hat{\Sigma}_k = \Sigma_k V_{11,k}^{-1} \tag{62c}$$

$$\Lambda_k = V_{w_k w_k,k} - V_{1 w_k,k} V_{11,k}^{-1} V_{w_k 1,k} \tag{62d}$$

[3] We note that this corresponds to Type II dependency (as in Figure 3, but v_{k-1} is replaced by v_k for notational convenience). Treatment of Type I dependency, which we have not included here, can be carried out similarly.

Table II
SUMMARY OF THE INFORMATION STEPS OF MPF UTILIZING A LINEAR-GAUSSIAN SUBSTRUCTURE

Prior	$p(x_k^l, X_k^p\|Y_k) = \underbrace{p(x_k^l\|X_k^p, Y_k)}\,\underbrace{p(X_k^p\|Y_k)}$
PF: TU	$\underbrace{p(X_k^p\|Y_k)}_{\text{prior}} \Rightarrow\ p(X_{k+1}^p\|Y_k)$
KF: Dyn MU	$\underbrace{p(x_k^l\|X_k^p, Y_k)}_{\text{prior}} \Rightarrow\ p(x_k^l\|X_{k+1}^p, Y_k)$
KF: TU	$p(x_{k+1}^l\|X_{k+1}^p, Y_k) = \int p(x_{k+1}^l\|X_{k+1}^p, x_k^l, Y_k)\,\underbrace{p(x_k^l\|X_{k+1}^p, Y_k)}_{\text{KF:Dyn MU}}\,dx_k^l$
y_{k+1} is available now	
PF: MU	$\underbrace{p(X_{k+1}^p\|Y_k)}_{\text{PF:TU}} \Rightarrow\ p(X_{k+1}^p\|Y_{k+1})$
KF: MU	$\underbrace{p(x_{k+1}^l\|X_{k+1}^p, Y_k)}_{\text{KF:TU}} \Rightarrow\ p(x_{k+1}^l\|X_{k+1}^p, Y_{k+1})$
Posterior	$p(x_{k+1}^l, X_{k+1}^p\|Y_{k+1}) = p(x_{k+1}^l\|X_{k+1}^p, Y_{k+1})p(X_{k+1}^p\|Y_{k+1})$

$V_{w_k w_k,k}$ is defined as the upper-left $d \times d$ sub-block of $V_k \in \mathbb{R}^{(d+1)\times(d+1)}$. The joint density of (μ_k, Σ_k) is of the form

$$p(\mu_k, \Sigma_k) = \mathrm{NiW}(\nu_k, V_k) \tag{63a}$$

$$= \frac{1}{c}|\Sigma_k|^{-\frac{\nu_k}{2}} \times \times \exp(-\frac{1}{2}\mathrm{tr}(\Sigma_k^{-1}[-I_d, \mu_k]V_k[-I_d, \mu_k]^T)), \tag{63b}$$

where c is the normalizing constant. Consequently, the posterior predictive distribution of w_k can be analytically obtained as a multivariate Student's t distribution of the form

$$\begin{aligned} p(w_k) &= t_{\tilde{\nu}_k}(\tilde{\mu}_k, \tilde{\Sigma}_k) \\ &= \frac{\Gamma(\tilde{\nu}_k/2 + d/2)}{\Gamma(\tilde{\nu}_k/2)} \frac{|\tilde{\Sigma}_k|^{-1/2}}{(\tilde{\nu}_k\pi)^{(d/2)}} \times \\ &\quad \times \left[1 + \frac{1}{\tilde{\nu}_k}(w_k - \tilde{\mu}_k)^T\tilde{\Sigma}_k^{-1}(w_k - \tilde{\mu}_k)\right]^{-(\frac{\tilde{\nu}_k+d}{2})} \end{aligned} \tag{64}$$

where $\tilde{\nu}_k = \nu_k - d + 1$, is the degree of freedom, $\tilde{\mu}_k = \mu_k$ and $\tilde{\Sigma}_k = \frac{(1+V_{11,k})}{(\nu_k-d+1)V_{11,k}}\Lambda_k$ are the location and the scale parameters of the above Student's t distribution, with $\Gamma(\cdot)$ as the Gamma function. If we now partition the variable w_k into two blocks (corresponding to v_k and e_k respectively), then the marginals of w_k (i.e. v_k and e_k) are also obtained as Student's t distributions [20]

$$v_k \sim t_{\tilde{\nu}_k}(\tilde{\mu}_{v,k}, \tilde{\Sigma}_{vv,k}) \tag{65a}$$

$$e_k \sim t_{\tilde{\nu}_k}(\tilde{\mu}_{e,k}, \tilde{\Sigma}_{ee,k}). \tag{65b}$$

