Marginalized Particle Filters for Bayesian Estimation of Gaussian Noise Parameters

Saikat Saha, Emre Özkan, Fredrik Gustafsson
Department of Electrical Engineering
Linköping University
Linköping, Sweden
{saha, emre, fredrik}@isy.liu.se

Václav Šmídl
Institute of Information Theory
and Automation
Prague, Czech Republic
smidl@utia.cas.cz

Abstract – *The particle filter provides a general solution to the nonlinear filtering problem with arbitrarily accuracy. However, the curse of dimensionality prevents its application in cases where the state dimensionality is high. Further, estimation of stationary parameters is a known challenge in a particle filter framework. We suggest a marginalization approach for the case of unknown noise distribution parameters that avoid both aforementioned problem. First, the standard approach of augmenting the state vector with sensor offsets and scale factors is avoided, so the state dimension is not increased. Second, the mean and covariance of both process and measurement noises are represented with parametric distributions, whose statistics are updated adaptively and analytically using the concept of conjugate prior distributions. The resulting marginalized particle filter is applied to and illustrated with a standard example from literature.*

Keywords: Unknown Noise Statistics, Adaptive Filtering, Marginalized Particle Filter, Bayesian Conjugate prior

1 Introduction

State space models are widely used in many engineering applications. Depending on the nature of the problem, these models could involve simple linear equations or complex nonlinearities. Estimating the unknown state based on the available measurements is an important and a well studied subject in the literature. Most of the estimation algorithms rely on the prior knowledge of the model and its parameters. In many scenarios, the model parameters, especially the noise/disturbance parameters, might not be known a priori and should be estimated on the run. This problem is referred to as noise adaptive filtering in the literature. The solution is typically given by the joint estimation of noise parameters together with the dynamic state. One very common approach is to augment the state vector with unknown parameters and redefine the problem as a filtering problem. This approach has readily been applied in the particle filtering context [16]. Such an approach has some major disadvantages as it requires artificial dynamics for the static parameters and it leads to an increase in the state dimension which is not preferable for particle filters. Many alternative approaches have also been proposed to circumvent such issues. In [3], the different approaches have been systematically classified into the following catagories: Bayesian, maximum likelihood, correlation and covariance matching. Traditionally the problem has been addressed for linear systems (see e.g., [1],[8]). A correlation based adaptive Kalman filter for noise identification using the weighted least squares criterion has been proposed in [2], while an asymptotic (in time) maximum likelihood estimate has been proposed in [5]. On the other hand, the Bayesian approach has been used, for example, in [6] and [7]. In [6], the nonstationary noise statistics are estimated using the so called IMM method, while an adaptive Kalman filter based on variational Bayesian methods is used in [7]. An adaptive sequential estimation with unknown noise statistics has been proposed in [4]. Estimation of state dependent covariance matrix using the marginalized particle filter approach has been considered by [9]. Here the covariance matrix is treated as additional state, for which a state transition equation has been defined.

In this article, we propose an efficient method in a Bayesian framework for approximating the joint density of the unknown parameters and the state based on the particle filters and marginalization concepts [10],[11]. Analytical substructures in the joint distribution of the state and the model parameters are important in applying the marginalization idea. We assume suitable prior distributions for the unknown noise parameters. Conditional on the particle filter output for the state, we define analytical posterior distribution for the unknown noise parameters and propagate the hyper-parameters of the posterior recursively. Among the previous studies, [14] and [15] are the most related ones to our work. The system considered in [15] is a specific model for a binary output and it is partially linear. The approach in [14] reflects a more general framework but only non-informative prior is used and it aims at estimating the state rather than the unknown parameters. In both studies, the noise sequences are assumed to have zero mean. The methodology we describe in this article defines a more general framework and it is applicable to noise sequences with unknown mean. In [13], estimating the unknown model and the noise parameters is aimed at by using the marginalization idea. Sampling for the unknowns is required in the update steps. In the method we propose, we integrate out the unknown parameters to compute the relevant distributions which results in a more efficient algorithm. Our experiments show that the proposed method is capable of estimating the unknown parameters of the measurement noise as well as the process noise even for highly nonlinear models.

