RESEARCH REPORT

Václav Šmídl

On Estimation of Unknown Disturbances of Non-Linear State-Space Model Using Marginalized Particle Filter

No. 2245 December 21, 2008

1 Introduction

This work has been motivated by the need for improvement of Bayesian filtering for sensor-less control of electric drives. This is a traditional area of application of Bayesian filtering, namely Extended Kalman Filter [1], [2, 3, 4, 5]. All current implementation are based on the fact that covariance matrices of the disturbances are known. This assumption is not met in this case and the covariances are often manually tuned for optimal performance [6]. Tuning of the EKF is a complex process usually discouraging the researchers from using this filter and strongly limiting its practical applications.

The idea of this report is to consider unknown variances as additional parameters and estimate their values. The problem of unknown disturbances of the state-space model has been addressed using many approaches. One important approach is based on the use of marginalized particle filter [7]. In this paper, we study their properties and suitability for sensor-less control.

1.1 State of the art

Severeness of the assumption of unknown covariance of disturbances in a state-space model has been realized very early [8]. Traditionally, the problem is solved for linear systems where many dedicated methods have been developed [9]. Methods for non-linear systems are also available [10], however, these require computationally heavy off-line processing are thus are unsuitable for an exploratory analysis of an unknown system.

Another common assumption is that the we estimate time invariant full matrix Q of the system. In this case, the derivation is rather complicated, however, proofs of convergence are available [11]. This approach is not suitable for our problem since the covariance matrices are assumed to be state-dependent.

More suitable approach is reformulation of the problem into the form of Bayesian filtering. The unknown covariance matrices are considered as a new state-variable for which a state-transition equation needs to be defined. This is an early formulation of the problem [8], which can now be fully explored with the availability of approximate Bayesian inference via particle filters [12]. The drawback of particle filtering is the need to sample from potentially large space. This can be remedied by using technique of Rao-Blackwellization [13], also known as marginalized particle filter [14]. We focus in this approach, which is shortly reviewed in the next Section.

2 Bayesian Filtering

Consider a sequence of observation models in the following manner:

$$d_t \sim f\left(d_t | x_t\right), \quad x_t \sim f\left(x_t | x_{t-1}\right). \tag{1}$$

Here, x_t is a vector known as the state variable and d_t are the observations. The first density in (1) is known as the *observation model*, the second is known as the *evolution model*. By *Bayesian Filtering*, we mean the recursive evaluation of the filtering

distribution, $f(x_t|D_t)$, using Bayes' rule [15, 13]:

$$f(x_t|D_{t-1}) = \int f(x_t|x_{t-1}) f(x_{t-1}|D_{t-1}) \, dx_{t-1}, \tag{2}$$

$$f(x_t|D_t) \propto f(d_t|x_t) f(x_t|D_{t-1}), \quad t = 1, 2, ... \tag{3}$$

where $f(\theta_1)$ is the prior distribution, and $D_t = [d_1, \ldots, d_t]$ denotes the set of all observations.

Bayesian filtering (BF) is analytically tractable if (i) marginalization over x_{t-1} is analytically tractable, and (ii) the resulting marginal distribution, $f(x_t|D_t)$, is in the same form as the previous step, $f(x_{t-1}|D_{t-1})$, allowing the procedure to be iterated. (i) and (ii) are satisfied in only a very limited class of models [16], most notable example of which is the Kalman filter [17]. For models outside of the tractable class, exact inference is computationally prohibitive and must be replaced by approximation. Two important approximation techniques are now shortly reviewed.

2.1 Extended Kalman Filter

Consider a non-linear Gaussian state-space model

$$f(x_t|x_{t-1}) = \mathcal{N}(g(x_{t-1}), Q), \tag{4}$$

$$f(d_t|x_t) = \mathcal{N}(h(x_{t-1}), R),$$

where $g(\cdot)$ and $h(\cdot)$ are vector functions of appropriate dimensions, $\mathcal{N}(\mu, R)$ denotes Gaussian probability density with mean value μ and variance R. Due to non-linear functions $g()$ and $h()$, operations (2)–(3) do not yield posterior density in the form of a Gaussian. However, it can be approximated for by a Gaussian distributed density of the parameter x_{t-1},

$$f(x_{t-1}|D_{t-1}) = \mathcal{N}(\hat{x}_{t-1}, P_{t-1|t-1}),$$

using Taylor expansion at the current point estimate, $\hat{\theta}_{t-1}$ as follows:

$$f(x_t|D_{t-1}) \approx \mathcal{N}(g(\hat{x}_{t-1}), P_t),$$

$$f(x_t|D_t) \approx \mathcal{N}(g(\hat{x}_{t-1}) - K(d_t - h(\hat{x}_{t-1})), P_{t|t}). \tag{5}$$

where $A = \frac{d}{dx_t} g(x_t)$, $C = \frac{d}{dx_t} h(x_t)$ and K, P_t are the Kalman gain and posterior covariance matrix well known from the Kalman filter.

$$R_y = C'P_{t-1}C + R,$$

$$K = P_{t-1}CR_y^{-1}$$

$$P_{t|t} = P_{t-1} - P_{t-1}C'R_y^{-1}CP_{t-1}, \tag{6}$$

$$P_t = AP_{t|t}A + Q.$$

The observation likelihood is then:

$$f(d_t|D_{t-1}) \approx \mathcal{N}(h(\hat{x}_{t-1}), R_y). \tag{7}$$

This approximation is performed in each step as is typical for local approximations which are known to diverge with time [18].

2.2 Particle filtering

Particle filtering (PF) [13] refers to a range of techniques for generating an empirical approximation of $f(X_t|D_t)$, where $X_t = [\theta_1, \ldots, x_t]$ is the state trajectory:

$$f(X_t|D_t) \approx f_\delta(X_t|D_t) = \frac{1}{n} \sum_{i=1}^{n} \delta\left(X_t - X_t^{(i)}\right), \qquad (8)$$

where $X_t^{(i)}$, $i = 1, \ldots, n$ are i.i.d. samples from the posterior and $\delta(\cdot)$ denotes the Dirac δ-function. We reserve the symbol $f_\delta(\cdot)$ for the empirical distribution. Therefore, this approach is feasible only if the we can sample from the exact posterior, $f(X_t|D_t)$. If this is not the case, we can draw samples from a chosen proposal distribution (importance function), $q(X_t|D_t)$, as follows:

$$f(X_t|D_t) \approx \frac{f(X_t|D_t)}{q(X_t|D_t)} \frac{1}{n} \sum_{i=1}^{n} \delta\left(X_t - X_t^{(i)}\right). \qquad (9)$$

Using the sifting property of the Dirac δ-function, the approximation can be written in the form of a *weighted* empirical distribution, as follows:

$$f_\delta(X_t|D_t) \approx \sum_{i=1}^{n} w_t^{(i)} \delta\left(X_t - X_t^{(i)}\right), \qquad (10)$$

$$w_t^{(i)} \propto \frac{f\left(X_t^{(i)}|D_t\right)}{q\left(X_t^{(i)}|D_t\right)}. \qquad (11)$$