Moreover, the conditional, $p(e_k|v_k)$ can be obtained as,

$$p(e_k|v_k) \sim t_{(\tilde{\nu}_k+d_e)}(\tilde{\mu}_{e|v,k}, \tilde{\Sigma}_{e|v,k}) \tag{65c}$$

with

$$\tilde{\mu}_{e|v,k} = \tilde{\mu}_{e,k} + \tilde{\Sigma}_{ve,k}^T\tilde{\Sigma}_{vv,k}^{-1}(v_k - \tilde{\mu}_{v,k}) \tag{65d}$$

$$\tilde{\Sigma}_{e|v,k} = h_{e|v,k}(\tilde{\Sigma}_{ee,k} - \tilde{\Sigma}_{ve,k}^T\tilde{\Sigma}_{vv,k}^{-1}\tilde{\Sigma}_{ve,k}) \tag{65e}$$

$$h_{e|v,k} = \frac{1}{(\tilde{\nu}_k + d_v)}[\tilde{\nu}_k + (v_k - \tilde{\mu}_{v,k})^T \times \times\ \tilde{\Sigma}_{vv,k}^{-1}(v_k - \tilde{\mu}_{v,k})]. \tag{65f}$$

B. Joint state and parameter estimation

Our interest lies in estimating $p(x_k|Y_k)$ and $p(\theta_k|Y_k)$ recursively over time. Suppose we are at time step $k-1$ and $p(x_{k-1}|Y_{k-1})$ is approximately given in the form of weighted particle cloud. The propagation of $p(x_{k-1}|Y_{k-1})$ to the next time step using a running PF is shown below:

1) Particle filter update: PF approximates $p(X_{k-1}|Y_{k-1})$ by the empirical measure as

$$p(X_{k-1}|Y_{k-1}) \simeq \sum_{i=1}^{N} \omega_{k-1}^{(i)}\delta_{X_{k-1}^{(i)}}(X_{k-1}). \tag{66}$$

Now we generate new samples $x_k^{(i)}$ from the proposal $q(x_k|\cdot)$ and form the trajectories $X_k^{(i)}$ as $X_k^{(i)} \triangleq [X_{k-1}^{(i)}, x_k^{(i)}]$ such that

$$p(X_k|Y_k) \simeq \sum_{i=1}^{N} \omega_k^{(i)}\delta_{X_k^{(i)}}(X_k). \tag{67}$$

The weight update of PF can be obtained recursively as

$$\omega_k^{(i)} = \omega_{k-1}^{(i)} \frac{p(y_k|X_k^{(i)}, Y_{k-1})p(x_k^{(i)}|X_{k-1}^{(i)}, Y_{k-1})}{q(x_k^{(i)}|\cdot)}. \tag{68}$$

So to obtain the new weights, we need to evaluate $p(y_k|X_k, Y_{k-1})$ and $p(x_k|X_{k-1}, Y_{k-1})$ respectively. Now,

from (57a)–(57b), $p([x_k - f(x_{k-1})]|X_{k-1}, Y_{k-1}) = p(v_k)$, which is given by (65a). So,

$$(x_k|X_{k-1}, Y_{k-1}) \quad \sim t_{(\tilde{\nu}_k+d_v)}(\tilde{\mu}^*_{v,k}, \tilde{\Sigma}_{vv,k}), \tag{69}$$

where $\tilde{\mu}^*_{v,k} = \tilde{\mu}_{v,k} + f(x_{k-1})$.

Similarly, $p([y_k - h(x_k)]|X_k, Y_{k-1}) = p(e_k|v_k)$, given by (65c). So $p(y_k|X_k, Y_{k-1})$ is another Student's *t* distribution as given by (65c), with mean modified as

$$\bar{\tilde{\mu}}^*_{e|v,k} = \tilde{\mu}_{e|v,k} + h(x_k). \tag{70}$$

Subsequently, from (67), one can approximate the marginal as

$$p(x_k|Y_k) \simeq \sum_{i=1}^{N} \omega_k^{(i)} \delta_{x_k^{(i)}}(x_k). \tag{71}$$

PF provides an approximation of the joint smoothing distribution recursively. However, as k increases, such particle filter suffers due to a progressively impoverished particle representation as a result of resampling. This problem is known as the particle path degeneracy problem [8]. On the other hand, uniform convergence in time of the particle filter is known under the mixing assumptions as in [18]. This property ensures that any error is forgotten exponentially with time and explains why the particle filter works for marginal filter density (as in (71)) in most practical applications.