2 Problem Definition

Consider the following nonlinear discrete time state space model relating a hidden state x_t to the observation y_t

$$x_t = f_t(x_{t-1}) + v_t \tag{1}$$

$$y_t = h_t(x_t) + w_t \tag{2}$$

Here t denotes the time index. $f(.)$ and $h(.)$ are possibly nonlinear functions of the state vector x_t. v_t and w_t are mutually independent Gaussian noise sequences with unknown mean and covariances.

$$v_t \overset{i.i.d.}{\sim} \mathcal{N}(\mu_v, \Sigma_v), \tag{3}$$

$$w_t \overset{i.i.d.}{\sim} \mathcal{N}(\mu_w, \Sigma_w). \tag{4}$$

The means and the covariances of the noise sequences are unknown and denoted by θ.

$$\theta \triangleq [\theta^v, \theta^w] \triangleq [\mu_v \Sigma_v, \mu_w \Sigma_w]. \tag{5}$$

The typical problem here is to infer sequentially the unobserved state x_t together with the unknown noise statistics based on a set of observation $y_{0:t}$. This problem appears quite naturally in many practical applications of interests where the exact knowledge of the noises are unavailable. The main difficulty arises from the fact that the estimation of the hidden state also depends on the unknown noise parameters θ. We aim to address this problem by estimating sequentially the joint density of the unknown noise parameters (μ_v, Σ_v, μ_w and Σ_w) and the state sequence $x_{0:t}$ given the set of measurements $y_{0:t}$.

3 Methodology

The method we present here heavily relies on the marginalization concept. We make use of the conjugate priors[1] for the unknown parameters such that, it is sufficient to keep only the hyper-parameters of the posterior distribution for each particle in order to express the joint distribution of the state and the unknown noise parameters. Moreover, within the same approach, it is possible to integrate out the unknown parameters and derive the marginal density for the state easily.

3.1 Posterior distribution for the conjugate prior

For multivariate Normal data with unknown mean μ and covariance Σ, a Normal-inverse-Wishart distribution defines a conjugate prior. Let us denote it as $[\mu_w, \sigma_w] \sim \mathrm{NiW}(k_0, \mu_0, v_0, \Lambda_0)$. Assuming Normal-inverse-Wishart distribution with parameters, $(k_0, \mu_0, v_0, \Lambda_0)$ defines a hierarchical Bayesian model given below:

$$z \sim \mathcal{N}(\mu, \Sigma) \tag{6}$$

$$\mu | \Sigma \sim \mathcal{N}(\mu_0, \frac{\Sigma}{k_0}) \tag{7}$$

$$\Sigma \sim \mathrm{iW}(v_0, \Lambda_0) \tag{8}$$

where iW(.) denotes Inverse Wishart distribution. The joint density of (μ, Σ) is of the form

$$p(\mu, \Sigma) = \mathrm{NiW}(k_0, \mu_0, v_0, \Lambda_0) \tag{9}$$

$$= \frac{1}{c} |\Sigma|^{-((\frac{v_0+d}{2})+1)}$$
$$\times \exp(-\frac{1}{2} tr(\Lambda_0 \Sigma^- 1) - \frac{k_0}{2} (\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0)), \tag{10}$$

where

$$c = \frac{2^{v_0 d/2} \Gamma_d(v_0/2)(2\pi/k_0)^{d/2}}{|\Lambda_0|^{v_0/2}}. \tag{11}$$

The parameters μ_0 and k_0 define the prior mean and the number of prior measurements, while v_0 and Λ_0 define the degrees of freedom and the scale matrix for the inverse-Wishart distribution. Notice that the mean and the covariance are dependent. A larger covariance results in a larger variance on μ whereas a smaller covariance will pull the mean towards μ_0. Further, the Gaussian distribution for z becomes a t-distribution for a NiW prior,

$$p(z|\mu, v, \Lambda) = t_{v-d+1}(\mu, \frac{(k+1)}{k(v-d+1)} \Lambda) \tag{12}$$

3.2 Recursive updates of the conjugate prior

Suppose we observe a set of m observations, $\{z_t\}_{t=1}^m$ from a multivariate Gaussian distribution for which

[1]A family of prior distributions is conjugate to a particular likelihood function if the posterior distribution belongs to the same family as the prior.

we assumed a normal-inverse-Wishart prior for the unknown mean and variance. Via conjugacy, the posterior distribution of the unknown parameters is again a normal-inverse-Wishart distribution with the updated hyper-parameters. The hyper-parameters of the posterior distribution are updated as follows [12],

$$\mu_{m+1} = \frac{k_0}{k_0+m}\mu_0 + \frac{m}{k_0+m}\bar{z}_m \tag{13a}$$

$$\Lambda_{m+1} = v_0\Lambda_0 + S_m + \frac{k_0 m}{k_0+m}(\bar{z}_m-\mu_0)(\bar{z}_m-\mu_0)^T \tag{13b}$$

