Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Estimation of State and Measurement Noise Covariance Matrices by Multi-Step Prediction ⋆

Jindřich Duník, Miroslav Šimandl

Department of Cybernetics & Research Centre Data-Algorithms-Decision Making, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, 306 14 Pilsen, Czech Republic
e-mails: dunikj@kky.zcu.cz (J. Duník), simandl@kky.zcu.cz (M. Šimandl)

Abstract: Estimation of noise covariance matrices for linear or nonlinear stochastic dynamic systems is treated. The novel off-line technique for estimation of the covariance matrices of the state and measurement noises is designed. The technique is based on the multi-step prediction error and on knowledge of the system initial condition and it takes an advantage of the well-known standard relations from the area of state estimation techniques and least square method. The theoretical results are illustrated in numerical examples.

1. INTRODUCTION

The problem of recursive state estimation of discrete-time stochastic dynamic systems has been a subject of considerable research interest for the last several decades. The general solution of the estimation problem is given by the Bayesian Recursive Relations (BRR's) for computation of the probability density functions (pdf's) of the state conditioned by the measurements.

Although the BRR's represent the full solution of the filtering problem, their exact solution is possible only for a few special cases, e.g. for linear Gaussian systems, which leads to the well-known Kalman Filter (KF) [Lewis, 1986]. In most of other situations it is necessary to apply some approximative technique to the BRR's solution. The approximative methods can be generally classified into the two groups: local and global methods [Sorenson, 1974].

The local methods are often based on approximation of the nonlinear functions in the state or measurement equation so that the technique of the Kalman filter design can be used for the BRR's solution for the nonlinear systems as well [Lewis, 1986, Nørgaard et al., 2000, Julier and Uhlmann, 2004, Duník et al., 2005, Xiong et al., 2006]. This approach causes that all conditional pdf's of the state estimate are given by the first two moments, i.e. by mean value and covariance matrix, which induces local validity of the state estimates and consequently impossibility to generally ensure the convergency of the local filter estimates. As an advantage of the local methods the simplicity of the BRR's solution can be mentioned. The Extended Kalman Filter, Unscented Kalman Filter, and Divided Difference Filters exemplify the local methods.

The global methods are rather based on approximation of the conditional pdf of the state estimate of some kind to accomplish better state estimates. These methods are more sophisticated but they have higher computational demands than the local methods. Further details and references can be found e.g. in [Šimandl and Straka, 2003, Duník et al., 2005].

The successful design and application of an arbitrary filter is, however, conditioned by knowledge of the sufficiently exact model of the real system, namely the "deterministic part" including the functions in the state and measurement equations and the "stochastic part" including the statistical properties of the system initial condition and of the state and measurement noises. Thus, the relatively close attention have been paid to the estimation of unknown parameters in the system description [Wan and Nelson, 2001, Mehra, 1972].

This work is addressed to the problem of estimating the second order noise statistics. The methods for estimation of the state and measurement noise covariance matrices can be generally divided into on-line and off-line methods. The on-line methods, usually called adaptive filtering methods, can be further classified into the various groups such as Bayesian estimation, maximum likelihood estimation, correlation methods, and covariance matching methods [Mehra, 1972]. These methods can be further modified into the form where, instead of computation of noise covariance matrices, the filter gain is directly computed [Mehra, 1970, Bos et al., 2005]. As an alternative to the adaptive filtering methods the methods based on the minimax approach have been proposed [Verdú and Poor, 1984]. The common disadvantage of the adaptive and minimax filtering methods is that they have been designed for the linear systems only.

The off-line estimation can be performed by e.g. the subspace methods [Palanthandalam-Madapusi et al., 2005] or the prediction error methods [Ljung, 1999]. These methods are, however, suitable mainly for linear systems or for special types of nonlinear systems, e.g. Wiener or Hammerstein. Recently, the novel technique for estimation of the system noise covariance matrices either for linear or nonlinear systems have been proposed [Šimandl and Duník, 2007]. The technique, which belongs, in essence, into the correlation methods, was based on knowledge of the system initial condition and on the possibility to measure a first few data repeatedly [1] . However, in some cases

⋆ The work was supported by the Ministry of Education, Youth and Sports of the Czech Republic, project No. 1M0572.

[1] Motivation for the technique was found in the problem of traffic control where the control law is based on the minimisation of the total lengths of car queues in all arms of an intersection [Homolová and Nagy, 2005]. Unfortunately, the queue lengths are hardly measurable and they have to be

the estimation of the noise covariance matrices on the basis of one measured realisation is advantageous.

The first aim of the paper is therefore to extend the technique given in [Šimandl and Duník, 2007] to estimate the state and measurement noise covariance matrices of linear systems on the basis of one measured realisation. The second aim is to analyse the stability of the technique for estimation of the noise covariance matrices of nonlinear systems on the basis of multiple measured data sets.

The paper is organised as follows. Section 2 is devoted to the system description and to the Bayesian solution of the estimation problem. Then, the thorough analysis of the multi-step prediction error is given and novel relations for estimation of the noise covariance matrices are derived in Section 3. The results from the area of linear systems are extended into the area of nonlinear systems in Section 4. Numerical illustrations and conclusion remarks are given in Section 5 and Section 6, respectively.

2. PROBLEM STATEMENT

Let the discrete-time nonlinear Gaussian stochastic system be considered

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k) + \mathbf{w}_k, k = 0, 1, 2, \ldots, \tag{1}$$

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, k = 0, 1, 2, \ldots, \tag{2}$$

where the vectors $\mathbf{x}_k \in \mathbb{R}^{n_x}$ and $\mathbf{z}_k \in \mathbb{R}^{n_z}$ represent the immeasurable state of the system and the measurement at time instant k, respectively, and $\mathbf{f}_k : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_x}$, $\mathbf{h}_k : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_z}$ are known vector functions. The pdf of the initial state $p(\mathbf{x}_0) = \mathcal{N}\{\mathbf{x}_0 : \bar{\mathbf{x}}_0, \mathbf{P}_0\}$ is known as well. The variables $\mathbf{w}_k \in \mathbb{R}^{n_x}$ and $\mathbf{v}_k \in \mathbb{R}^{n_z}$ are the state and the measurement white noises with Gaussian pdf's $p(\mathbf{w}_k) = \mathcal{N}\{\mathbf{w}_k : \mathbf{0}, \mathbf{Q}\}$ and $p(\mathbf{v}_k) = \mathcal{N}\{\mathbf{v}_k : \mathbf{0}, \mathbf{R}\}$, $\forall k$, respectively, where the covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ are supposed to be unknown. The noises are mutually independent and independent of the initial state.