2) Updates on noise parameters: We note from ((57a)–(57b)) that knowing the sequence (X_{k-1}, Y_{k-1}) would lead us to the completely observed noise sequence $w_{1:k-1}$. Now suppose that $p(\theta_{k-1}|X_{k-1}, Y_{k-1})$ is given by a Normal-inverse-Wishart prior[4] as

$$p(\theta_{k-1}|X_{k-1}, Y_{k-1}) = p(\theta_{k-1}|w_{k-1}) = \mathrm{NiW}(\nu_{k-1}, V_{k-1}). \tag{72}$$

Since we assume the noise parameters to be slowly varying, we approximate the time update step using principle of exponential forgetting [25] as

$$p(\theta_k|X_{k-1}, Y_{k-1}) = p(\theta_k|w_{k-1}) = \mathrm{NiW}(\lambda\nu_{k-1}, \lambda V_{k-1}), \tag{73}$$

where $\lambda \in [0, 1]$ is the forgetting factor used [5]. Via Bayesian conjugacy, the posterior $p(\theta_k|X_k, Y_k)$ is again a Normal-inverse-Wishart distribution as

$$p(\theta_k|X_k, Y_k) = p(\theta_k|w_k) = \mathrm{NiW}(\nu_k, V_k), \tag{74}$$

with

$$V_k = \lambda V_{k-1} + \begin{pmatrix} w_k \\ 1 \end{pmatrix} \begin{pmatrix} (w_k)^T & 1 \end{pmatrix} \tag{75}$$

$$\nu_k = \lambda\nu_{k-1} + 1, \tag{76}$$

where

$$w_k = \begin{pmatrix} v_k \\ e_k \end{pmatrix} = \begin{pmatrix} x_k - f(x_{k-1}) \\ y_k - h(x_k) \end{pmatrix}. \tag{77}$$

[4]Treating both μ_k and Σ_k to be unknown might have implications on the observability and identifiability of the model. When Σ_k is the only unknown parameter, a suitable conjugate prior is given by the inverse Wishart distribution and the posterior predictive is again a Student's *t* distribution [27].

[5]Similar argument was put forward in [26], page 369, where λ is called a discount factor.

We again stress that the path dependency of the parameter posterior leads to accumulation of errors over time. However, by using the principle of exponential forgetting, this is less critical here. Now, we define $T_k(x_k) := p(\theta_k|x_k, Y_k)$. Since $T_k(x_k)$ can be written as

$$T_k(x_k) = \int p(\theta_k|X_k, Y_k) p(X_{k-1}|Y_{k-1}, x_k) dX_{k-1}, \tag{78}$$

we can establish the recursive relation for $T_k(x_k)$ as

$$\begin{aligned} T_k(x_k) &= \iint \frac{p(\theta_k|X_k, Y_k) p(\theta_{k-1}|X_{k-1}, Y_{k-1})}{p(\theta_{k-1}|X_{k-1}, Y_{k-1})} \times \\ &\quad \times p(X_{k-2}|Y_{k-2}, x_{k-1}) p(x_{k-1}|Y_{k-1}, x_k) \times \\ &\quad \times dX_{k-2}\, dx_{k-1} \\ &= \int \frac{p(\theta_k|X_k, Y_k)}{p(\theta_{k-1}|X_{k-1}, Y_{k-1})} T_{k-1}(x_{k-1}) \times \\ &\quad \times p(x_{k-1}|Y_{k-1}, x_k) dx_{k-1}. \end{aligned} \tag{79}$$

From the forward particle filter, we have

$$p(x_{k-1}|Y_{k-1}, x_k) \approx \frac{\sum_{j=1}^{N} \omega_{k-1}^{(j)} p(x_k|x_{k-1}^{(j)}) \delta_{x_{k-1}^{(j)}}(x_{k-1})}{\sum_{l=1}^{N} \omega_{k-1}^{(l)} p(x_k|x_{k-1}^{(l)})}. \tag{80}$$