$$k_{m+1} = k_0 + m \tag{13c}$$

$$v_{m+1} = v_0 + m \tag{13d}$$

where

$$\bar{z}_m = \frac{1}{m}\sum_{t=1}^{m} z_t, \tag{13e}$$

$$S_m = \sum_{t=1}^{m}(z_t-\bar{z})(z_t-\bar{z})^T. \tag{13f}$$

3.3 Marginalization in nonlinear filtering

Let us define NiW priors for the unknown process noise and the measurement noise sequences of the system defined by (1) and (2). Let $\Phi_0 = [\phi_0^w \phi_0^v]$ denote the initial hyper-parameters describing the prior distributions ($\phi_0^w = [(k_0^w, \mu_0^w, v_0^w, \Lambda_0^w)]$ for the process noise and $\phi_0^v = [(k_0^v, \mu_0^v, v_0^v, \Lambda_0^v)]$ for the measurement noise). Our aim is to approximate the joint density for $p(x_{0:t}, \theta|y_{0:t})$ and allow marginalization if possible. The joint distribution of the states and the unknown parameters can be decomposed into conditional distributions:

$$p(x_{0:t}, \theta|y_{0:t}) = p(\theta|x_{0:t}, y_{0:t})p(x_{0:t}|y_{0:t}). \tag{14}$$

Suppose we approximate the distribution $p(x_{0:t}|y_{0:t})$ by a set of N particles and their weights as

$$p(x_{0:t}|y_{0:t}) \simeq \sum_{i=1}^{N}\omega_t^{(i)}\delta_{x_{0:t}^{(i)}}(.). \tag{15}$$

For each particle we can compute analytical expressions for the posterior distribution of the unknown parameters. Notice that given the state trajectory $x_{0:t}$ and the measurements $y_{0:t}$, the measurement and process noise parameters become independent. Hence the hyper-parameter update for the posterior distributions can be done separately. The posteriors follow normal-inverse-Wishart distribution and the hyper-parameters are updated according to the equations (13a)-(13d). The following expressions are substituted with the pseudo measurement z_t in the update equations.

$$z_t^v \triangleq x_t^{(i)} - f_t(x_{t-1}^{(i)}) \quad \text{for the process noise update} \tag{16}$$

$$z_t^w \triangleq y_t - h_t(x_t^{(i)}) \quad \text{for the measurement noise update.} \tag{17}$$

Using the sequential importance sampling scheme for propagating the particle approximation (15) leads to the standard weight update equation:

$$\omega_t^{(i)} = \omega_{t-1}^{(i)}\frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{t-1}^{(i)}, y_t)}, \tag{18}$$

where $q(.)$ is the importance distribution from which we sample $x_t^{(i)}$.

3.4 Likelihood marginalization

In order to compute the likelihood $p(y_t|x_t^{(i)})$, we can utilize the posterior distribution of the unknown parameters that we computed for each particle. One important advantage of using conjugate priors reveals itself here as it is possible to integrate out unknown noise parameters as they follow normal-inverse-Wishart distribution.

$$p(y_t|x_t) = \int p(y_t|\theta^v, x_{0:t})p(\theta^v|x_{0:t})d\theta^v. \tag{19}$$

In accordance with the notations described in equations (13a)-(13d), the resulting predictive distribution is a multivariate Student-t distribution as follows from (12),

$$p(z_t|z_{1:t-1}, k, \mu, v, \Lambda) = t_{v_t-d+1}(\mu_t, \frac{(k_t+1)}{k_t(v_t-d+1)}\Lambda_t) \tag{20}$$

where $t_v(\mu, \lambda)$ is the student-t distribution with v degrees of freedom, located at μ with scale parameter λ. The likelihood can be computed using the above expression together with (17).

3.5 State prediction

In most of the cases it is not possible to sample from the optimal importance distribution. The state transition density $p(x_t|x_{t-1})$ can be used as the importance distribution. Once again the unknown process noise can be integrated out.

$$p(x_t|x_{0:t-1}) = \int p(x_t|\theta^w, x_{0:t-1})p(\theta^w|x_{0:t-1})d\theta^w. \tag{21}$$

The resulting predictive distribution is a multivariate Student-t distribution similar to (12).

3.6 Posterior distribution for the noise parameters

The marginal posterior density of the unknown parameters can be computed by integrating out the states in the joint density.

$$\begin{aligned} p(\theta|y_{1:t}) &= \int p(\theta|x_{0:t}, y_{1:t}) p(x_{0:t}|y_{1:t}) dx_{0:t} \\ &\approx \sum_{i=1}^{N} \omega_t^{(i)} p(\theta|x_{0:t}^{(i)}, y_{1:t}). \end{aligned} \quad (22)$$

Then the estimate of the unknown parameters could be computed according to a chosen criterion. As an example, according to minimum mean square error (MMSE) criterion, the noise variance estimate at time t could be computed as

$$\widehat{\Sigma_t} = \sum_{i=1}^{N} \omega_t^{(i)} \frac{\Lambda_t^{(i)}}{v_t - d - 1}, \quad (23)$$

where the weights are inherited from the particles.