Under this *importance sampling* procedure, the true posterior distribution need only be evaluated point-wise. Furthermore, normalizing constant of $f(\cdot)$ is not required, since (10) can be normalized trivially via a constant $c = \sum_{i=1}^{n} w_t^{(i)}$. Weights (11) may be written in the following recursive form:

$$w_t^{(i)} \propto \frac{f\left(d_t|x_t^{(i)}\right) f\left(x_t^{(i)}|x_{t-1}^{(i)}\right)}{q\left(x_t^{(i)}|X_{t-1}^{(i)}, D_t\right)} w_{t-1}^{(i)}. \qquad (12)$$

where, now, $x_t^{(i)}$ are drawn from the denominator of (12), which can be chosen as $f(x_t|x_{t-1})$. Thus, the form of the posterior distribution is preserved by Bays rule as is characteristic for global approximation, [18]. Successful application of the particle filter requires more steps—such as re-sampling and appropriate choice of the importance function—which are beyond scope of this paper. See [13] for more details.

2.3 Marginalized Particle Filtering (MPF)

The main advantage of importance sampling is its generality. However, it may be computationally prohibitive to draw samples from the possibly high dimensional state space

of x_t. Furthermore, it is necessary to generate large numbers of such particles in these cases in order to achieve an acceptable error of approximation. These problems can be overcome in cases where the structure of the model (1) allows analytical marginalization over a subset, $\overline{x}_t$, of the full state vector $x_t' = [\overline{x}_t', z_t']$ [13, 14, 7]. Therefore, we consider the factorization

$$f(X_t|D_t) = f\left(\overline{X}_t|Z_t, D_t\right) f(Z_t|D_t), \tag{13}$$

where $f\left(\overline{X}_t|Z_t, D_t\right)$ is analytically tractable, while $f(Z_t|D_t)$ is not. We replace the latter by a weighted empirical distribution, in analogy to (9), yielding

$$f(X_t|D_t) \approx \sum_{i=1}^{n} w_t^{(i)} f\left(\overline{X}_t|Z_t^{(i)}, D_t\right) \delta\left(Z_t - Z_t^{(i)}\right), \tag{14}$$

$$w_t^{(i)} \propto \frac{f\left(Z_t^{(i)}|D_t\right)}{q\left(Z_t^{(i)}|D_t\right)}. \tag{15}$$

Note that we now only have to sample from the space of z_t. The weights can, once again, be evaluated recursively:

$$w_t^{(i)} \propto \frac{f\left(d_t|z_t^{(i)}\right) f\left(z_t^{(i)}|z_{t-1}^{(i)}\right)}{q\left(z_t|Z_{t-1}^{(i)}, D_t\right)} w_{t-1}^{(i)}. \tag{16}$$

Hence, the model (1) must admit a partition, $[\overline{x}_t, z_t]$, for which $\overline{x}_{t-1}$ can be integrated analytically in (2) and the resulting $f(\overline{x}_t|Z_t, D_{t-1})$ is of the same form as in the previous step. Then, the marginalized particle filter (14)–(16) can be evaluated exactly. This requirement is always fulfilled when the model can be decomposed in linear and non-linear parts [14], and possibly even for wider class [16]. (14)–(16) is sometimes referred to as the Rao-Blackwellized particle filter [13].

3 MPF for Models with Unknown Covariance Matrix

Note that formulation (13)–(15) readily offers a way how to estimate covariance matrices for linear state-space models with time-invariant covariance matrices, i.e. linear $g()$ and $h()$ in (4). The original state x_t becomes $\overline{x}_t$ and the unknown covariance matrices, Q and R, form the extension $z_t = [\text{vec}(Q_t)', \text{vec}(R_t)']'$. The state space model (4) has to be extended by an evolution model on the covariance matrices, $f(z_t|z_{t-1})$. Then, the resulting filter acts as a bank of n Kalman filters, each with different covariance matrix, $z_t^{(i)} == [\text{vec}(Q_t^{(i)})', \text{vec}(R_t^{(i)})']'$.

If the MPF formulation is followed strictly for a non-linear model (4), the nonlinear part of $\overline{x}_t$ (in the sense of [14]) would have to be estimated using empirical approximation as well. This would complicate future use of the result. Therefore, we use formulation of [7], where MPF arise as an approximation of term $f(Z_t|D_t)$ in (13) by an empirical density. In the same spirit, we can replace $f(X_t|Z_t, D_t)$ by a suitable approximation, $\tilde{f}(X_t|Z_t, D_t)$, for example that of the EKF (4). The resulting posterior density is then:

$$f\left(\overline{x}_t, z_t | D_t\right) \approx \sum_{i=1}^{n} w_i \tilde{f}\left(\overline{x}_t | z_t^{(i)}, D_t\right) \delta(z_t^{(i)} - z_t). \tag{17}$$

3.1 Scalar variance

In one-dimensional case, the covariance matrix is just a non-negative scalar. Note that posterior density of unknown variance of a Normal model is inverse-Gamma distributed [15]. Therefore, we propose to model evolution of scalar variance, for example $f(r_t | r_{t-1})$, by an inverse-Gamma random walk

$$f(r_t | r_{t-1}) = i\mathcal{G}(a, b) = \frac{b^a}{\Gamma(a)} r_t^{-a-1} \exp(-\frac{b}{r_t}), \tag{18}$$

where $a > 0, b > 0$ are shaping parameters of the density, and $\Gamma(a)$ is the Gamma function [19]. The walk will be defined via moments of (18), as follows:

$$\mathsf{E}(r_t) = \frac{b}{a-1} \equiv r_{t-1}, \quad \mathsf{std}[r_t] = \frac{b}{(a-1)\sqrt{a-2}} \equiv k r_{t-1}. \tag{19}$$

The first choice is typical for a random walk, the second choice makes standard deviation of the process proportional to its mean value by a chosen constant of proportionality k. Low values of k allows only small variations of the process, higher values of k model rapid variations of the variance. Substituting choices (19) into (18) we get:

$$f(r_t | r_{t-1}) = i\mathcal{G}\left(\frac{1}{k^2} + 2, \ r_{t-1}\left(\frac{1}{k^2} + 1\right)\right). \tag{20}$$

Notes:

- r_t can be sampled using Gamma sampler, using relation $1/r_t \sim \mathcal{G}(a, b)$.

- Random walk of this type may yield unstable results. Stabilization can be achieved using geometric combination of r_{t-1} and some reference point r_{ref}

$$\mathsf{E}(r_t) = \frac{b}{a-1} \equiv r_{t-1}^l r_{\text{ref}}^{1-l}, \tag{21}$$

where $l \in <0, 1>$ is a chosen weight of the geometric combination.