Now using (80) we can approximate (79) as

$$\begin{aligned} T_k(x_k^{(i)}) = \sum_{j=1}^{N} \frac{\mathrm{NiW}(\nu_k^{(ij)}, V_k^{(ij)})}{\mathrm{NiW}(\nu_{k-1}^{(j)}, V_{k-1}^{(j)})} T_{k-1}(x_{k-1}^{(j)}) \times \\ \times \frac{\omega_{k-1}^{(j)} p(x_k^{(i)}|x_{k-1}^{(j)})}{\sum_{l=1}^{N} \omega_{k-1}^{(l)} p(x_k^{(i)}|x_{k-1}^{(l)})}, \end{aligned} \tag{81}$$

with

$$V_k^{(ij)} = \lambda V_{k-1}^{(j)} + \begin{pmatrix} w_k^{(ij)} \\ 1 \end{pmatrix} \begin{pmatrix} (w_k^{(ij)})^T & 1 \end{pmatrix} \tag{82}$$

$$\nu_k^{(ij)} = \lambda\nu_{k-1}^{(j)} + 1, \tag{83}$$

where we define

$$w_k^{(ij)} = \begin{pmatrix} x_k^{(i)} - f(x_{k-1}^{(j)}) \\ y_k - h(x_k^{(i)}) \end{pmatrix}. \tag{84}$$

Finally, using (81) and (71), we get

$$\begin{aligned} p(\theta_k|Y_k) &\approx \int T_k(x_k) p(x_k|Y_k) dx_k \\ &= \sum_{i=1}^{N} \omega_k^{(i)} T_k(x_k^{(i)}). \end{aligned} \tag{85}$$

C. Algorithmic summary

In this section, we give a summary of one step of the main algorithm.

Algorithm 2: **[Estimating the statistics of unknown dependent Gaussian noises]**

- At time step $(k-1)$:
  we have $\left\{x_{k-1}^{(i)}, \omega_{k-1}^{(i)}, T_{k-1}(x_{k-1}^{(i)})\right\}_{i=1}^{N}$, such that

$$p(x_{k-1}|Y_{k-1}) \simeq \sum_{i=1}^{N} \omega_{k-1}^{(i)} \delta_{x_{k-1}^{(i)}}(x_{k-1}),$$
$$p(\theta_{k-1}|Y_{k-1}) \simeq \sum_{i=1}^{N} \omega_{k-1}^{(i)} T_{k-1}(x_{k-1}^{(i)}).$$

- At time step (k):
  with new observation y_k
  - Generate particles from the proposal $x_k^{(i)} \sim q(x_k|\cdot)$.
  - Recursive weight update: $\omega_k^{(i)}$ according to (68) using (69)–(70).

  Recursive computation $T_k(x_k^{(i)})$ using (81)–(84).

To summarize this section, the problem of estimating the unknown state from a general state space model involving unknown Gaussian noise characteristics is presented. Specifically, we addressed the online joint state and noise parameter estimation problem involving additive dependent Gaussian noises using a PF based approach.

VI. Sampling Continuous Time Models

Dependency between process and observation noises in a general state space model occurs naturally in many situation, although it is often ignored in favor of modeling conveniences. This section motivates the importance of dependent noise processes for a wide range of applications starting with a continuous time models.

Consider the following continuous time linear model with a noisy input $u(t)$,

$$\dot{x}(t) = Ax(t) + B\big(u(t) + v(t)\big), \tag{86a}$$
$$y(t_k) = Cx(t_k) + e(t_k), \tag{86b}$$
$$e(t_k) \in \mathcal{N}(0, R). \tag{86c}$$

where $x(t)$ is the continuous time state, $y(t_k)$ is the discrete time observation and $v(t)$ is a zero mean white Gaussian noise process, with $E(v(t)v(s)^T) = Q\delta(t-s)$, where $E(\cdot)$ denotes the expectation operator and $\delta(\cdot)$ is the Dirac delta function. For the continuous time model, it is natural to assume that $v(t)$ and $e(t_k)$ are independent. However, discretization of the model to a sampled discrete time model defined at the time instants t_k may introduce dependence. Tables III and IV summarize different sampling strategies (see Chapter 13.1 in [5]) for single and double integrators, respectively. The corresponding discrete time model is given by,

$$x_{k+1} = Fx_k + G\big(u_k + v_k\big) = Fx_k + Gu_k + \bar{v}_k, \tag{87a}$$
$$y_k = Hx_k + J\big(u_k + v_k\big) + e_k = Hx_k + Ju_k + \bar{e}_k, \tag{87b}$$
$$\begin{pmatrix} \bar{v}_k \\ \bar{e}_k \end{pmatrix} \in \mathcal{N}\left(0, \begin{bmatrix} GQG^T & GQJ^T \\ JQG^T & JQJ^T + R \end{bmatrix}\right). \tag{87c}$$

We note from Tables III and IV that we get dependence ($J \neq 0$ and $G \neq 0$) in all cases except for ZOH, where a piecewise constant process noise is assumed. The following application areas surveyed in [16] are important cases:

- *Navigation* – The input is one of or a combination of acceleration and angular rates as measured by accelerometers and gyroscopes. The basic form of continuous time dynamics is in both cases given by a double integrator $\ddot{p}(t) = u(t) + v(t)$. The sensors are typically low-pass filtered before sampling to avoid aliasing, and thus the true acceleration/angular rate $u(t)+v(t)$ is not piecewise constant.
- *Odometry* – Here the angular speeds of two wheels are measured and used to compute the position based on the principle of dead-reckoning. The basic form of continuous time dynamics here is a single integrator $\dot{p}(t) = u(t)+v(t)$. The angular rate encoders are typically low-pass filtered before sampling.
- *Tracking* – The input is an unobserved force, so $u(t) = 0$ and $v(t)$ models the force input. The dynamics is given by a double integrator $\ddot{p}(t) = v(t)$. A suitable model for $v(t)$ is subject to debate, but all cases except when it is assumed piecewise constant synchronized with the external position sensor lead to dependent noise.

VII. Conclusions

The fact that the process noise is dependent to the measurement noise in sampled models is often neglected in literature. There might be several reasons for this. The process noise is often instrumental for the tuning, so its physical interpretation is of less importance. Another reason is that the correlation can be quite small for fast sampling compared to the time constant of the system. A final reason might be that the PF theory is not yet adapted to dependent noise, in contrast to the KF literature where this is more of a standard assumption with a rather simple remedy.

We have extended the particle filter theory in three ways. First, the important choice of proposal density is examined. Both the prior and optimal proposal are derived for dependent noise, for two different cases of dependence structures. For Gaussian noise, the optimal proposal gets an analytical expression, which further simplifies to a Gaussian for the prior proposal. Second, the marginalized particle filter, that is instrumental for real-world applications to mitigate the curse of dimensionality, was derived for dependent noise. Third, the less studied problem of estimating parameters in the noise distributions was addressed for the case of Gaussian dependent noise.

Acknowledgment

The authors would like to thank the Linnaeus research environment CADICS, funded by the Swedish Research Council for the financial support.

Appendix
Detailed steps of MPF using a linear-Gaussian substructure

At time zero, $p(x_0^l|X_0^p, Y_0) = \mathcal{N}(\hat{x}_{0|0}^l, P_{0|0})$. Under favorable mixing condition, we assume that

Table III
CORRELATION DUE TO SAMPLING OF THE STATE $x(t) = (p(t),\ v(t))^T$ IN A DOUBLE INTEGRATOR USING ZERO-ORDER HOLD (ZOH, ASSUMING PIECEWISE CONSTANT INPUT), FIRST-ORDER HOLD (FOH, ASSUMING PIECEWISE LINEAR INPUT), AND BILINEAR TRANSFORMATION (BIL), WHICH IS AN OFTEN USED APPROXIMATION FOR BAND-LIMITED SIGNALS.

Continuous time	$A = \begin{pmatrix} 0_n & I_n \\ 0_n & 0_n \end{pmatrix}$	$B = \begin{pmatrix} 0_n \\ I_n \end{pmatrix}$	$C = (I_n,\ 0_n)$	$D = 0_n$
ZOH	$F = \begin{pmatrix} I_n & TI_n \\ 0_n & I_n \end{pmatrix}$	$G = \begin{pmatrix} \frac{T^2}{2} I_n \\ TI_n \end{pmatrix}$	$H = (I_n,\ 0_n)$	$J = 0_n$
FOH	$F = \begin{pmatrix} I_n & TI_n \\ 0_n & I_n \end{pmatrix}$	$G = \begin{pmatrix} T^2 I_n \\ TI_n \end{pmatrix}$	$H = (I_n,\ 0_n)$	$J = \frac{T^2}{6} I_n$
BIL	$F = \begin{pmatrix} I_n & TI_n \\ 0_n & I_n \end{pmatrix}$	$G = \begin{pmatrix} \frac{T^2}{4} I_n \\ \frac{T}{2} I_n \end{pmatrix}$	$H = \left(I_n,\ \frac{T}{2} I_n\right)$	$J = \frac{T^2}{2} I_n$

Table IV
SIMILAR TO TABLE III, BUT FOR A SINGLE INTEGRATOR USING THE STATE $x(t) = p(t)$.