3.7 Marginalized particle filter

In the proposed method, each particle keeps its own estimate for the parameters of the unknown process noises and measurement noise. In the importance sampling step, the particles use their own posterior distribution of the unknown parameters. The weight update of the particles is made according to the measurement likelihood. It is our expectation that the particles, keeping the unknown parameters which best explains/fits to the observed measurement sequence will survive in time. The methodology followed here is described in the next paragraph as a pseudo code.

Algorithm:

- <u>`Initialization:`</u>
- For each particle $i = 1, .., N$ do
  - Sample $x_0^{(i)} \sim p_0(x_0)$
  - Set initial weights $\omega_0^{(i)} = \frac{1}{N}$
  - Set initial noise hyper-parameters $[\phi_0^w \phi_0^v]$ corresponding to each particle
- <u>`Iterations:`</u>
- For $t = 1, 2, \ldots$ do
  - For each particle $i = 1, .., N$ do
    - sample $x_t^{(i)} \sim q(x_t^{(i)}|y_t, x_{t-1}^{(i)})$
  - For $i = 1, .., N$, update the weights

    $$\widetilde{\omega}_t^{(i)} = \omega_{t-1}^{(i)} \frac{p(y_t|x_t^{(i)}) p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{t-1}^{(i)}, y_t)}$$
  - Update hyper-parameters of the process noise, using the pseudo measurement $z_t^v = x_t^{(i)} - f_t(x_{t-1}^{(i)})$ (Equations (13a)-(13d)).
  - Update hyper-parameters of the measurement noise, using the pseudo measurement $z_t^w = y_t - g_t(x_t^{(i)})$ (Equations (13a)-(13d)).
  - Normalize weights, $\omega_t^{(i)} = \frac{\widetilde{w}_t^{(i)}}{\sum_{i=1}^{N} \widetilde{w}_t^{(i)}}$.
  - Compute $N_{\text{eff}} = \frac{1}{\sum_{i=1}^{N} (\omega_t^{(i)})^2}$.
    - If $N_{\text{eff}} \leq \eta$, Resample the particles. Copy the corresponding hyperparameters and set $\omega_t^{(i)} = 1/N$.

4 Simulations

We use the following benchmark scalar nonlinear time series model for our illustrations:

$$x_t = \frac{x_{t-1}}{2} + \frac{25x_{t-1}}{1 + x_{t-1}^2} + 8\cos(1.2t) + v_t, \quad (24)$$

$$y_t = \frac{x_t^2}{20} + w_t, \quad v_t \perp w_t,\ t = 1, 2, \ldots \quad (25)$$

where $v_t \sim N(0, \Sigma_v)$ and $w_t \sim N(0, \Sigma_w)$ and both Σ_v and Σ_w are unknown. For simulated data, we use $\Sigma_v = 10$ and $\Sigma_w = 1$.

In the first subsection, we assume that only the variances of the noises are unknown, whereas in the second subsection, we consider both the means and the variances to be unknown.

4.1 Unknown Variances

We note that when the mean of the Gaussian noise is known, the conjugate prior for the covariance matrix is given by Inverse Wishart distribution (iW), which for the scalar case reduces to Inverse Gamma distribution (iG) [12]. At time step $t = 0$, we set the priors as $p(x_0) = N(0, 5)$, $p(\Sigma_v) = \text{iG}(\alpha_0, \beta_0)$ and $p(\Sigma_w) = \text{iG}(\lambda_0, \delta_0)$. We take $\alpha_0 = 2.01$, $\beta_0 = 5$, $\lambda_0 = 2.01$ and $\delta_0 = 5$. The hyper-parameters update equations for the iG are shown in the appendix A. In the particle filtering step, the state transition density is taken as the proposal and the resampling is done whenever the effective sample size falls below the one-third of the original sample size. The expressions for the mean and variance of the posterior are shown in appendix B. Realizations of the estimates of Σ_v and Σ_w with particle size $N = 500$ and $N = 5000$ are respectively shown in Figures 1–2. We observe that the estimation procedure works quite well. Next, keeping the particle size $N = 5000$, we repeat the estimates over 100 Monte Carlo runs. The Monte Carlo average of the estimates of Σ_v and Σ_w are shown in Figures 3–4. Here, the