The aim of the filtering, as a special part of the state estimation, is to find the state estimate in the form of the conditional pdf of the state $\mathbf{x}_k$ conditioned by measurements up to the time k, i.e. $p(\mathbf{x}_k|\mathbf{z}^k)$ is looked for, where $\mathbf{z}^k = [\mathbf{z}_0, \ldots, \mathbf{z}_k]$. The solution of the filtering problem, given by the BRR's [Lewis, 1986], is conditioned by knowledge of the state and the measurement noise pdf's, i.e. by $p(\mathbf{w}_k)$ and $p(\mathbf{v}_k)$, $\forall k$. However, the covariance matrices of the state and measurement noises are unknown and they have to be estimated.

The first goal of the paper is to design the off-line technique for estimation of the noise covariance matrices for linear Gaussian systems based on the one measured realisation and so to extend the technique proposed in [Šimandl and Duník, 2007]. The technique comes out from the statistical properties of the multi-step measurement prediction error and it allows to determine set of independent equations for estimation of all elements of the noise covariance matrices. The second aim is to apply the proposed technique into the area of nonlinear systems as well with stress on the bounded multi-step prediction error. To ensure bounded prediction error for nonlinear systems the stability analysis of the local filters will be exploited.

estimated on the basis of measured quantities. However, there are many (on the order of hundreds) of measured daily courses of passing vehicles through the intersection, e.g. measurable intensity of traffic flow, at disposal. Thus, the proposed technique is suitable not only for systems with possibility to measure a first few data repeatedly, but also it is easily applicable for periodic processes.

3. ESTIMATION OF NOISE COVARIANCE MATRICES OF LINEAR SYSTEMS

In this section the novel technique for estimation of the state and measurement noise covariance matrices will be discussed. This technique will be based on the known initial condition of the system and on the analysis of properties of measurement multi-step prediction error.

3.1 Multi-Step Prediction

Let the linear t-invariant Gaussian system (1), (2) with known initial condition $p(\mathbf{x}_0)$, where $\mathbf{f}_k(\mathbf{x}_k) = \mathbf{F}\mathbf{x}_k$ and $\mathbf{h}_k(\mathbf{x}_k) = \mathbf{H}\mathbf{x}_k$, be considered. Then, the multi-step predictor, with initial condition $p(\mathbf{x}_0|\mathbf{z}^{-1}) = p(\mathbf{x}_0) = \mathcal{N}\{\mathbf{x}_0 : \hat{\mathbf{x}}_{0|-1}, \mathbf{P}_{0|-1}\}$; $\hat{\mathbf{x}}_{0|-1} = \bar{\mathbf{x}}_0$, $\mathbf{P}_{0|-1} = \mathbf{P}_0$, is given as

$$\hat{\mathbf{x}}_{k+1|-1} = \mathbf{F}\hat{\mathbf{x}}_{k|-1}, \tag{3}$$

$$\mathbf{P}_{k+1|-1} = \mathbf{F}\mathbf{P}_{k|-1}\mathbf{F}^T + \mathbf{Q}^P, \tag{4}$$

where $\hat{\mathbf{x}}_{k+1|-1}$ is the $(k+1)$-th state prediction with covariance matrix $\mathbf{P}_{k+1|-1}$. The predictive state estimate is thus based on the known initial condition and it does not depend on the measurement, i.e. the predictive mean and the covariance matrix are conditioned by $\mathbf{z}^{-1}$. If the true state noise covariance matrix is known it is used in the predictor design, i.e. $\mathbf{Q}^P = \mathbf{Q}$. In the situation when there is no such information, let the matrix $\mathbf{Q}^P$ in the predictor algorithm be chosen arbitrarily.

To find a relation of the measurement multi-step prediction error it is advantageous to rewrite (1) and (3) as

$$\mathbf{x}_{k+1} = \mathbf{F}^{k+1}\mathbf{x}_0 + \sum_{i=0}^{k} \mathbf{F}^i \mathbf{w}_{k-i}, \tag{5}$$

$$\hat{\mathbf{x}}_{k+1|-1} = \prod_{i=0}^{k} \mathbf{F}\hat{\mathbf{x}}_{0|-1} = \mathbf{F}^{k+1}\hat{\mathbf{x}}_{0|-1}. \tag{6}$$

Then, with respect to the linear measurement equation, the prediction error at time instant $(k+1)$ is given by

$$\begin{aligned}\mathbf{e}_{k+1} &= \mathbf{z}_{k+1} - \hat{\mathbf{z}}_{k+1|-1} = \mathbf{H}(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|-1}) + \mathbf{v}_{k+1} = \\ &= \mathbf{H}\mathbf{F}^{k+1}(\mathbf{x}_0 - \hat{\mathbf{x}}_{0|-1}) + \mathbf{H}\sum_{i=0}^{k} \mathbf{F}^i \mathbf{w}_{k-i} + \mathbf{v}_{k+1},\end{aligned} \tag{7}$$

where $\hat{\mathbf{z}}_{k+1|-1} = \mathbf{H}\hat{\mathbf{x}}_{k+1|-1}$ represents $(k+1)$-th prediction of the measurement. And as will be shown, (7) is the key relation for estimation of the system noise covariance matrices.