Continuous time	$A = 0_n$	$B = I_n$	$C = I_n$	
ZOH	$F = I_n$	$G = TI_n$	$H = I_n$	$J = 0_n$
FOH	$F = I_n$	$G = TI_n$	$H = I_n$	$J = \frac{T}{2} I_n$
BIL	$F = I_n$	$G = I_n$	$H = TI_n$	$J = \frac{T}{2} I_n$

$p(x_k^l|X_k^p, Y_k) = \mathcal{N}(\hat{x}_{k|k}^l, P_{k|k})$ at an arbitrary time, k. Now we outline here one complete cycle of propagating the joint density of the state conditional on the available observation.

PF time update (PF: TU)

At this stage, it is required to generate N new particles (samples) $\left\{x_{k+1}^{p(i)}\right\}_{i=1}^N$ from the appropriate importance function $q(x_{k+1}^p|\cdot)$.

KF dynamic measurement update (KF : DYN MU)

$$p(x_k^l|X_{k+1}^p, Y_k) = \frac{p(x_{k+1}^p|X_k^p, x_k^l, Y_k)p(x_k^l|X_k^p, Y_k)}{\int p(x_{k+1}^p|X_k^p, x_k^l, Y_k)p(x_k^l|X_k^p, Y_k)dx_k^l}. \tag{88}$$

From the prior, we have $p(x_k^l|X_k^p, Y_k) = \mathcal{N}(\hat{x}_{k|k}^l, P_{k|k})$. Now, at this stage, $Z_k^{(1)}$ is available. Let $p(x_k^l|X_{k+1}^p, Y_k) = \mathcal{N}(\hat{x}_{k|k}^{l*}, P_{k|k}^*)$. Then following the proof (part 2) of [19],

$$N_k^* = \bar{A}_k^p P_{k|k}(\bar{A}_k^p)^T + G_k^p \Lambda_k^{\bar{p}}(G_k^p)^T \tag{89a}$$
$$L_k = P_{k|k}(\bar{A}_k^p)^T (N_k^*)^{-1} \tag{89b}$$
$$\hat{x}_{k|k}^{l*} = \hat{x}_{k|k}^l + L_k(Z_k^{(1)} - \bar{A}_k^p \hat{x}_{k|k}^l - G_k^p \Gamma_k^{py} Z_k^{(2)}) \tag{89c}$$
$$P_{k|k}^* = P_{k|k} - L_k(N_k^*)(L_k)^T \tag{89d}$$

KF time update (KF: TU)

$$p(x_{k+1}^l|X_{k+1}^p, Y_k) = \int p(x_{k+1}^l|X_{k+1}^p, x_k^l, Y_k) \times \times p(x_k^l|X_{k+1}^p, Y_k)dx_k^l, \tag{90}$$

where $p(x_k^l|X_{k+1}^p, Y_k) = \mathcal{N}(\hat{x}_{k|k}^{l*}, P_{k|k}^*)$. It follows that $p(x_{k+1}^l|X_{k+1}^p, Y_k) = \mathcal{N}(\hat{x}_{k+1|k}^l, P_{k+1|k})$ with

$$\hat{x}_{k+1|k}^l = \bar{f}_k^l + \bar{A}_k^l \hat{x}_{k|k}^{l*} \tag{91a}$$
$$P_{k+1|k} = \bar{A}_k^l P_{k|k}^*(\bar{A}_k^l)^T + G_k^l \Lambda_k^{\bar{l}}(G_k^l)^T \tag{91b}$$

PF measurement update (PF: MU)

With new measurement y_{k+1}, we get $p(X_{k+1}|Y_{k+1}) \simeq \sum_{i=1}^N \omega_{k+1}^{(i)} \delta_{X_{k+1}^{(i)}}(X_{k+1})$, where the weights of the particle filter can be recursively updated according to :

$$\omega_{k+1}^{(i)} = \omega_k^{(i)} \frac{p(y_{k+1}|X_{k+1}^{p(i)}, Y_k)p(x_{k+1}^{p(i)}|X_k^{p(i)}, Y_k)}{q(x_{k+1}^{p(i)}|\cdot)}. \tag{92}$$

The transition density $p(x_{k+1}^p|X_k^p, Y_k)$ can be obtained as

$$p(x_{k+1}^p|X_k^p, Y_k) = \int p(x_{k+1}^p|X_k^p, x_k^l, Y_k)p(x_k^l|X_k^p, Y_k)dx_k^l,$$

where, $p(x_k^l|X_k^p, Y_k) = \mathcal{N}(\hat{x}_{k|k}^l, P_{k|k})$, as obtained from the prior. Since at this stage $Z_k^{(2)}$ is known, we have from (55b) and (55d), $p(x_{k+1}^p|X_k^p, Y_k) = \mathcal{N}(\mu_{k+1}^{tr}, \Sigma_{k+1}^{tr})$, where

$$\mu_{k+1}^{tr} = f_k^p + \bar{A}_k^p \hat{x}_{k|k}^l + G_k^p \Gamma_k^{py} Z_k^{(2)} \tag{93a}$$
$$\Sigma_{k+1}^{tr} = \bar{A}_k^p P_{k|k}(\bar{A}_k^p)^T + G_k^p \Lambda_k^{\bar{p}}(G_k^p)^T. \tag{93b}$$