Remark 1 (Log-normal random walk). Alternatively, we may sample r_t from a log-Normal density $r_t \sim \log \mathcal{N}(\mu, \sigma)$. Using conditions (19) with substituted moments of the log-Normal density, parameters μ, σ must satisfy:

$$\exp(\mu + \frac{1}{2}\sigma^2) = r_{t-1}, \ \sqrt{e^{\sigma^2} - 1} \exp(\mu + \frac{1}{2}\sigma^2) = k r_{t-1}.$$

Hence, $\sigma^2 = \ln(k^2 + 1)$, $\mu = \ln(r_{t-1}) - \frac{1}{2}\sigma^2$.

3.2 Multi-dimensional covariance matrix

In analogy to uni-variate case, we can define random walk on multivariate matrix using inverse Wishart density, since it is the form of posterior density of unknown covariance matrix of multi-variate normal density [15]. The evolution model of $p \times p$ matrix Q_t is

$$f(Q_t|Q_{t-1}) = i\mathcal{W}(\Psi, \nu) = \frac{|\Psi|^{\frac{1}{2}\nu}|Q_t|^{-\frac{1}{2}(\nu+p+1)}}{2^{\frac{\nu p}{2}}\Gamma_p(\frac{\nu}{2})} \exp\left\{ \operatorname{tr}\left(-\frac{1}{2}\Psi Q_t^{-1} \right) \right\},$$

where $\Gamma_p()$ is the multivariate Gamma function [19]. Moments are:

$$\mathsf{E}(Q_t) = \frac{\Psi}{\nu - p - 1} \equiv Q_{t-1}, \ \operatorname{var}(q_{ij}) = \frac{(\nu - p + 1)\psi_{ij}^2 + (\nu - p - 1)\psi_{ii}\psi_{jj}}{(\nu - p)(\nu - p - 1)^2(\nu - p - 3)}. \tag{22}$$

Remark 2 (Univariate special case). of the Wishart density is inverse-Gamma which was used in the previous Section. For diagonal elements (22) reduces to

$$\mathsf{E}(q_{ii,t}) = \frac{\psi_{ii,t}}{\nu - p - 1} \equiv q_{ii,t-1}, \ \operatorname{std}(q_{ii,t}) = \frac{\sqrt{2}\psi_{ii,t}}{(\nu - p - 1)\sqrt{\nu - p - 3}} \equiv kq_{ii,t}.$$

Here, we imposed the same requirements as in (19) which leads to the following choice

$$\nu = \frac{2}{k^2} + p + 3, \ \ \psi_{ii,t} = (\nu - p - 1)q_{ii,t-1}. \tag{23}$$

Generalizing (23) to the full rank we obtain $\Psi_t = (\nu - p - 1)Q_{t-1}$.

Sampling from inverse-Wishart density can be achieved using samples from Wishart identity since

$$Q_t \sim i\mathcal{W}(\Psi_t, \nu) \iff Q_t^{-1} \sim \mathcal{W}(\Psi_t^{-1}, \nu).$$

An algorithm for sampling from Wishart density has been proposed in [20]. It can be adapted to inverse-Wishart density as follows:

1. Generate a random $p \times p$ lower triangular matrix A such that:

   - $a_{ii} = (\chi_{n-i+1}^2)^{1/2}$, i.e. a_{ii} is the square root of a sample taken from a chi-square distribution χ_{n-i+1}^2
   - $a_{i,j}$, for $j < i$, is sampled from a standard normal distribution $\mathcal{N}(0, 1)$.

2. Compute the matrix $P_t = LAA'L'$ where L is Choleski decomposition of Ψ_t^{-1}, $\Psi_t^{-1} = LL'$. P_t^{-1} is a sample from the inverse-Wishart distribution $\mathcal{W}(\Psi_t, \nu)$.

Note the Choleski decomposition of matrix Ψ_t and thus Q_{t-1} playes an important role in this scheme. This decomposition is also important in square-root implementations of the Kalman filter [21]. Hence, combination of these two approaches can be improve computational efficiency of the scheme.

Remark 3 (Square-root EKF). Square root algorithms are common for Kalman filters, however, these must be modified in order to be used for EKF. Specifically, the only part of the filter that can be computed in square-root form is the Kalman gain, K (6), since predictions are computed using non-linear evolution model. However, the standard square-root algorithm typically computes numerical values of $\overline{K} = AK$. This is advantagegous for computing predictions of the standard Kalman filter, but disadvantageous for EKF. The K needed in EKF my be obtained as $K = A^{-1}\overline{K}$ but this is computationally expensive. A better solution is still to be found.

4 Experimental Verification

In this Section, we apply the approach defined above to two problems: a toy problem where model disturbances are simulated, and PMSM drive model with disturbances simulated by a realistic simulator operating at 1μs.

4.1 PMSM model

For all experiments, we will use the following non-linear model of a PMSM electrical drive discretized at $\Delta t = 125\mu$s:

$$
\begin{aligned}
i_{s\alpha}(t+1) &= (1 - \frac{R_s}{L_s}\Delta t)i_{s\alpha}(t) + \frac{\Psi_{pm}}{L_s}\Delta t\omega_{me}(t)\sin\vartheta_e(t) + u_{s\alpha}(k)\frac{\Delta t}{L_s}, \\
i_{s\beta}(t+1) &= (1 - \frac{R_s}{L_s}\Delta t)i_{s\beta}(k) - \frac{\Psi_{pm}}{L_s}\Delta t\omega_{me}(k)\cos\vartheta_e(k) + u_{s\beta}(t)\frac{\Delta t}{L_s}, \quad (24) \\
\omega_{me}(t+1) &= \omega_{me}(t), \\
\vartheta_e(t+1) &= \vartheta_e(t) + \omega_{me}(t)\Delta t.
\end{aligned}
$$

Here, $i_{s\alpha}$, $i_{s\beta}$, $u_{s\alpha}$ and $u_{s\beta}$ represent stator current and voltage in the stationary reference frame, respectively; ω_{me} is electrical rotor speed and ϑ_e is electrical rotor position. R_s and L_s is stator resistance and inductance respectively, Ψ_{pm} is the flux of permanent magnets on the rotor, B is friction and T_L is load torque, J is moment of inertia, p_p is the number of pole pairs, k_p is the Park constant.