3.2 Properties of Multi-Step Prediction Error

In this part the prediction error is analysed which allows to find relations for estimation of the noise covariance matrices. The mean of the prediction error, with respect to the zero means of the state and measurement noises, is

$$E[\mathbf{e}_k] = \mathbf{H}E[\mathbf{x}_k - \hat{\mathbf{x}}_{k|-1}] + \mathbf{v}_k = \mathbf{0}, \forall k, \tag{8}$$

where $E[\mathbf{x}_0] = \bar{\mathbf{x}}_0 = \hat{\mathbf{x}}_{0|-1}$. As far as the second-order statistics are considered, the covariance matrix of $\mathbf{e}_k$ and cross-correlation matrices of $\mathbf{e}_l$ and $\mathbf{e}_k$, where $l < k$, can be easily computed as well, i.e.

$$\begin{aligned}\mathrm{cov}[\mathbf{e}_k] &= \mathrm{cov}[\mathbf{z}_k - \hat{\mathbf{z}}_{k|-1}] = \mathbf{H}\mathbf{F}^k\mathbf{P}_{0|-1}(\mathbf{F}^k)^T\mathbf{H}^T + \\ &+ \sum_{i=0}^{k-1} \mathbf{H}\mathbf{F}^i\mathbf{Q}(\mathbf{F}^i)^T\mathbf{H}^T + \mathbf{R},\end{aligned} \tag{9}$$

$$\mathrm{cov}[\mathbf{e}_l;\mathbf{e}_k] = \mathrm{cov}[\mathbf{z}_l - \hat{\mathbf{z}}_{l|-1};\mathbf{z}_k - \hat{\mathbf{z}}_{k|-1}] = \tag{10}$$
$$= \mathbf{H}\mathbf{F}^l\mathbf{P}_{0|-1}(\mathbf{F}^k)^T\mathbf{H}^T + \sum_{i=1}^{l}\mathbf{H}\mathbf{F}^{l-i}\mathbf{Q}(\mathbf{F}^{k-i})^T\mathbf{H}^T.$$

Relations (9), (10), which do not depend on chosen $\mathbf{Q}^P$, can be used for estimation of the true covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ based on the multiple measurements of a first few steps. This will be discussed later in detail. However, to be able to estimate the true covariance matrices of the noises on the basis of one measured data realisation it is necessary to specify further relations.

The first relation is based on the possibility to find a relation for computation of the covariance matrix of a mixture based on particular prediction errors. The basic idea will be discussed in the following proposition.

Proposition 1. Let two variables $\mathbf{x} \sim \mathcal{N}\{\mathbf{x} : \bar{\mathbf{x}}, \mathbf{P}_x\}$ and $\mathbf{y} \sim \mathcal{N}\{\mathbf{y} : \bar{\mathbf{y}}, \mathbf{P}_y\}$ with the known (true) means ($\bar{\mathbf{x}} = \bar{\mathbf{y}} = \mathbf{0}$) and covariance matrices be supposed and two sets of samples from particular variables $\mathbf{x}^{(1:N)} = [\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}]$ and $\mathbf{y}^{(1:N)} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(N)}]$ be available. The estimated mean of merged sets of samples, denoted as $\mathbf{z}^{(1:2N)} = [\mathbf{x}^{(1:N)}, \mathbf{y}^{(1:N)}]$, is then

$$\hat{\mathbf{z}} = \frac{1}{2N}\sum_{i=1}^{2N}\mathbf{z}^{(i)} = \frac{1}{2}\Big(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}^{(i)} + \frac{1}{N}\sum_{i=1}^{N}\mathbf{y}^{(i)}\Big) = \frac{1}{2}(\hat{\mathbf{x}} + \hat{\mathbf{y}}). \tag{11}$$

Applying mean value operator to (11) leads to

$$\bar{\mathbf{z}} = E[\hat{\mathbf{z}}] = \frac{1}{2}(\bar{\mathbf{x}} + \bar{\mathbf{y}}) = \alpha_x\bar{\mathbf{x}} + \alpha_y\bar{\mathbf{y}} = \mathbf{0}. \tag{12}$$

where $\alpha_x = \alpha_y = \frac{N}{2N} = \frac{1}{2}$. With respect to the known means of variables $\mathbf{x}$ and $\mathbf{y}$, the estimator of the covariance matrix is

$$\hat{\mathbf{P}}_z = \frac{1}{2N}\sum_{i=1}^{2N}(\mathbf{z}^{(i)} - \bar{\mathbf{z}})(\mathbf{z}^{(i)} - \bar{\mathbf{z}})^T = \frac{1}{2}(\hat{\mathbf{P}}_x + \hat{\mathbf{P}}_y), \tag{13}$$

where $\hat{\mathbf{P}}_x = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}^{(i)}(\mathbf{x}^{(i)})^T$ and $\hat{\mathbf{P}}_y = \frac{1}{N}\sum_{i=1}^{N}\mathbf{y}^{(i)}(\mathbf{y}^{(i)})^T$ are unbiased estimators of the covariance matrices $\mathbf{P}_x$ and $\mathbf{P}_y$ [Papoulis and Pillai, 2002]. Then

$$\mathbf{P}_z = E[\hat{\mathbf{P}}_z] = \frac{1}{2}(\mathbf{P}_x + \mathbf{P}_y) = \alpha_x\mathbf{P}_x + \alpha_y\mathbf{P}_y. \tag{14}$$

In other words, variable $\mathbf{z}$ can be understood as the mixture of variables $\mathbf{x}$ and $\mathbf{y}$. Thus, the true mean $\bar{\mathbf{z}}$ and covariance matrix $\mathbf{P}_z$ of variable $\mathbf{z}$ can be computed as weighted sum of means and covariance matrices of particular Gaussian terms and (11), (13) represent their unbiased estimators. Note that the weights α_x, α_y of particular terms are given by the number of samples.

The idea given in the proposition can be adopted for computation of the covariance matrix of the mixture of the particular prediction errors $\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_n$. The mixture is denoted as $\mathbf{e}$.

Theorem 1. Let the prediction errors $\mathbf{e}_0, \ldots, \mathbf{e}_n$ be supposed. Then, the covariance matrix of the variable $\mathbf{e}$, representing the mixture of the particular prediction errors $\mathbf{e}_k$, $\forall k$, is given by

$$\mathrm{cov}[\mathbf{e}] = \frac{1}{n+1}\sum_{k=0}^{n}\mathbf{H}\mathbf{F}^k\mathbf{P}_{0|-1}(\mathbf{F}^k)^T\mathbf{H}^T + $$
$$+ \sum_{k=0}^{n-1}\frac{n-k}{n+1}\mathbf{H}\mathbf{F}^k\mathbf{Q}(\mathbf{F}^k)^T\mathbf{H}^T + \mathbf{R}. \tag{15}$$

Proof 1. The proof of relation (15) is based on relation (14), which can be understood as a weighted sum of covariance matrices (9) of the particular prediction errors $\mathbf{e}_k$, $\forall k$, i.e.