We now obtain the likelihood density $p(y_{k+1}|X_{k+1}^p, Y_k)$ as

$$p(y_{k+1}|X_{k+1}^p, Y_k) = \int p(y_{k+1}|X_{k+1}^p, x_{k+1}^l, Y_k) \times \times p(x_{k+1}^l|X_{k+1}^p, Y_k)dx_{k+1}^l. \tag{94}$$

Let $p(y_{k+1}|X_{k+1}^p, Y_k) = \mathcal{N}(\mu_{k+1}^L, \Sigma_{k+1}^L)$. Then

$$\mu_{k+1}^L = h_{k+1} + C_{k+1}\hat{x}_{k+1|k}^l \tag{95a}$$
$$\Sigma_{k+1}^L = C_{k+1}P_{k+1|k}(C_{k+1})^T + \Sigma_{k+1}^{yy}. \tag{95b}$$

KF measurement update (KF: MU)

From KF time update stage, we have $p(x_{k+1}^l|X_{k+1}^p, Y_k) = \mathcal{N}(\hat{x}_{k+1|k}^l, P_{k+1|k})$. As y_{k+1} is available now, it implies that $Z_{k+1}^{(2)} = (y_{k+1} - h_{k+1})$ is also available at this stage. Now

following the proof (part 1) of [19], we have

$$p(x_{k+1}^l|X_{k+1}^p, Y_{k+1}) = \frac{p(y_{k+1}|X_{k+1}^p, x_{k+1}^l, Y_k)p(x_{k+1}^l|X_{k+1}^p, Y_k)}{\int p(y_{k+1}|X_{k+1}^p, x_{k+1}^l, Y_k)p(x_{k+1}^l|X_{k+1}^p, Y_k)dx_{k+1}^l}. \quad (96)$$

Using the fact that the measurement noise and thereby $p(y_{k+1}|X_{k+1}^p, x_{k+1}^l, Y_k)$ is Gaussian, and using KF, We can show that

$p(x_{k+1}^l|X_{k+1}^p, Y_{k+1}) = \mathcal{N}(\hat{x}_{k+1|k+1}^l, P_{k+1|k+1})$, where

$$M_{k+1} = C_{k+1}P_{k+1|k}(C_{k+1})^T + \Sigma_{k+1}^{yy} \quad (97a)$$

$$K_{k+1} = P_{k+1|k}C_{k+1}(M_{k+1})^{-1} \quad (97b)$$

$$\hat{x}_{k+1|k+1}^l = \hat{x}_{k+1|k}^l + K_{k+1}(Z_{k+1}^{(2)} - C_{k+1}\hat{x}_{k+1|k}^l) \quad (97c)$$

$$P_{k+1|k+1} = P_{k+1|k} - K_{k+1}M_{k+1}(K_{k+1})^T \quad (97d)$$

□

REFERENCES

[1] F. Gustafsson, "Adaptive Filtering and Change Detection", *John Wiley and Sons*, 2000.

[2] V. Smidl and A. Quinn, "The Variational Bayes Method in Signal Processing", *Springer*, 2006.

[3] A. Doucet, S. Godsill, C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering", *Statistics and Computing*,Vol 10, No. 3, pp. 197–208, 2000.

[4] T. Kailath, A. H. Sayed and B. Hassibi, "Linear Estimation", *Prentice-Hall, Upper Saddle River, NJ*, 2000.

[5] F. Gustafsson, *Statistical Sensor Fusion*, Studentlitteratur, 2010.

[6] S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A Tutorial on Particle Filters for online nonlinear/non-Gaussian Bayesian tracking", *IEEE Transaction on Signal Processing*, vol. 50(2), pp. 174–188, Feb. 2002.

[7] J. V. Candy, "Bootstrap particle filtering", *IEEE Signal Processing Magazine*, vol. 24(4), pp. 73–85, 2007.

[8] O. Cappe, S. J. Godsill and E. Moulines, "An overview of existing methods and recent advances in sequential Monte Carlo", *IEEE Proceedings*, vol. 95(5), pp. 899–924, 2007.