4.2 Simulated diagonal covariance matrix

We have simulated model (24) with the following covariance matrices: $Q_0 = \text{diag}([1e-6, 1e-6, 1e-3, 1e-4])$ and $R_0 = 1e-8\,I_2$, where I_2 is the 2×2 identity matrix. The covariance matrices Q_t were rapidly switched at times $t = [1000, 3000, 5000, 7000]$, at each time, one of the diagonal element was multiplied 10 times and after 1000 steps it was returned to its original value. See Fig. 1 for illustration, where simulated values are displayed in tandem with posterior mean values. The results were obtained using evolution model (21) with parameters $Q_{ref} = 10Q_0, l = 0.999, k = 0.1$. Note that the estimates follow simulated values even for 20 particles which is quite promising from computational point of view.
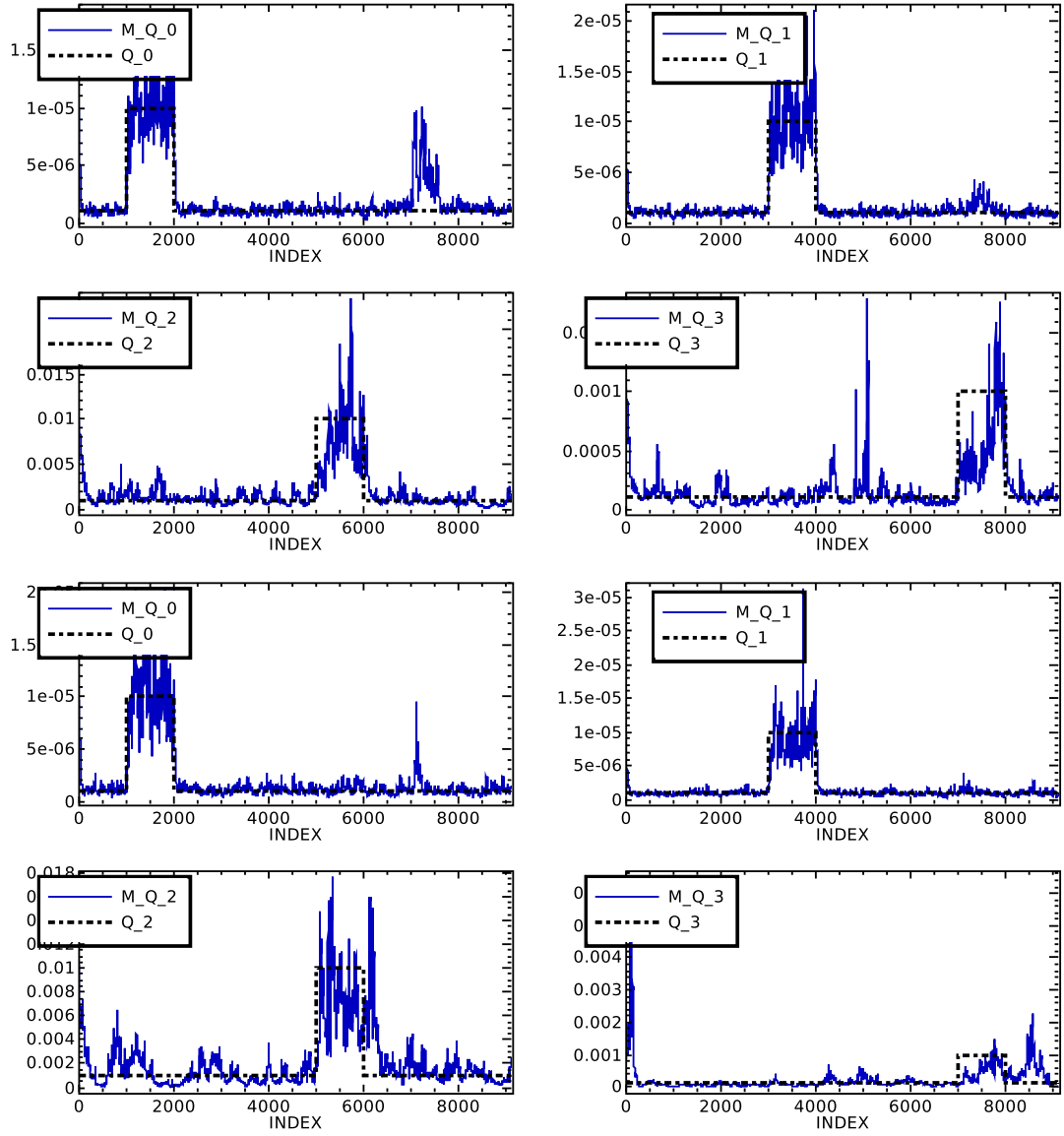
Figure 1: Mean values of estimated diagonal matrix Q by MPF using EKF for PMSM model and random walk model (21) with parameters $Q_{ref} = 10Q_0, l = 0.999, k = 0.1$. Two runs for 200 particles (top) and 20 particles (bottom) are displayed.

9

In this experiment, the stabilization of the evolution model via (21) was necessary, since the original model (20) may become unstable. Results of the same experiment as above except for $l = 0.9999$, are displayed in Fig. 2. Note that for 200 particles, the filter provides reasonable estimates of the covariance matrix, while for 20 particles, the filter got stuck in a local minima, with $Q_{3,3,t}$ and $Q_{4,4,t}$ falling down to zero. This can be explained by the chosen evolution model. In fact, they estimates of $Q_{3,3,t}$ were of orders $1e - 10$ for which the chosen evolution model allows standard deviation of similar order and thus the particle at time $t+1$ remain in this part of space. With large particle population, the algorithm was able to reject those while with 20 particles there was no reasonable alternative particle to choose.

Remarks:

- Qualitatively similar results were obtained using log-normal evolutions model, Remark 1. However, we will consider gamma sampling in further experiments, since it is slightly less computationally demanding.

- The only tuning parameter in the evolution models is parameter k. Its low values encourage small variations in the variance evolution yielding smoother curve estimates. Results of two simulation runs with $k = 0.1$ and $k = 0.3$ are displayed in Fig. 3. Note that the evolution model with $k = 0.1$ allows generation of a narrow range of particles (visualized by bounds on minimal and maximal value of the particle range), hence the estimates are not able to react to the step function quickly enough. The evolution model with $k = 0.3$ generates a wider range of particles, thus being able to react faster to step function. However, it tends to produce rather 'spiky' estimates in contrast to smoother curves provided by the model with lower k.

- All experiments so far were done for systems with high signal-to-noise ratio (SNR) with variance of both observation and state noise being much lower than absolute values of the state variable. Specifically the variance of the observation noise was of order $1e - 8$ which is much lower than the lowest variance on the state variable, i.e. $1e - 6$. Results of estimation of Q_t for two simulation runs with $R_t = 1e - 6\,I_2$ and $R = 1e - 4\,I_2$ are displayed in Figure 4.

  Note that for $R_t = 1e - 6\,I_2$ the step on $Q_{1,1,t}$ was not detected, but it was compensated by increase of variance $Q_{2,2,t}$. For $R_t = 1e - 4$ steps in variance were not detected at all without any compensation. The result is expected since the first two state variables are directly observed, with variance of the observation noise being 100 times higher that that of the noise on the evolution model of $Q_{1,1,t}$ and $Q_{2,2,t}$.