$$\mathrm{cov}[\mathbf{e}] = \sum_{k=0}^{n}\alpha_k\mathrm{cov}[\mathbf{e}_k] = \frac{1}{n+1}\sum_{k=0}^{n}\Big(\mathbf{H}\mathbf{F}^k\mathbf{P}_{0|-1}(\mathbf{F}^k)^T\mathbf{H}^T + $$
$$+ \sum_{i=0}^{k-1}(\mathbf{H}\mathbf{F}^i\mathbf{Q}(\mathbf{F}^i)^T\mathbf{H}^T) + \mathbf{R}\Big) = \mathbf{R} + \frac{1}{n+1}\sum_{k=0}^{n}\mathbf{H}\mathbf{F}^k \times$$
$$\times \mathbf{P}_{0|-1}(\mathbf{F}^k)^T\mathbf{H}^T + \frac{1}{n+1}\sum_{k=0}^{n}\sum_{i=0}^{k-1}\mathbf{H}\mathbf{F}^i\mathbf{Q}(\mathbf{F}^i)^T\mathbf{H}^T, \tag{16}$$

which can be rewritten to the form (15). The weights $\alpha_k = \frac{1}{n+1}$, are the same for all k because the set of samples are supposed to have the same number of elements. However, the sets can generally have different number of samples which causes the different weights. In any case it holds that $\sum_{k=0}^{n}\alpha_k = 1$. ■

Further relations, suitable for estimation of the noise covariance matrices of stable linear systems only, are based on the fact, that the powers of the stable state transition matrix $\mathbf{F}^k$ become insignificant for increasing k. In other words it means that

$$\mathrm{cov}[\mathbf{e}_k] \approx \mathrm{cov}[\mathbf{e}_{k+1}], k > L_1, \tag{17}$$

where $L_1 \in \mathbb{Z}^+$ and $\mathbb{Z}^+$ is a set of positive integers. Similar relation can be found for the cross-correlation matrices

$$\mathrm{cov}[\mathbf{e}_l;\mathbf{e}_k] \approx \mathrm{cov}[\mathbf{e}_{l+1};\mathbf{e}_{k+1}], l > L_1. \tag{18}$$

It should be mentioned that the analogous relations describing properties of the prediction error can be derived also for a linear t-variant system (1), (2), where $\mathbf{f}_k(\mathbf{x}_k) = \mathbf{F}_k\mathbf{x}_k$, $\mathbf{h}_k(\mathbf{x}_k) = \mathbf{H}_k\mathbf{x}_k$.

Utilisation of the derived relations concerning statistical properties of the prediction error in the estimation of the state and measurement noise covariance matrices is discussed in the following parts.

3.3 Estimation by Multiple Realisations

The estimation of the system noise covariance matrices based on the multiple measured data sets (realisations) is considered as a first technique. Suppose that N sets of data were repeatedly measured for $(n+1)$ steps, i.e. data $\mathbf{z}_0^{(1:N)}, \ldots, \mathbf{z}_n^{(1:N)}$ are available, where $\mathbf{z}_k^{(1:N)} = [(\mathbf{z}_k^{(1)})^T, \ldots, (\mathbf{z}_k^{(N)})^T]^T$, $k = 0, 1, \ldots, n$. Then, the computation of multi-step prediction $\hat{\mathbf{z}}_{k|-1}$, $\forall k$, (6) on the basis of known mean of the system initial condition, allows to determine sets of the prediction error sequences $\mathbf{e}_k^{(1:N)} = [(\mathbf{e}_k^{(1)})^T, \ldots, (\mathbf{e}_k^{(N)})^T]^T$ according to (7). By means of these sets, the sample covariance matrices $\hat{\mathbf{P}}_{e_k}$, $\forall k$, can be found as

$$\hat{\mathbf{P}}_{e_k} = \frac{1}{N}\sum_{j=1}^{N}\mathbf{e}_k^{(j)}(\mathbf{e}_k^{(j)})^T \approx \mathrm{cov}[\mathbf{e}_k]. \tag{19}$$

Similarly the sample cross-correlation matrix

$$\hat{\mathbf{P}}_{e_l,e_k} = \frac{1}{N}\sum_{j=1}^{N}\mathbf{e}_l^{(j)}(\mathbf{e}_k^{(j)})^T \approx \mathrm{cov}[\mathbf{e}_l;\mathbf{e}_k] \tag{20}$$

and the sample covariance matrix of the mixture of particular prediction errors

$$\hat{\mathbf{P}}_e = \frac{1}{(n+1)\times N}\sum_{k=0}^{n}\sum_{j=1}^{N}\mathbf{e}_k^{(j)}(\mathbf{e}_k^{(j)})^T \approx \mathrm{cov}[\mathbf{e}] \tag{21}$$

can be computed. After substitution of the sample covariance or cross-correlation matrices (19)–(21) into relations (9), (10), (15), the estimates of the state and measurement noise covariance matrices $\hat{\mathbf{Q}}$ and $\hat{\mathbf{R}}$ can be found by means of the standard least square method. Therefore, it is also easily possible to compute a covariance matrix of estimated parameters representing elements of the system noise covariance matrices [Ljung, 1999]. The number of estimated elements of the noise covariance matrices and the fact, that each of equations (19)–(21) allows to determine $n_z(n_z+1)/2$ independent equations, leads to the minimal length of repeatedly measured data set n. It should be also noted that if the system initial covariance matrix $\mathbf{P}_0$ is unknown it can be analogously estimated as well.

In some cases it is nevertheless inevitable to estimate the noise covariance matrices on the basis of one measured realisation, i.e. $N=1$. This situation is considered in the following part.