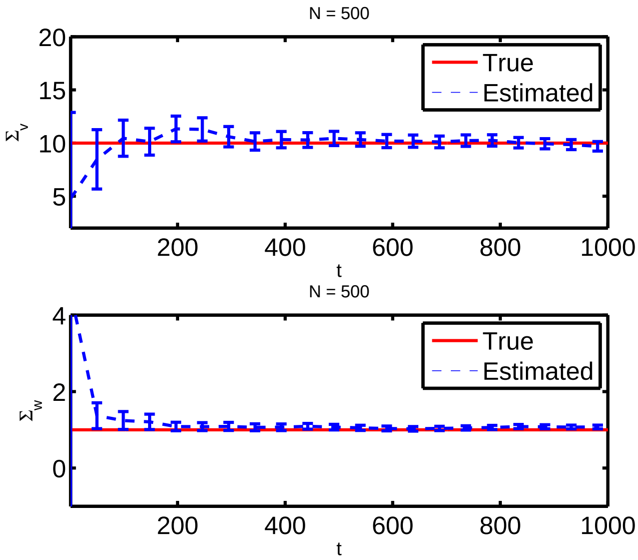

Figure 1: Posterior estimate of Σ_v and Σ_w visualized via mean value and two standard deviation bounds.

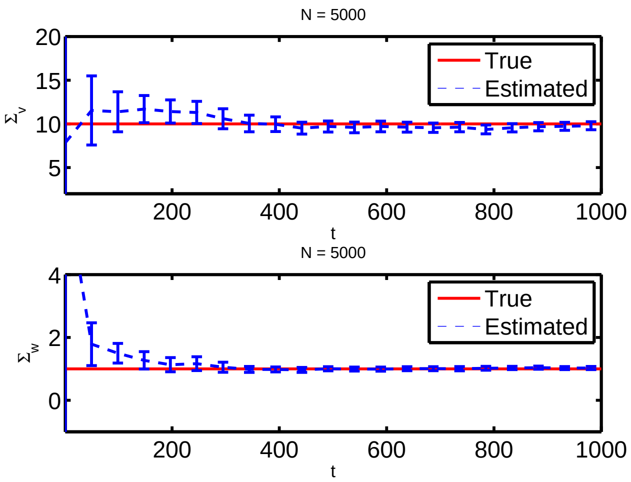

Figure 2: Posterior estimate of Σ_v and Σ_w visualized via mean value and two standard deviation bounds.

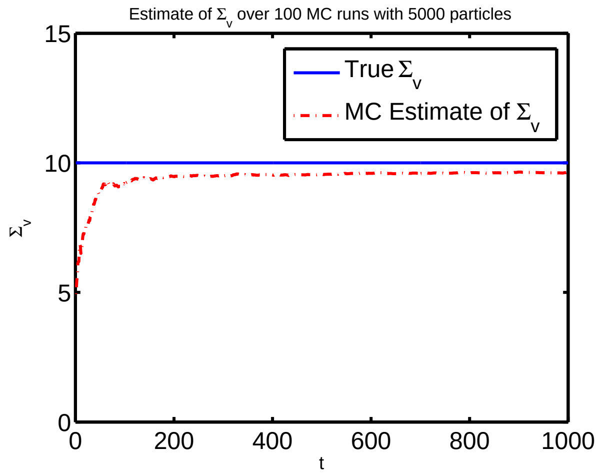

Figure 3: Estimated Σ_v with 5000 particles over 100 Monte Carlo runs

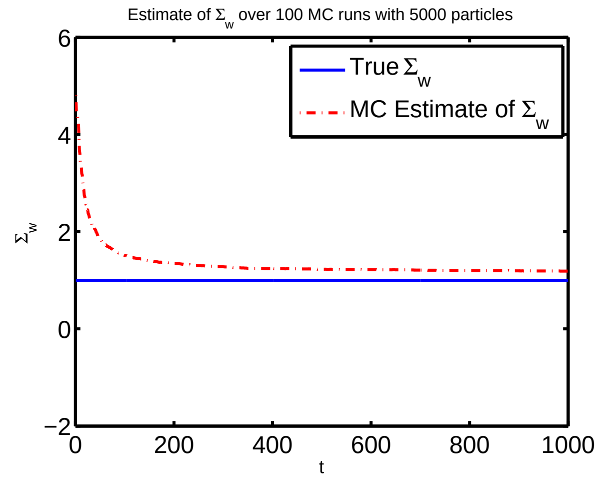

Figure 4: Estimated Σ_w with 5000 particles over 100 Monte Carlo runs

estimates appear to be slightly biased. We also compute the root mean squared error (RMSE) estimates of Σ_v and Σ_w at each time step t (over $M = 100$ Monte Carlo runs) given by $\left(\frac{1}{M}\sum_{j=1}^{M}(\hat{z}_t^j - z_t^j)^2\right)^{\frac{1}{2}}$. Here z_t^j is the true parameter for time t in the j'th run and $\hat{z}_t^j$ is the corresponding estimate. The results are shown in Figures 5–6. Next, for a typical realization with 5000 particles, we also plot the posterior densities $p(\Sigma_v|y_{1:T})$ and $p(\Sigma_w|y_{1:T})$ at final time $T = 1000$. This is shown in Figure 7. Subsequently, we compute the mean and the maximum *a posteriori* (MAP) estimates of both Σ_v and Σ_w at final time step $T = 1000$. Here, MAP is obtained by maximizing the argument of the respective posterior density. Each posterior is given by the weighted mixture of inverse Gamma densities. We maximize the posterior numerically using Matlab's `fminunc` command, with the starting value taken to be the corresponding