4.3 Simulated full covariance matrix

An experiment with full covariance matrix was performed with data simulated in the same way as in the previous experiments, i.e. step functions on diagonal of the covariance matrix Q_t. However, the covariance matrix was estimated using inverse-Wishart random
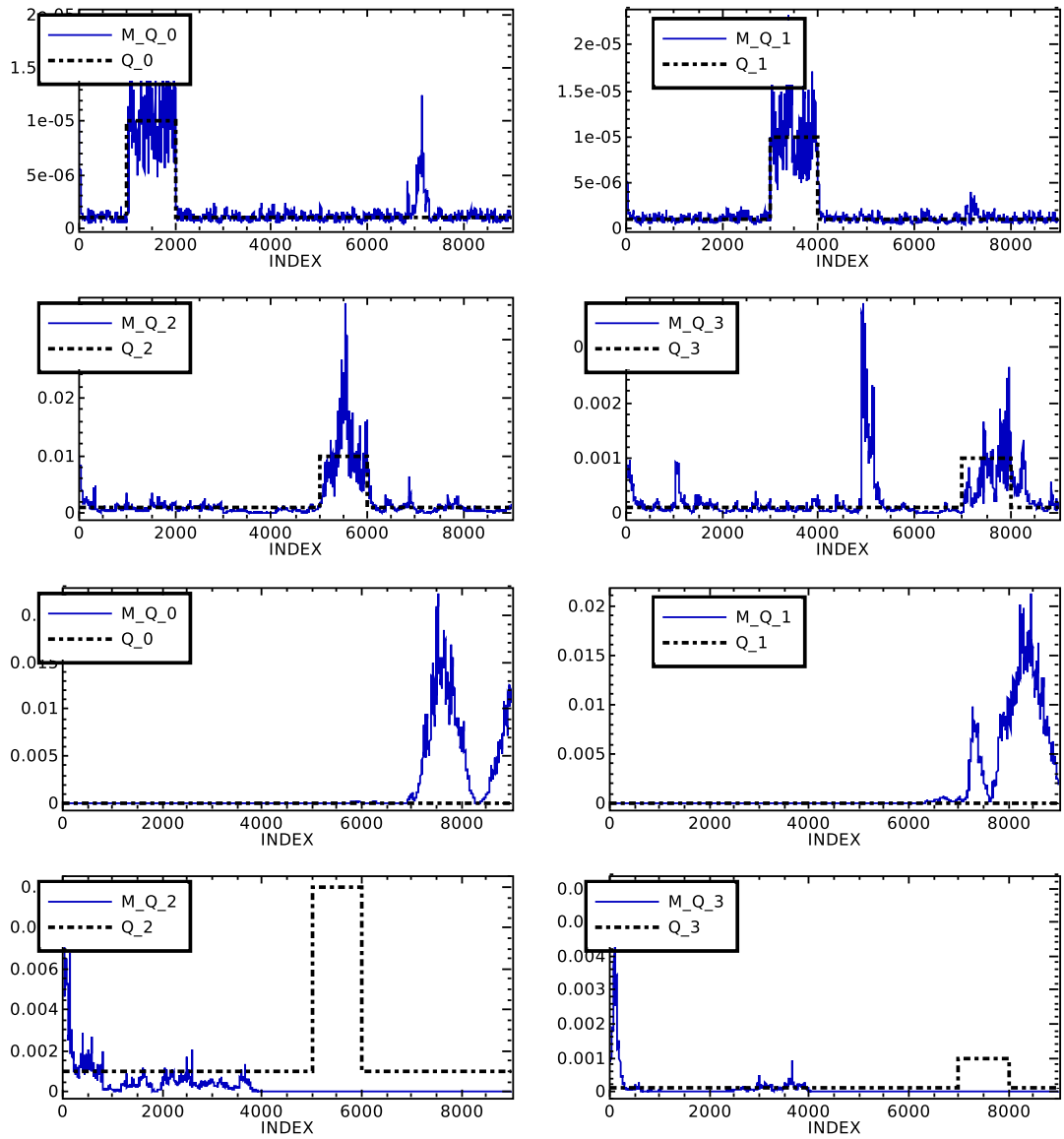
Figure 2: Mean values of estimated diagonal matrix Q by MPF using EKF for PMSM model and random walk model (21) with parameters $Q_{ref} = 10Q_0, l = 0.999, k = 0.1$. Two runs for 200 particles (top) and 20 particles (bottom) are displayed.
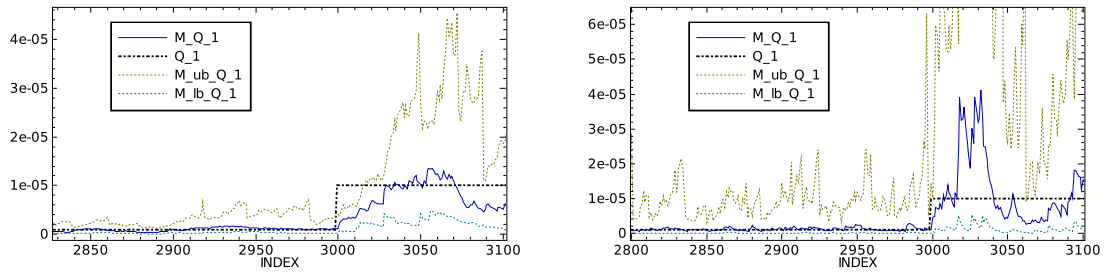
Figure 3: Details of results for two simulation runs with $k = 0.1$ (left) and $k = 0.3$ (right). Simulated values of $Q_{2,2,t}$ (thick dashed line) are displayed in tandem with mean value of posterior density (solid line) and upper- and lower-bound on the posterior density (thin dashed line).

walk (23). Results of the experiment are displayed in Figure 5. Interpretation of the results is more demanding than in the previous case and detailed analysis is beyond scope of this report. We only note that:

- Once again, Wishart random walk as defined in (23) is not numerically stable and estimation procedure using thsi random walk diverges similarly to that on Fig. 2. Geometric combination of reference vector and diagonal element in the Choleski decomposition of Q_{t-1} in the style of (21) was required to stabilize the procedure.

- More particles is now required to achieve better results due to the increase of dimensionality of the state vector. Results in Fig. 5 were obtained for 2000 particles.

- The problem is now highly ill-posed since 2 observations at time t are used to estimate 14 parameters. The posterior density on the paremeters is expected to be rather flat and correlated.

- Note that all four step functions on the diagonal elements of Q_t are reflected in the posterior mean value. However, they are less accurate that posterior mean values obtained using pure diagonal model, Fig. 1. This is expected since more candidates for explanation of the observations are now evaluated.

- All non-diagonal elements were supposed to be zero, yet non-zero values were estimated, especially during transient periods. Assessment of significance of these values is rather complicated. In Fig. 5 the scales of off-diagonal elements were adjusted such that maximum on the y-axix of i, jth element is square-root of product of maximas on ith and jth diagonal element. In this scale, the estimated values appears to be reasonably close to zero.