3.4 Estimation by Single Realisation

For stable linear systems it is possible to estimate all elements of the noise covariance matrices on the basis of single measured data realisation because the sufficient number of independent equations can be set. Thus, suppose that single data set was measured for $(n+1)$ steps and prediction error sequence was computed, i.e. single sequence $[\mathbf{e}_0^{(1)}, \ldots, \mathbf{e}_n^{(1)}]$ is at disposal. Then, the sample covariance matrices, with respect to relations (17), (18), and (15), are given by

$$\hat{\mathbf{P}}_{e_{L_1}} = \frac{1}{n-L_1} \sum_{k=L_1}^{n} \mathbf{e}_k^{(1)} (\mathbf{e}_k^{(1)})^T \approx \mathrm{cov}[\mathbf{e}_{L_1}], \quad (22)$$

$$\hat{\mathbf{P}}_{e_{L_1},e_{L_2}} = \frac{1}{n-L_1} \sum_{k=L_1}^{n} \mathbf{e}_k^{(1)} (\mathbf{e}_{k+i}^{(1)})^T \approx \mathrm{cov}[\mathbf{e}_{L_1}; \mathbf{e}_{L_2}], \quad (23)$$

$$\hat{\mathbf{P}}_{e} = \frac{1}{(n+1)} \sum_{k=0}^{n} \mathbf{e}_k^{(1)} (\mathbf{e}_k^{(1)})^T \approx \mathrm{cov}[\mathbf{e}], \quad (24)$$

where $L_2 = L_1 + i$, $i = 1, 2, \ldots$. Analogously to previous part, estimates of matrices $\hat{\mathbf{Q}}$ and $\hat{\mathbf{R}}$ can be found by means of the least square method by substitution of the sample matrices into (9), (10), (15) with respect to (17), (18). Moreover, it can be easily seen that for the stable systems the requirement concerning the known system initial condition can be omitted. For unstable linear systems, relations (17) and (18) are not valid (powers of $\mathbf{F}^k$ do not become insignificant for increasing k) and thus one equation, namely (15), for estimating elements of $\mathbf{Q}$ and $\mathbf{R}$ is available which allows to determine $n_z(n_z+1)/2$ independent equations only.

3.5 Some Aspects of Proposed Technique

In the rest of the section three comments concerning the proposed technique are given. Firstly, the basic idea behind the proposed technique can be stated as follows: The predicted measurement $\hat{\mathbf{z}}_{k|-1}$ represents a "deterministic part" of the real measurement $\mathbf{z}_k$ and therefore the prediction error includes information about a "stochastic part" of the system (and thus information about the noise covariance matrices). Secondly, similar relations can be also derived by means of the KF with suitably chosen matrices $\mathbf{Q}^F$ and $\mathbf{R}^F$ for the KF algorithm. If $\mathbf{Q}^F \ll \mathbf{R}^F$ then the Kalman gain is insignificant and the KF works nearly as a multi-step predictor. In this case the variable $\mathbf{e}_k$ is referred as an innovation. This interpretation is interesting for comparison with other correlation techniques for estimating noise covariance matrices, such as the Adaptive Kalman Filter (AKF) [Mehra, 1970], from the theoretical point of view, but unfortunately it is necessary to set down more restriction on the system. Estimation of noise covariance matrices by means of the filter was discussed in detail in [Šimandl and Duník, 2007] and it will be also discussed in the section devoted to nonlinear systems. Thirdly, the proposed technique, except for estimation of noise covariance matrices of an unstable linear system on the basis of one measured realisation, allows to determine an "arbitrary" number of equations and thus to estimate all elements of $\mathbf{Q}$ and $\mathbf{R}$, contrary to the other correlation techniques, such as the AKF, which allows to estimate $n_x \times n_z$ elements of $\mathbf{Q}$ only.

4. ESTIMATION OF NOISE COVARIANCE MATRICES OF NONLINEAR SYSTEMS

The aim of this section is to extend the proposed technique given in the previous section for linear systems into the area of nonlinear systems. Whereas the technique for estimation of the noise covariance matrices was designed exactly for linear systems, for nonlinear systems it is necessary to employ some approximations in design of a multi-step predictor. Application of the approximations, however, can cause that the multi-step prediction error does not remain bounded. The simplest way to ensure the bounded prediction error for nonlinear systems is to exploit the results from the stability analysis of the local filters. Thus, the design of the estimation technique via the local filter will be preferred in this section. Mention that the estimation of the noise covariance matrices of the discrete-time nonlinear systems was initially motivated by the articles [Xiong et al., 2006, Wu et al., 2007, Xiong et al., 2007] where it was proved that the estimation error of the local filter remains bounded (and local filter is thus stable) if certain conditions are satisfied.

4.1 Extended Kalman Filter and its Innovation Sequence

The algorithms of the local filters have the same structure as the algorithm of the KF, where the first two moments of the state estimates are recursively computed. As a representative of these filters the Extended Kalman Filter (EKF) was chosen. The EKF is based on the linearisation of the nonlinear functions in the state and measurement equations (1), (2) by means of the first order Taylor expansion and can be written in the form where the filtering mean and covariance matrix are given as

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \mathbf{e}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}), \quad (25)$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{H}_k(\hat{\mathbf{x}}_{k|k-1}) \mathbf{P}_{k|k-1}, \quad (26)$$

and the predictive ones as

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{f}_k(\hat{\mathbf{x}}_{k|k}), \quad (27)$$

$$\mathbf{P}_{k+1|k} = \mathbf{F}_k(\hat{\mathbf{x}}_{k|k}) \mathbf{P}_{k|k} \mathbf{F}_k^T(\hat{\mathbf{x}}_{k|k}) + \mathbf{Q}^F, \quad (28)$$

$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T(\hat{\mathbf{x}}_{k|k-1})\big(\mathbf{H}_k(\hat{\mathbf{x}}_{k|k-1})\mathbf{P}_{k|k-1}\mathbf{H}_k^T(\hat{\mathbf{x}}_{k|k-1}) + \mathbf{R}^F\big)^{-1}$ is the Kalman gain, $\hat{\mathbf{z}}_{k|k-1} = \mathbf{h}_k(\hat{\mathbf{x}}_{k|k-1})$, $\mathbf{e}_k$ is the innovation, and $\mathbf{F}_k(\hat{\mathbf{x}}_{k|k}) = \frac{\partial \mathbf{f}_k(\mathbf{x}_k)}{\partial \mathbf{x}_k}|_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k}}$ and $\mathbf{H}_k(\hat{\mathbf{x}}_{k|k-1}) = \frac{\partial \mathbf{h}_k(\mathbf{x}_k)}{\partial \mathbf{x}_k}|_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k-1}}$ are Jacobians of the nonlinear functions $\mathbf{f}_k(\cdot)$ and $\mathbf{h}_k(\cdot)$ in the best available state estimates $\hat{\mathbf{x}}_{k|k}$ and $\hat{\mathbf{x}}_{k|k-1}$, respectively. As far as the stability is concerned, the EKF, as a local filter, will be stable if the system (1), (2) is observable, the initial conditions of the system and filter are sufficiently close, and the filter noise covariance matrices are greater than true ones, i.e. $\mathbf{Q}^F > \mathbf{Q}$, $\mathbf{R}^F > \mathbf{R}$ [Xiong et al., 2006, Wu et al., 2007, Xiong et al., 2007]. Thus, estimation of the noise covariance matrices is, for nonlinear systems, conditioned by exact knowledge of the system initial condition, i.e. $\hat{\mathbf{x}}_{0|-1} = \mathbf{x}_0$, which allows to set small filter initial covariance matrix, i.e. $\mathbf{P}_{0|-1} = \delta\mathbf{I}$, where $\delta \to 0$ and $\mathbf{I}$ is identity matrix.