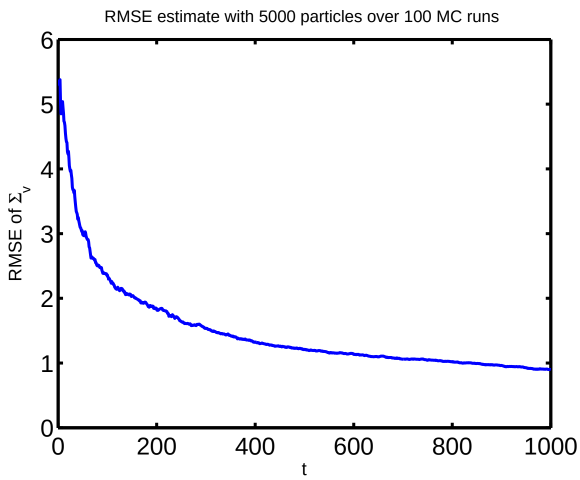

Figure 5: RMSE of Σ_v with 5000 particles over 100 Monte Carlo runs

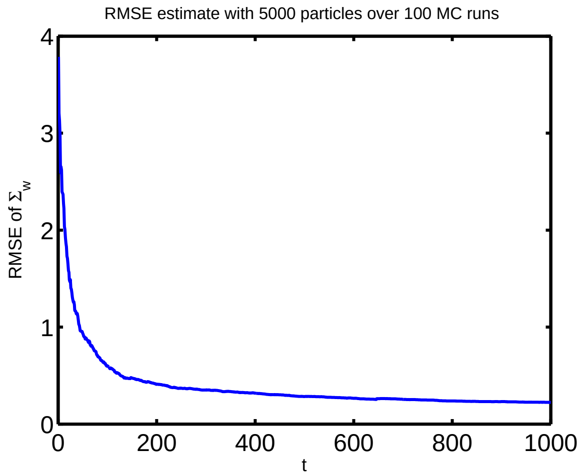

Figure 6: RMSE of Σ_w with 5000 particles over 100 Monte Carlo runs

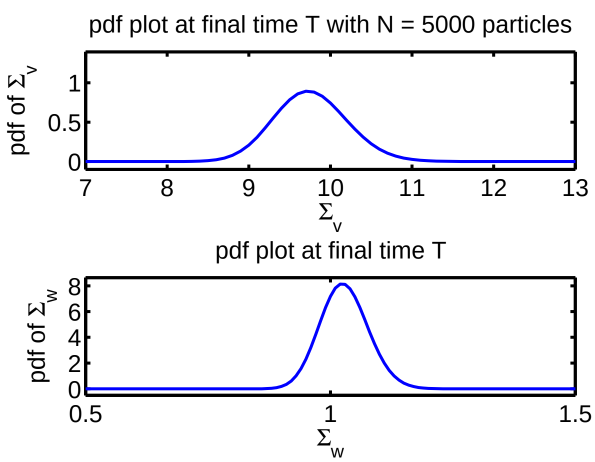

Figure 7: posterior densities of Σ_v and Σ_w

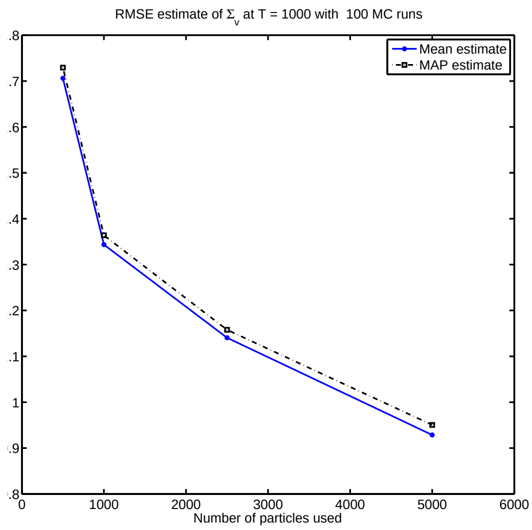

Figure 8: RMSE of Σ_v at final time step versus number of particles

mean estimate. The RMSE estimates over 100 Monte Carlo runs for different particle sizes are shown in Figures 8–9. We observe that the mean estimate performs better for Σ_v while both MAP and mean estimates are similar for Σ_w.