- Similarly to the previous experiment, the results deteriorate with decreasing SNR.

12
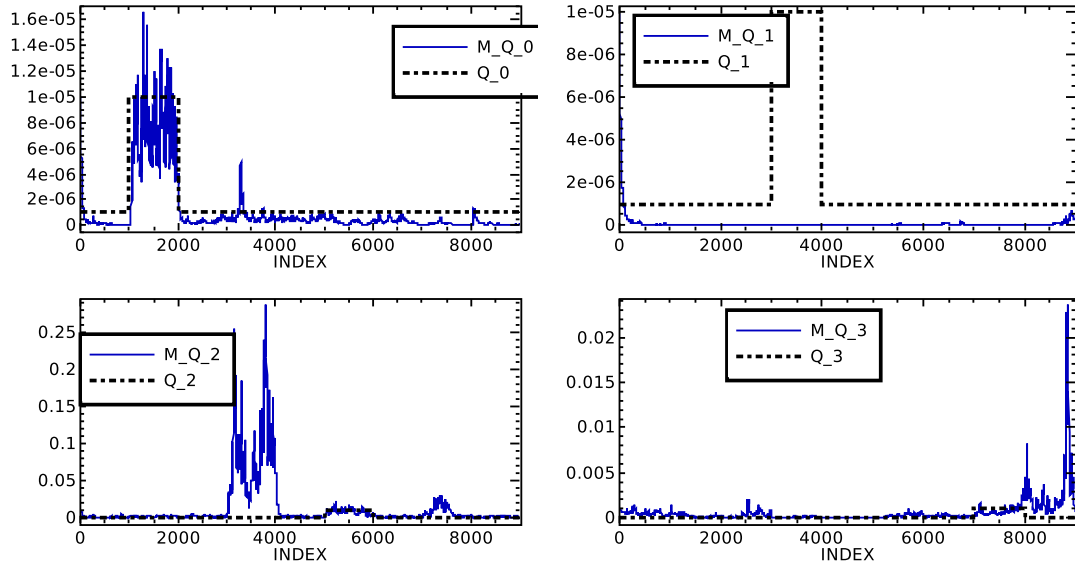
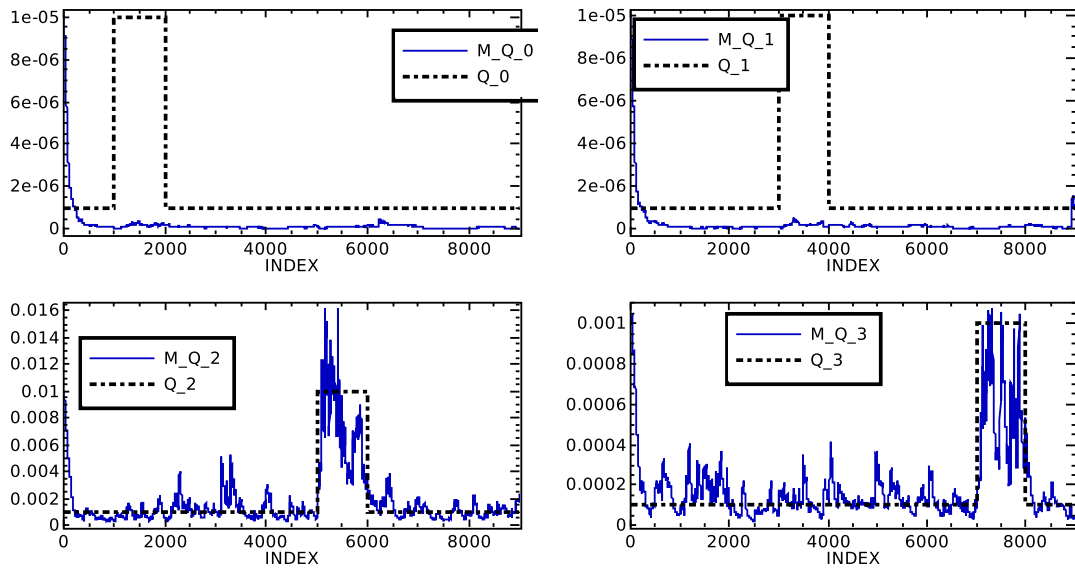Variance $R_t = 1e - 6\, I_2$

Variance $R_t = 1e - 4\, I_2$
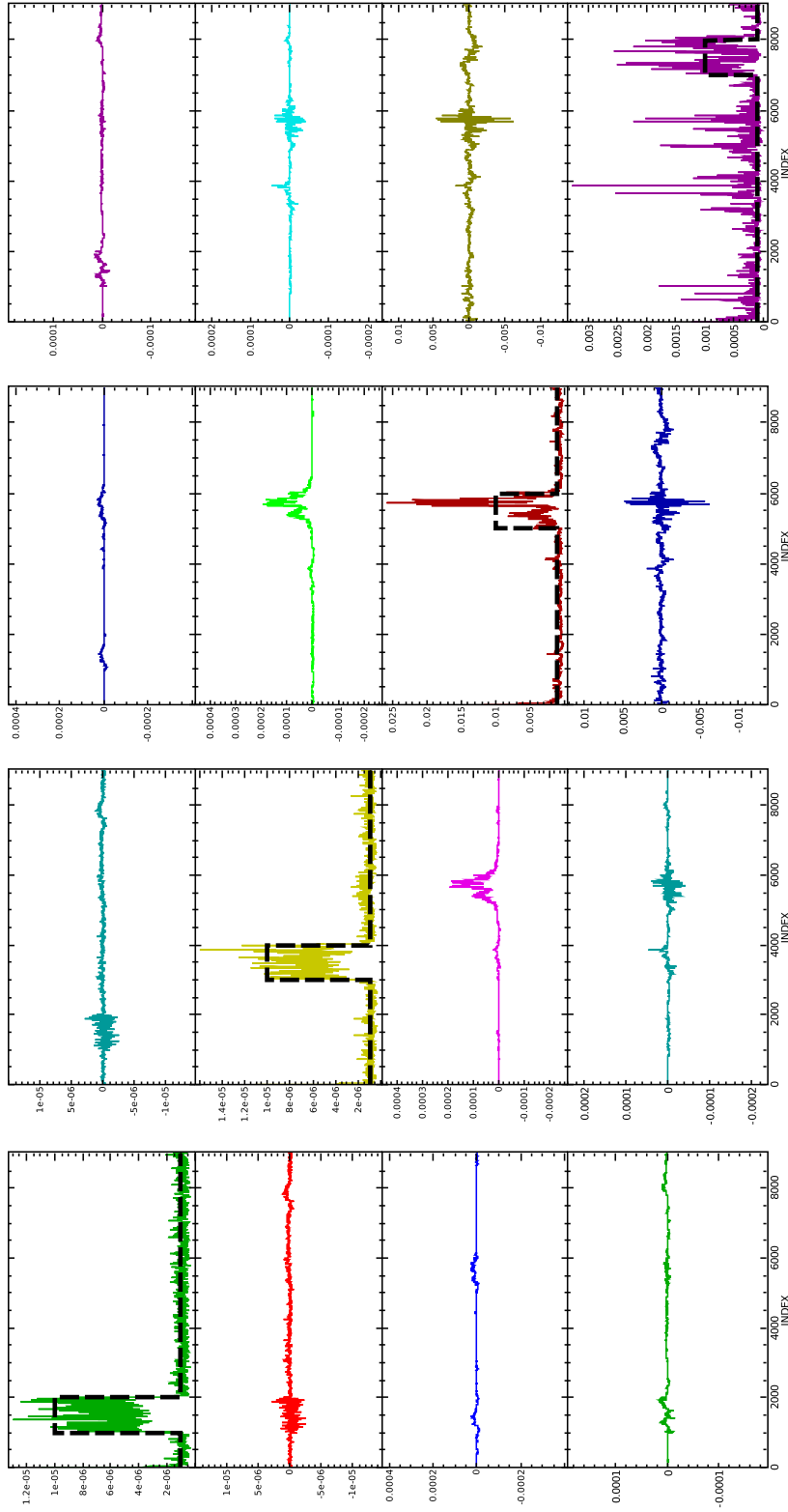
Figure 4: Results of estimation for different SNR.