Following the basic idea of the noise covariance matrices estimation technique, given in Section 3.5 for linear systems, it is necessary to make the EKF to work as a multi-step predictor with bounded prediction error. This can be achieved by the suitable Kalman gain $\mathbf{K}_k$ which depends on chosen $\mathbf{Q}^F$ and $\mathbf{R}^F$. Thus, impact of the chosen matrices $\mathbf{Q}^F$ and $\mathbf{R}^F$ on the Kalman gain and statistics of the innovation sequence are given below.

The mean of the innovation sequence $\mathbf{e}_k$, $\forall k$, of the EKF is zero for an arbitrary choice of the filter noise covariance matrices meeting conditions $\mathbf{Q}^F > \mathbf{Q}$ and $\mathbf{R}^F > \mathbf{R}$ [Šimandl and Duník, 2007]. Then, let matrix $\mathbf{R}^F$ be chosen "sufficiently large" which causes the insignificant Kalman gain $\mathbf{K}_0$ and thus, the elimination of the filtering step (25), (26) in the EKF algorithm. The covariance matrices of the innovations $\mathbf{e}_0$ and $\mathbf{e}_1$ can be approximated by following relations

$$\begin{aligned}\mathrm{cov}[\mathbf{e}_0] &\approx \mathrm{cov}[\mathbf{h}(\mathbf{x}_0) + \mathbf{v}_0 - \mathbf{h}(\hat{\mathbf{x}}_{0|-1})] \approx \\ &\approx \mathbf{H}_0(\hat{\mathbf{x}}_{0|-1})\mathbf{P}_{0|-1}\mathbf{H}_0^T(\hat{\mathbf{x}}_{0|-1}) + \mathbf{R} \approx \mathbf{R}, \qquad (29)\end{aligned}$$

$$\begin{aligned}\mathrm{cov}[\mathbf{e}_1] \approx E\Big[&\Big(\mathbf{H}_1(\hat{\mathbf{x}}_{1|0})\mathbf{F}_0(\hat{\mathbf{x}}_{0|0})(\mathbf{x}_0 - \hat{\mathbf{x}}_{0|-1}) + \qquad (30)\\ &+ \mathbf{H}_1(\hat{\mathbf{x}}_{1|0})\mathbf{w}_0 + \mathbf{v}_1\Big)\Big(\mathbf{H}_1(\hat{\mathbf{x}}_{1|0})\mathbf{F}_0(\hat{\mathbf{x}}_{0|0})(\mathbf{x}_0 - \hat{\mathbf{x}}_{0|-1}) + \\ &+ \mathbf{H}_1(\hat{\mathbf{x}}_{1|0})\mathbf{w}_0 + \mathbf{v}_1\Big)^T\Big] = \mathbf{H}_1(\hat{\mathbf{x}}_{1|0})\mathbf{Q}\mathbf{H}_1^T(\hat{\mathbf{x}}_{1|0}) + \mathbf{R},\end{aligned}$$

where $\mathbf{f}_k(\mathbf{x}_k) \approx \mathbf{f}_k(\hat{\mathbf{x}}_{k|k}) + \mathbf{F}_k(\hat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})$ and $\mathbf{h}_k(\mathbf{x}_k) \approx \mathbf{h}_k(\hat{\mathbf{x}}_{k|k-1}) + \mathbf{H}_k(\hat{\mathbf{x}}_{k|k-1})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})$, $k = 0, 1$. Analogously to the linear system and relation (10) the matrix $\mathrm{cov}[\mathbf{e}_0; \mathbf{e}_1]$ can be computed which allows, together with (29) and (30), to determine $\frac{3n_z(n_z+1)}{2}$ independent equations for estimation elements of matrices $\mathbf{Q}$ and $\mathbf{R}$. The covariance matrix $\mathrm{cov}[\mathbf{e}]$, as the mixture of $\mathbf{e}_0$ and $\mathbf{e}_1$, can be computed as well, but it is linearly dependent on the matrices $\mathrm{cov}[\mathbf{e}_0]$ and $\mathrm{cov}[\mathbf{e}_1]$. Note that the covariance matrices of $\mathbf{e}_0$ and $\mathbf{e}_1$ were determined on assumptions of small $\mathbf{P}_{0|-1}$ and large $\mathbf{R}^F$ only and they do not depend on the chosen $\mathbf{Q}^F$. Thus, $\mathbf{Q}^F$ can be chosen sufficiently large as well, i.e. the inequality $\mathbf{Q}^F > \mathbf{Q}$ surely holds, to ensure the stability of the local filter. However, if it is necessary the additional equations can be determined for example from

$$\begin{aligned}\mathrm{cov}[\mathbf{e}_k] \approx \sum_{i=1}^{k-1}\mathbf{H}_k(\hat{\mathbf{x}}_{k|k-1})\prod_{j=i}^{k-1}\mathbf{F}_m(\hat{\mathbf{x}}_{m|m})\mathbf{Q}\Big(\mathbf{H}_k(\hat{\mathbf{x}}_{k|k-1}) \times \\ \times \prod_{j=i}^{k-1}\mathbf{F}_m(\hat{\mathbf{x}}_{m|m})\Big)^T + \mathbf{H}_k(\hat{\mathbf{x}}_{k|k-1})\mathbf{Q}\mathbf{H}_k^T(\hat{\mathbf{x}}_{k|k-1}) + \mathbf{R}, \qquad (31)\end{aligned}$$

where $m = k - j + i - 1$, $k = 2, 3 \ldots$. Unfortunately, for general time instant, matrix $\mathbf{Q}^F$ should be chosen sufficiently small to ensure negligible Kalman gain. This choice of $\mathbf{Q}^F$ can cause the instability of the local filter as will be illustrated in the numerical example. The alternative way is to keep sufficiently large $\mathbf{Q}^F$ and to take "neglected" terms into account which significantly increases the computational demands.