4.2 Unknown Means and Variances

Here, we investigate the case where both the mean and the variance of the noise sequences are unknown. The same system defined by the equations (24)-(25) is used once more. The true parameters of the noises are set to: $\mu_v = 3$, $\sigma_v^2 = 4$, $\mu_w = 1$, $\sigma_w^2 = 6$. NiW distribution is used as the prior and the initial hyper-parameters are set to $\phi_0^w = [(k_0^w, \mu_0^w, v_0^w, \Lambda_0^w)] = [(5, 0, 5, 10)]$ for the process noise and $\phi_0^v = [(k_0^v, \mu_0^v, v_0^v, \Lambda_0^v)] = [(5, 0, 5, 10)]$ for the measurement noise. In Figure 10, the estimates for the measurement and the process noise covariances and the means are depicted together.

5 Conclusions and Discussions

A new method for estimation of unknown noise parameters in general state space models is presented in this article. The method is defined in Bayesian framework where we define conjugate priors for the unknown noise parameters. We also make use of the marginalization idea in order to keep the algorithm implementation simple and efficient, with analytic posterior distributions for the noise parameters. The methodology described here is generic and can be extended and generalized in several ways: (i) A larger class of noises from the exponential family with suitable conjugate priors can be used. (ii) The independence assumption on the process noise and the measurement noise sequences can be relaxed. (iii) The principle of exponential forgetting

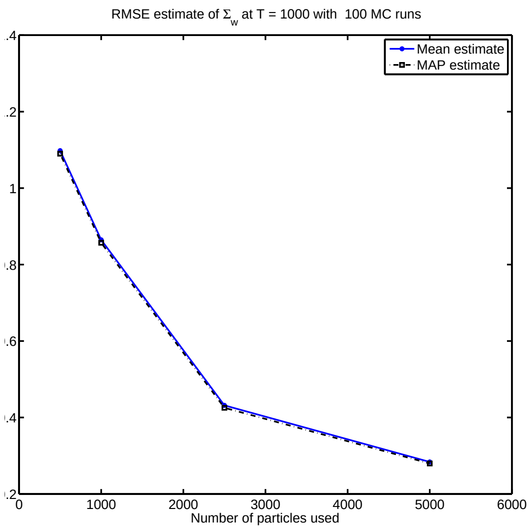

Figure 9: RMSE of Σ_w at final time step versus number of particles

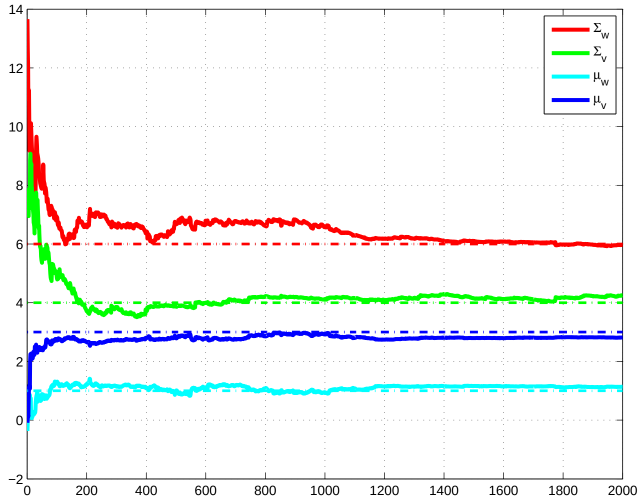

Figure 10: Estimated mean and covariance for the measurement and the process noises. The algorithm is run with 5000 particles.

can be applied to allow for time varying noise characteristics. Such modifications are left as future work.

6 Acknowledgement

Both S. Saha and E. Özkan are supported by grants from the Linnaeus Center for Control, Autonomy, and Decision-making in Complex Systems (CADICS).

A Updating hyper-parameters of Inverse Gamma distribution

Suppose the prior $p(\Sigma) \sim \text{iG}(a_0, b_0)$. Note that iG is a special case of iW for scalar variables. If the observations $X = [x_1, \ldots, x_n]$ are independent Gaussian variables drawn from $N(\mu, \Sigma)$ distribution, then conjugacy implies that the posterior distribution $p(\Sigma|X)$ is also inverse Gamma distributon, i.e. $p(\Sigma|X) \sim \text{iG}(a_n, b_n)$, where the hyper parameters can be updated as

$$a_n = a_0 + \frac{n}{2} = a_{n-1} + \frac{1}{2}, \tag{26a}$$

$$b_n = b_0 + \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2 = b_{n-1} + \frac{(x_n - \mu)^2}{2}. \tag{26b}$$