Figure 5: Mean value of the posterior density on covariance matrix Q_t. Each element of the 4×4 matrix of figures displayes one element of the Q_t matrix. Y-axis scales of the off-diagonal elements, $i \neq j$, were scaled such that $y_{i,j,\max} = \sqrt{y_{i,i,\max} y_{j,j,\max}}$.
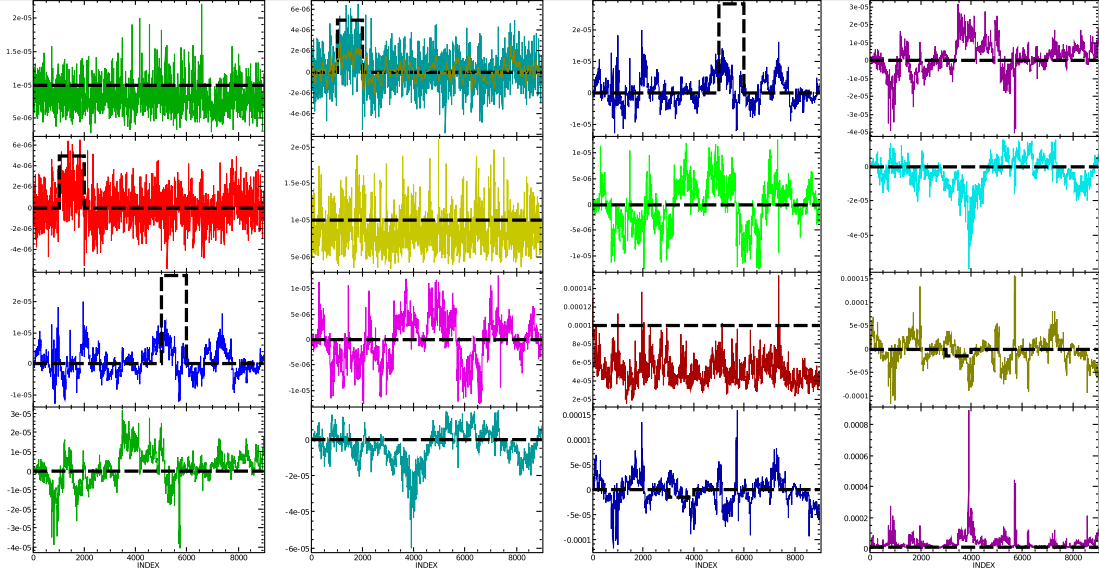
14

Figure 6: Posterior expected values of Q_t for exeriment with varying off-diagonal elements.

The ability to estimate non-diagonal elements was tested by simulating $Q_t = Q_0$ with reference values Q_0, $l = 0.99$ complemented by step functions on off-diagonal elements

$$q_{1,2} = q_{2,1} = \frac{1}{2}\sqrt{q_{11}q_{22}} \text{ for } t \in [1000, 2000],$$

$$q_{3,4} = q_{4,3} = -\frac{1}{2}\sqrt{q_{44}q_{33}} \text{ for } t \in [3000, 4000],$$

$$q_{1,3} = q_{3,1} = 0.9\sqrt{q_{11}q_{33}} \text{ for } t \in [5000, 6000],$$

Results of this experiment are displayed in Figure 6. Note that the first step on $q_{1,2}$ is reflected in the estimates, while the other steps on $q_{3,4}$ and $q_{1,3}$ are completely ignored. We conjecture that this is due to the use of EKF as approximation in (17). The only way how matrix Q_t enters posterior likelihood (7) of the EKF is via the Kalman gain K (6) and matrix R_y. However, due to diagonal matrix R_t and sparse matrix C, the influence of Q_t on the posterior is greatly reduced.

4.4 Realistic Simulator of a PMSM drive

The main motivation for this research was the need to estimate covariance structure of PMSM drive. The most challenging scenario for the drive is the area of very low speed. A comparison of estimates of rotor speed, ω_t, and rotor position, θ_t, using (i) EKF with expert-selected covariance matrix, and (ii) MPF with wishart random walk with reference set the the same covariance matrix as EKF, is displayed in Figure 7. Note that both estimators fail to estimate the true positions and speed of the rotor. Note, however, that the true values remain within uncertainty bounds of the MPF for most of
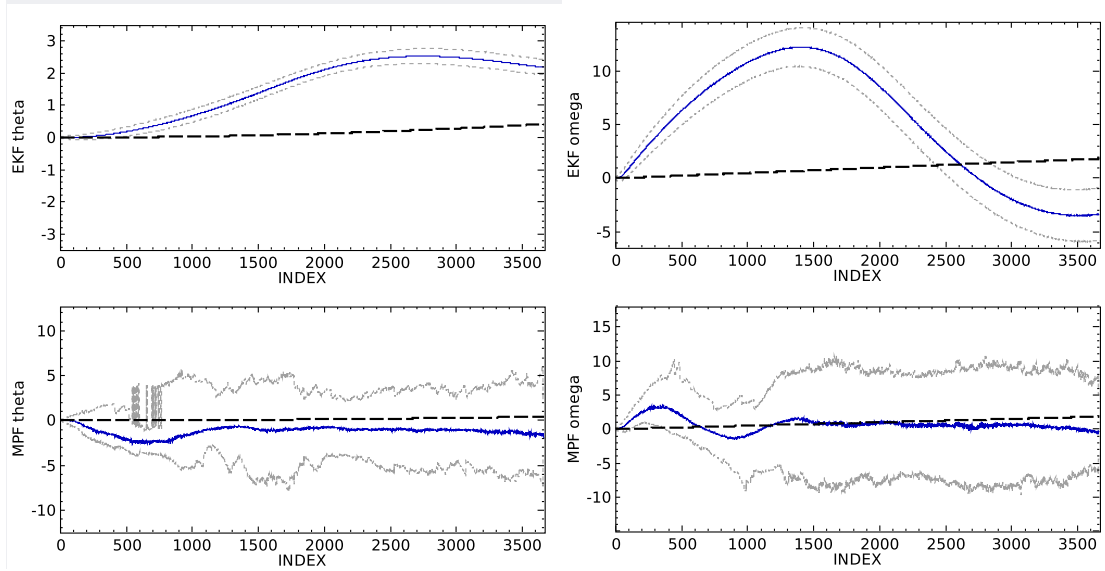
Figure 7: Posterior density of rotor position, θ_t, (left column) and rotor speed, ω_t, (right column) obtained using EKF with expert-chosen covariance (upper row) and MPF with inverse-Wishart random walk (lower row). Dashed black line denotes simulated values, solid blue denotes mean value of the posterior density and dashed gray denotes unceratinty bounds on the estimate.