4.2 Estimation of Noise Covariance Matrices

With respect to the stability of the local filter the set of independent equations allowing estimation of the noise covariance matrices of nonlinear systems can be determined. The set of equations is based on the covariance matrices of $\mathbf{e}_k$ for few first time instants and thus the estimation of the noise covariance matrices will be based on the multiple measured short data sets. On the basis of these sets the sample covariance and cross-correlation matrices of the innovation sequence can be computed and the noise covariance matrices can be estimated along the same line as in the linear case (see Section 3.3). However, the key difference from the linear case is that the linearised matrices can be generally different for each filter run due to the different linearisation points. To ensure the unvarying linearised system matrices the constrain on the known and constant system initial condition, by the small initial covariance matrix, was set. Then, with proper choice of $\mathbf{R}^F$ (and $\mathbf{Q}^F$) it surely holds that $\mathbf{F}_k^{(i)}(\cdot) = \mathbf{F}_k^{(j)}(\cdot)$ and $\mathbf{H}_k^{(i)}(\cdot) = \mathbf{H}_k^{(j)}(\cdot)$, $i, j = 1, \ldots, N$.

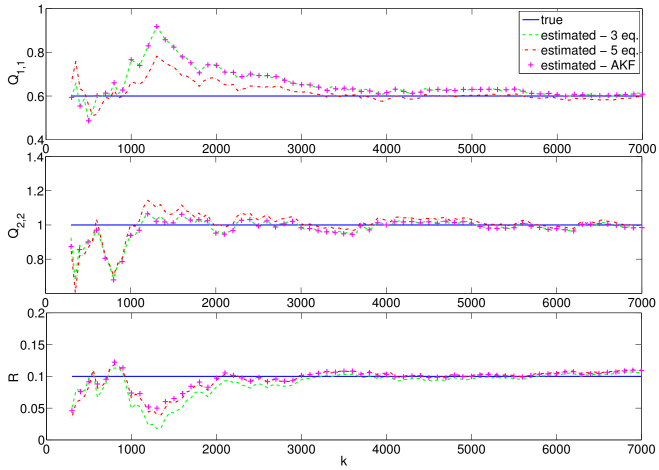

Fig. 1. Theoretical and estimated elements of system noise covariance matrices on the basis of one measured data set.

5. NUMERICAL ILLUSTRATIONS

Example 1: As the first example of estimation of the noise covariance matrices on basis of one measured data set, the linear stable system, defined by

$$\mathbf{x}_{k+1} = \begin{bmatrix} -0.8 & 0.9 \\ 0.1 & 0.5 \end{bmatrix}\mathbf{x}_k + \mathbf{w}_k, \qquad (32)$$

$$z_k = [0.4 \ \ 0.1]\mathbf{x}_k + v_k, \qquad (33)$$

was chosen, where $k = 0, \ldots, 7000$. Both noises were described by the normal distribution with $p(\mathbf{w}_k) = \mathcal{N}\{\mathbf{w}_k : [0, 0]^T, \mathrm{diag}[0.6, 1]\}$ and $p(v_k) = \mathcal{N}\{v_k : 0, 0.1\}$, $\forall k$. The initial condition was given as $p(\mathbf{x}_0) = p(\mathbf{x}_0|z^{-1}) = \mathcal{N}\{\mathbf{x}_0 : [20, 20]^T, \mathbf{I}\}$ and the covariance matrix $\mathbf{Q}^P = \mathbf{I}$. The function diag[$\mathbf{x}$] stands for diagonal matrix with vector $\mathbf{x}$ on diagonal.

The aim is to estimate the diagonal elements of the matrix $\mathbf{Q}$ and R. Thus, at least three equations has to be set, e.g. equations for $\hat{P}_{e_{L_1}}$ (22), $\hat{P}_{e_{L_1}, e_{L_1+1}}$, and $\hat{P}_{e_{L_1}, e_{L_1+2}}$ (23), where $L_1 = 200$. Arbitrary number of additional equations can be further set to refine the estimates, e.g. $\hat{P}_{e_{L_1}, e_{L_1+3}}$ (23) and $\hat{P}_{e_{0:7000}}$ (24). The resultant estimates based on the three and five equations can be found in Fig. 1. These estimates are compared with those given by the Adaptive Kalman Filter with the initial estimates of noise covariance matrices $\mathbf{Q}^F = \mathbf{I}$ and $\mathbf{R}^F = 1$. Contrary to the proposed technique, the AKF allows to determine a limited set of equations only and thus to estimate a limited number elements of the matrix $\mathbf{Q}$ (in this case 2 elements only.)

Example 2: In the second example, the nonlinear Gaussian system, described by

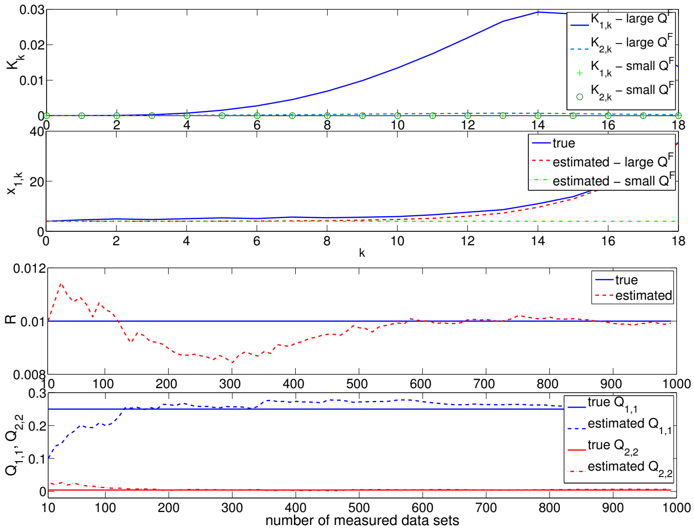

Fig. 2. Kalman gain, true and estimated state for different choices of matrix $\mathbf{Q}^F$, and true and estimated noise matrices for increasing number of measured data sets.