[9] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo and J. Miguez, "Particle filtering", *IEEE Signal Processing Magazine*, vol. 20(5), pp. 19–38, 2003.

[10] F. Gustafsson, "Particle filter theory and practice with positioning applications", *IEEE Aerospace and Electronic Systems Magazine*, vol. 25(7), pp. 53–82, 2010.

[11] A. Doucet and A. M. Johansen, "A Tutorial on Particle Filtering and Smoothing: Fifteen years Later", *Handbook of Nonlinear Filtering (eds. D. Crisan et B. Rozovsky)*, Oxford University Press, 2011.

[12] D. Simon, "Optimal State Estimation", Wiley Interscience, 2006.

[13] N. Caylus, A. Guyader and F. Legland, "Particle filters for partially observed Markov chains", *IEEE workshop on Statistical Signal Processing*, Saint-Louis, 2003.

[14] F. Desbouvries and W. Pieczynski, "Particle filtering with pairwise Markov processes", *IEEE ICASSP*, Hong-Kong, 2003

[15] S. Saha, "Topics in Particle Filtering and Smoothing", *PhD thesis, Univ. of Twente*, 2009, ISBN 978-90-365-2864-1.

[16] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson and P. Nordlund, "Particle Filters for Positioning, Navigation and Tracking", *IEEE Transaction on Signal Processing*, vol. 50(2), Feb. 2002.

[17] R. Chen and J. S. Liu, "Mixture Kalman filters", *Journal of the Royal Statistical Society*, vol. 62(3), pp. 493–508, 2000.

[18] D. Crisan and A. Doucet, "A Survey of Convergence Results on Particle Filtering Methods for Practitioners," *IEEE Transaction on Signal Processing*, vol. 50(3), pp. 736–746, 2002.

[19] T. Schön, F. Gustafsson and P. J. Nordlund, "Marginalized Particle Filters for Mixed Linear/Nonlinear State-Space Models", *IEEE Transactions on Signal Processing*,Vol 53, No. 7, pp. 2279–2289, Jul. 2005.

[20] K. P. Murphy, "Conjugate Bayesian Analysis of the Gaussian distribution", *Technical Report, UBC*, 2007.

[21] F. Gustafsson and S. Saha, "Particle filtering with dependent noises", *13th International Conference on Information Fusion (Fusion 2010)*, Edinburgh, UK.

[22] S. Saha, E. Özkan, F. Gustafsson and V. Smidl, "Marginalized Particle Filters for Bayesian Estimation of Gaussian Noise Parameters", *13th International Conference on Information Fusion (Fusion 2010)*, Edinburgh, UK.

[23] F. Gustafsson, S. Saha and U. Orguner, "The Benefits of Down-Sampling in the Particle Filter", *14th International Conference on Information Fusion (Fusion 2011)*, Chicago, USA.

[24] S. Saha and F. Gustafsson, "Marginalized Particle Filter for Dependent Gaussian Noise Processes", *IEEE Aerospace Conference, 2012*, Montana, USA.

[25] R. Kulhavý and M. B. Zarrop, "On a general concept of forgetting", *International Journal of Control*, Vol 58, No. 4, pp. 905–924, 1993.

[26] M. West and J. Harrison, *Bayesian Forecasting and Dynamic Models*, Springer-Verlag, 1989.

[27] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

Saikat Saha is an Assistant Professor at the Division of Automatic Control, Linköping University, Sweden. He previously held a postdoctoral position in the same division between 2009 and 2011. He received his MSc (Engg) from Indian Institute of Science in 2003 and PhD from University of Twente, the Netherlands in 2009. His research interests include statistical signal processing, information fusion, system identification and computational finance.

Fredrik Gustafsson is professor in Sensor Informatics at Department of Electrical Engineering, Linköping University, since 2005. He received the M.Sc. degree in electrical engineering 1988 and the Ph.D. degree in Automatic Control, 1992, both from Linköping University. During 1992-1999 he held various positions in automatic control, and in 1999 he got a professorship in Communication Systems. In 2004, he was awarded the Arnberg prize by the Royal Swedish Academy of Science (KVA) and in 2007 he was elected member of the Royal Academy of Engineering Sciences (IVA). His research interests are in stochastic signal processing, adaptive filtering and change detection, with applications to communication, vehicular, airborne and audio systems, where the current focus is on sensor fusion algorithms for navigation and tracking problems. He was an associate editor for IEEE Transactions of Signal Processing 2000-2006 and is currently associate editor for EURASIP Journal on Applied Signal Processing and IEEE Transactions on Aerospace and Electronic Systems. He is an IEEE fellow.