the time. The main conclusion from this experiment is that the observed data do not carry sufficient information allowing correct estimation of random-walk evolution model on covariance matrix Q_t.

5 Discussion and conclusion

We have shown in simulation, that combination of extended Kalman filter with particle filter via marginalized particle filter is applicable to the problem of estimation of unknown covariance matrices. The method is a combination of two constituents:

1. Bayesian filter of the non-linear state-space model, potentially approximate,

2. Extension of the state for elements of Q_t and definition of its evolution model.

In this report, we have investigated the use of EKF as the Bayesian filter and random-walk evolution models for Q_t. Potential of the method was clearly demonstarted on the toy problem with diagonal matrices, Section 4.2, where 20 particles (each complemented by an associated Kalman filter) were sufficient to correctly estimate diagonal of the co-variance matrix Q_t. However, further extension of the approach to completely unknown matrix Q_t presents a challenge. Note that dimension of the state extended by the un-known matrix Q_t is significantly greater than dimension of the observations. Hence, the problem becomes ill-posed and yielding only flat posterior density on the state.

16

For a practical problem, such as sensor-less control of PMSM drives, the only feasible way to address the problem appears to be construction of a detailed model of covariance structure. Such a model should be based on all available expert knowledge about laws of physics of the process. All unknowns and uncertainties within the model shoulde be aggregated into a *small number* of shaping parameters. For this aprticular case, deviations from the mean values (given by the difference equations (24)) aer known to arise from various sources:

PWM (pulse width modulation) due to dead-times effects and Volt-Ampere characteristics of the power electronics. These deviations are ever-present but bounded, so that their variance can be easily deduced.

Discrete sampling: differential equations of the motor are converted to the form of difference equations using a fixed width sampling rate. It is assumed that profile of the signal during one sample step is known (e.g. rectangular of linear). This assumption is justified when the system is close to linear (e.g. when the revolutions are slow, approximation of two subsequent $\sin(x)$ by a linear function is justifiable. However, the accuracy of this approximation heavily depends the operating regime. For example, the error of approximation of function $\sin(x)$ is growing with growing speed of the rotor .

Unknown parameters: most likely, parameters of interest, such as R, L etc, are not known exactly. A deviation from their true values would cause another (systematic) error in the equations.

These sources can be parameterized with one or two parametetrs each. Marginalized particle filter, as presented in Section 2.3, presents a promising tool for estimation of such shaping parameters since it allows their non-linear evolution models.

References

[1] P. Vas, *Sensorless vector and direct torque control.* Oxford University Press New York, 1998.

[2] R. Dhaouadi, N. Mohan, and L. Norum, "Design and implementation of an extended Kalman filter for thestate estimation of a permanent magnet synchronous motor," *Power Electronics, IEEE Transactions on*, vol. 6, no. 3, pp. 491–497, 1991.

[3] S. Bolognani, R. Oboe, and M. Zigliotto, "Sensorless full-digital PMSM drive with EKF estimation of speed and rotor position." *IEEE Trans. on Industrial Electronics*, vol. 46, pp. 184 – 191, 1999.

[4] S. Bolognani, M. Zigliotto, and M. Zordan, "Extended-range PMSM sensorless speed drive based on stochasticfiltering," *Power Electronics, IEEE Transactions on*, vol. 16, no. 1, pp. 110–117, 2001.

[5] J. Holtz, "Sensorless control of induction motors and pm synchronous machines. tutorial." in *IEEE International Symposium on Industrial Electronics (ISIE)*, Dubrovnik, Croatia, 2005.

[6] Z. Peroutka, "Design considerations for sensorless control of PMSM drive based on extended Kalman filter," in *11th European Conference on Power Electronics and Applications (EPE)*, Dresden, Germany, 2005.

[7] V. Šmídl and A. Quinn, "Variational Bayesian filtering," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5020–5030, 2008.

[8] R. Mehra, "Approaches to adaptive filtering," *Automatic Control, IEEE Transactions on*, vol. 17, no. 5, pp. 693–698, 1972.

[9] 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.

[10] J. Duník and Šimandl M., "Estimation of state and measurement noise covariance matrices by multi-step prediction," Seoul, Korea, 2008, pp. 3689–3684.

[11] D. Wiberg, T. Powell, and D. Ljungquist, "An online parameter estimator for quick convergence and time-varying linear systems," *Automatic Control, IEEE Transactions on*, vol. 45, no. 10, pp. 1854–1863, 2000.

[12] P. Djuric and J. Miguez, "Sequential particle filtering in the presence of additive Gaussian noise with unknown parameters," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, 2002.

[13] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice.* Springer, 2001.

[14] 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, pp. 2279–2289, 2002.

[15] V. Peterka, "Bayesian approach to system identification," in *Trends and Progress in System identification*, P. Eykhoff, Ed. Oxford: Pergamon Press, 1981, pp. 239–304.

[16] E. Daum, "New exact nonlinear filters," in *Bayesian Analysis of Time Series and Dynamic Models*, J. Spall, Ed. New York: Marcel Dekker, 1988.

[17] R. Kalman, "A new approach to linear filtering and prediction problem," *Trans. ASME, Ser. D, J. Basic Eng.*, vol. 82, pp. 34–45, 1960.

[18] H. Sorenson, "On development of practical nonlinear filters," *Information Sciences*, vol. 7, pp. 253–270, 1974.

[19] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*. New York: Dover Publications, 1972.

[20] W. Smith and R. Hocking, "Wishart variate generator," *Applied Statistics*, vol. 21, pp. 341–345, 1972.

[21] P. Park and T. Kailath, "New square-root algorithms for Kalman filtering," *Automatic Control, IEEE Transactions on*, vol. 40, no. 5, pp. 895–899, 1995.