$$x_{1,k+1} = x_{1,k}x_{2,k} + w_{1,k}, \tag{34}$$
$$x_{2,k+1} = x_{2,k} + w_{2,k}, \tag{35}$$
$$z_k = x_{1,k}^2 + v_k, \tag{36}$$

is considered, where $k = 0, \ldots, 18$, $p(\mathbf{w}_k) = \mathcal{N}\{\mathbf{w}_k : [0,0]^T, \text{diag}[0.25, 0.004]\}$, and $p(v_k) = \mathcal{N}\{v_k : 0, 0.01\}$, $\forall k$. The initial condition is supposed to be $p(\mathbf{x}_0) = p(\mathbf{x}_0|z^{-1}) = \mathcal{N}\{\mathbf{x}_0 : [40, 0.95]^T, 10^{-8} \times \mathbf{I}\}$. As a state estimator the EKF was used with $R^F = 10^8$. Matrix $\mathbf{Q}^F$ was chosen "large" ($\mathbf{Q}^F = 40 \times \mathbf{I}$) to ensure stability of the filter and "small" ($\mathbf{Q}^F = 10^{-8} \times \mathbf{I}$) to ensure the negligible Kalman gain.

An example of the Kalman gain and state estimate from one particular run of the EKF with different choice of matrix $\mathbf{Q}^F$ is shown in Fig. 2. It can be easily seen that large $\mathbf{Q}^F$ ensures the stability of the EKF but it causes the significant Kalman gain and thus some terms cannot be neglected during derivation of (31). The insignificant Kalman gain is reached by the small $\mathbf{Q}^F$ which can, however, cause the instability of the local filter (the prediction error is not bounded and thus the nonlinear system is not linearised in the point of the state space close to the true system state). Nevertheless, note that for a first few time instants, namely for first four instants, the Kalman gain is negligible even for large $\mathbf{Q}^F$. In Fig. 2 the estimates of R and diagonal elements of $\mathbf{Q}$ for increasing number of the measured data sets are shown as well. These estimates were computed on the basis of cov$[e_0]$, cov$[e_1]$, and cov$[e_0; e_1]$ and large $\mathbf{Q}^F$.

6. CONCLUSION

The paper dealt with the off-line estimation of the state and measurement noise covariance matrices of linear and nonlinear dynamic systems. The novel technique for estimation of the noise covariance matrices of linear systems was developed for both situations when several data sets are available or single data set is available. The technique allows to determine a sufficient number of independent equations to estimate all elements of the system noise covariance matrices, contrary to other correlation methods such as the Adaptive Kalman Filter. The estimation of the noise covariance matrices was also discussed for a general nonlinear system with stress on the stability of the local filters. The noise covariance matrices estimation technique, based on the multiple measured data sets, was derived for the Extended Kalman Filter as an representative of the local filters.

REFERENCES

R. Bos, X. Bombois, and P. M. J. Van den Hof. Designing a Kalman filter when no noise covariance information is available. In *Proceedings of the 16th IFAC World Congress*, Prague, Czech Republic, July 2005.

J. Duník, M. Šimandl, O. Straka, and L. Král. Performance analysis of derivative-free filters. In *Proceedings of the 44th IEEE Conference on Decision and Control, and European Control Conference ECC'05*, pages 1941–1946, Seville, Spain, December 2005.

J. Homolová and I. Nagy. Traffic model of a microregion. In *Proceedings of the 16th IFAC World Congress*, Prague, Czech Republic, July 2005.

S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–421, 2004.

F. L. Lewis. *Optimal Estimation*. John Wiley & Sons, New York, 1986.

L. Ljung. *System Identification: Theory for the User*. Upper-Saddle River, NJ: Prentice-Hall, 1999.

R. K. Mehra. On the identification of variances and adaptive filtering. *IEEE Trans on AC*, 15(2):175–184, 1970.

R. K. Mehra. Approaches to adaptive filtering. *IEEE Transactions on Automatic Control*, 17(10):693–698, 1972.

M. Nørgaard, N. K. Poulsen, and O. Ravn. New developments in state estimation for nonlinear systems. *Automatica*, 36 (11):1627–1638, 2000.

H. J. Palanthandalam-Madapusi, S. Lacy, J. B. Hoagg, and D. S. Bernstein. Subspace-based identification for linear and nonlinear systems. In *Proceedings of the American Control Conference 2005*, pages 2320–2334, USA, June 2005.

A. Papoulis and S. U. Pillai. *Probability, Random Variables and Stochastic Processes*. Mc Graw Hill, fourth edition, 2002.

H. W. Sorenson. On the development of practical nonlinear filters. *Inf. Sci.*, 7:230–270, 1974.

S. Verdú and H. V. Poor. Minimax linear observers and regulators for stochastic systems with uncertain second-order statistics. *IEEE Trans on AC*, 29(6):499–511, 1984.

M. Šimandl and J. Duník. Off-line estimation of system noise covariance matrices by a special choice of the filter gain. In *Proceedings of the IEEE International Symposium on Intelligent Signal Processing*, Alcalá de Henares, Spain, October 2007.

M. Šimandl and O. Straka. Nonlinear filtering methods: Some aspects and performance evaluation. In *Proceedings of the IASTED International Conference On Modelling, Identification and Control*, Innsbruck, February 2003.

E. A. Wan and A. T. Nelson. *Dual Extended Kalman Filter Methods*, pages 123–163. Wiley Publishing, Eds. S. Haykin, 2001.

Y. Wu, D. Hu, and X. Hu. Comments on "Performance evaluation of UKF-based nonlinear filtering". *Automatica*, 43(3):567–568, 2007.

K. Xiong, H. Y. Zhang, and C. W. Chan. Performance evaluation of UKF-based nonlinear filtering. *Automatica*, 42(2): 261–270, 2006.

K. Xiong, H. Y. Zhang, and C. W. Chan. Author's reply to "Comments on 'Performance evaluation of UKF-based nonlinear filtering"'. *Automatica*, 43(3):569–570, 2007.