B Mean and variance for the posterior of Σ

The posterior of Σ at time step t is computed as

$$\begin{aligned} p(\Sigma|y_{1:t}) &= \int p(\Sigma|x_{0:t}, y_{1:t})p(x_{0:t}|y_{1:t})dx_{0:t} \\ &\approx \sum_{i=1}^{N} p(\Sigma|x_{0:t}^{(i)}, y_{1:t})\omega_t^{(i)}, \end{aligned} \tag{27}$$

where $p(\Sigma|x_{0:t}^{(i)}, y_{1:t})$ is $\text{iG}(a_t^{(i)}, b_t^{(i)})$ with mean and variance as

$$\text{E}(\Sigma|x_{0:t}^{(i)}, y_{1:t}) = \frac{b_t^{(i)}}{(a_t^{(i)} - 1)} \quad (for \quad a_t^{(i)} > 1) \tag{28}$$

$$\text{Var}(\Sigma|x_{0:t}^{(i)}, y_{1:t}) = \frac{(b_t^{(i)})^2}{(a_t^{(i)} - 1)^2(a_t^{(i)} - 2)} \quad (for \quad a_t^{(i)} > 2). \tag{29}$$

Then the mean and variance for the posterior of Σ are given by

$$\text{E}(\Sigma|y_{1:t}) \approx \sum_{i=1}^{N} \text{E}(\Sigma|x_{0:t}^{(i)}, y_{1:t})\omega_t^{(i)} \tag{30}$$

$$\begin{aligned} \text{Var}(\Sigma|y_{1:t}) \approx & \sum_{i=1}^{N} \omega_t^{(i)} \Big\{ \text{Var}(\Sigma|x_{0:t}^{(i)}, y_{1:t}) + \\ & + \ \big\{\text{E}(\Sigma|x_{0:t}^{(i)}, y_{1:t}) - \text{E}(\Sigma|y_{1:t})\big\}^2 \Big\} \end{aligned} \tag{31}$$

References

[1] S. Kosanam, and D. Simon, "Kalman Filtering with Uncertain Noise Covariances," *Intelligent Systems and Control*, Honolulu, HI, Aug. 2004, pp. 375–379.

[2] M. Oussalah, and J. De Schutter, "Adaptive Kalman filter for noise identification," *25th International Conference on Noise and Vibration Engineering*, Leuven, Belgium, Sept. 2000, pp. 1225–1232.

[3] R. Mehra,"On the identification of variances and adaptive Kalman filtering," *IEEE Transactions on*

Automatic Control,Vol 17, No. 5, pp. 693–698, 1972.

[4] K. Myers, and B. Tapley,"Adaptive sequential estimation with unknown noise statistics,"*IEEE Transactions on Automatic Control*,Vol 21, No. 4, pp. 520–523, 1976.

[5] R. E. Maine, and K. W. Iliff,"Formulation and implementation of a practical algorithm for parameter estimation with process and measurement noise,"*SIAM Journal of Applied Mathematics*,Vol 41, No. 3, pp. 558–579, 1981.

[6] X. R. Li, and Y. Bar-Shalom,"A recursive multiple model approach to noise identification,"*IEEE Transactions on Aerospace and Electronic Systems*,Vol 30, No. 3, 1994.

[7] S. Särkkä, and A. Nummenmaa,"Recursive Noise Adaptive Kalman Filtering by Variational Bayesian Approximations,"*IEEE Transactions on Automatic Control*,Vol 54, No. 3, Mar. 2009.

[8] Y. Liang, D. An, D. Zhou, and Q. Pan,"A finite-horizon adaptive Kalman filter for linear systems with unknown disturbances ,"*Signal Processing*,Vol 84, No. 11, pp. 2175–2194, 2004.

[9] V. Šmidl,"On Estimation of Unknown Disturbances of Non-Linear State-Space Model Using Marginalized Particle Filter,"*Research Report*, No. 2245, Dec. 2008.

[10] 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.

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

[12] A. Gelman and J.B. Carlin and H.S. Stern and D.B. Rubin, *Bayesian data analysis*, CRC Press 2004.

[13] G. Storvik, "Particle filters for state-space models with the presence of unknown static parameters", *IEEE Transactions on Signal Processing*, Vol 50, No. 2, pp. 281–289, Feb. 2002.

[14] P.M. Djuric, J. Miguez,"Sequential particle filtering in the presence of additive Gaussian noise with unknown parameters", *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2002* Vol 2, pp. 1621–1624.

[15] C.J. Bordin, M.G.S. Bruno, "Bayesian blind equalization of time-varying frequency-selective channels subject to unknown variance noise" *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2008* pp. 3449–3452.

[16] J. Liu, M. West,"Combined parameter and state estimation in simulation-based filtering", in A. Doucet and N. De Freitas and N. Gordon editors *Sequential Monte Carlo Methods in Practice*, Springer-Verlag 2001.