Noise Covariance Matrices Estimation for Systems with Time-Varying Availability of Sensors

Oliver Kost, Jindřich Duník, and Ondřej Straka
European Centre of Excellence - NTIS & Department of Cybernetics
Faculty of Applied Sciences, University of West Bohemia
Univerzitní 8, 306 14 Pilsen, Czech Republic
E-mail: kost@ntis.zcu.cz, dunikj@kky.zcu.cz, straka30@kky.zcu.cz

Abstract. The paper deals with the estimation of the noise covariance matrices of a time-varying system described by a state-space model with a time-varying set of sensors. In particular, the measurement difference autocovariance method, is proposed. The method is based on a statistical analysis of linearity transformed measurements leading to a system of linear matrix equations. The theoretical results are discussed and illustrated using a numerical study.

Keywords: Noise covariance matrices; State-space models; Estimation; Identification; Least-square method; Positioning

1. Introduction
State estimation plays a key role in many applications including satellite navigation, target tracking, signal processing, fault detection, and adaptive and optimal control problems. It constitutes an essential part of any decision-making process. Estimation for discrete-time stochastic dynamic systems is a very important field of study, which has been considerably developed for the past fifty years. In the beginning, the development was focused on linear dynamic systems and it has been further developed until today with the stress on the state estimation of nonlinear stochastic dynamic systems.

The state estimation, however, requires the knowledge of not only the system matrices or functions of the deterministic part of the model but also the knowledge of the noise statistics. The knowledge of noise statistics is, however, questionable in many cases. Incorrect description of the noise statistics may cause significant worsening of estimation or control quality or even failure of the underlying algorithms.

Therefore, several methods for estimation of the noise statistics have been developed, which can be divided into several categories; correlation methods [1–6], Bayesian estimation methods [7,8], maximum likelihood estimation methods [9,10], covariance matching methods [11], or the Kalman filter working as a parameter estimator [12]. Besides the noise covariance matrix estimation methods, alternative approaches directly estimating the gain of a linear estimator have been developed as well [1,13,14]. A characterisation of the methods with their assumptions, properties, and limitations can be found in e.g., [3,15–17].

The correlation methods have received quite considerable attention in the past as they may provide unbiased estimates with acceptable computational requirements even for high-dimensional systems. The methods are based on an analysis of the innovation sequence of “non-optimal” linear state and
measurement predictors\(^1\). The methods have been pioneered in \([1, 2]\) for linear time-invariant (LTI) and linear time-varying (LTV) models respectively. In \([3–5, 18, 19]\) the correlation methods have been further advanced (in terms of computation and algorithmic complexity) and a set of methods, called the autocovariance least-squares (ALS) methods have been proposed. The ALS methods were developed for both linear and nonlinear systems. However, for LTI models and only for certain classes of LTV and nonlinear models, the methods have been proven to provide \textit{asymptotically} unbiased estimates \([3, 20]\). Moreover, the above mentioned methods have been proposed for systems with a constant set of sensors. The assumption is, however, limiting for certain applications and areas, for example for navigation, tracking, and positioning, where the number of sensors or their availability varies over time \([21]\).

In this paper, the measurement difference autocovariance (MDA) least-squares-based method \([22]\), which was originally proposed for estimation of the noise covariance matrices for LTV models described by a state-space model, with a constant set of sensors, is extended for a state-space model, with linear state equation and linear or nonlinear measurement equation with a time-varying set of sensors. In particular, the models with the measurement function Jacobian with the rank greater or equal to the state vector dimension are considered. The proposed method is thoroughly analysed and conditions for \textit{unbiased} estimate are discussed.

The rest of the paper is organised as follows. In Section 2, the system description and the problem statement are given. The MDA method is introduced and extended in Section 3. Section 4 discusses implementation details of the MDA method. Then, in Section 5 the performance of the MDA method is illustrated using a set of simulations and Section 6 concludes the paper.

2. System description and goal of paper

Let a discrete-time time-varying dynamic stochastic system described by the state-space model with a time-varying set of available sensors be considered. Its model consists of the state and measurement equations. The state equation is given by

\[
x_{k+1} = F_k x_k + G_k u_k + w_k, \quad k = 0, 1, 2, \ldots, T,
\]

where the vectors \(x_k \in \mathbb{R}^{n_x}\) and \(u_k \in \mathbb{R}^{n_u}\) represent the system state and control at the time instant \(k\), respectively. The vector \(w_k \in \mathbb{R}^{n_x}\) is the state noise, which is supposed to be zero-mean and white with \textit{unknown} covariance matrix \(Q \in \mathbb{R}^{n_x \times n_x}\). The measurement equations corresponding to \(n_s\) available sensors are given by

\[
z_{k,i} = h_{k,i}(x_k) + v_{k,i}, \quad i = 1, 2, 3, \ldots, n_s.
\]

where, \(z_{k,i} \in \mathbb{R}^{n_z,i}\) is generally a vector measurement of the \(i\)-th sensor, \(h_{k,i}(x_k)\) is a known function, and \(v_{k,i} \in \mathbb{R}^{n_z,i}\) is the measurement noise with the following properties

\[
\mathbb{E}[v_{k,i}] = 0_{n_z,i \times 1}, \quad \forall k, i = 1, 2, 3, \ldots, n_s
\]

\[
\text{cov}[v_{k,i}, v_{l,j}] = \delta_{k,l} R_{i,j}, \quad i = 1, 2, 3, \ldots, n_s, j = 1, 2, 3, \ldots, n_s
\]

where \(\delta_{k,l}\) denotes the Kronecker delta function and \(0_{n \times m}\) stands for the zero matrix of dimension \(n \times m\).

The set of the sensor indices is \(Z_k^C = \{1, 2, 3, \ldots, n_s\}\), and the measurement vector corresponding to all sensors is given

\[
z_k = [z_{k,1}^T, z_{k,2}^T, \ldots, z_{k,n_s}^T]^T \in \mathbb{R}^{n_z}.
\]

where \(n_z = \sum_{i=1}^{n_s} n_z,i\).

\(^1\) As the noise covariance matrices are not known, the optimum (Kalman) gain cannot be computed. The predictor gain is, therefore, considered as a user-defined parameter, hence the prediction is not optimal in mean-square error.
At each time instant, however, a only subset $Z_k \subset Z_k^c$ of all possible sensors is available. The set of available sensors $Z_k$ corresponds to the measurement vector

$$z_k = S_k \bar{z}_k,$$  \hspace{1cm} (6)

where $S_k \in \mathbb{R}^{n_z \times n_z}$ is a selection matrix defined as

$$S_k \triangleq \text{blkdiag}(S_{k,1}, S_{k,2}, \ldots S_{k,n_z}),$$  \hspace{1cm} (7)

and

$$S_{k,i} = \begin{cases} I_{n_z,i} & \text{if } i\text{-th sensor is available at time instant } k, \text{ i.e., } i \in Z_k \\ 0_{n_z,i \times n_z,i} & \text{otherwise.} \end{cases}$$  \hspace{1cm} (8)

In (7), the function $\text{blkdiag}(S_{k,1}, S_{k,2}, \ldots S_{k,n_z})$ denotes a block diagonal matrix and the notation $I_n$ stands for the identity matrix of dimension $n$. Hence, the measurements corresponding to the unavailable sensors are zeroed by the selection matrix. The measurement equation at time $k$ is then given by

$$z_k = S_k \bar{z}_k = S_k (h_k(x_k) + v_k), \quad k = 0, 1, \ldots, T,$$  \hspace{1cm} (9)

where $z_k \in \mathbb{R}^{n_z}$ and $v_k \in \mathbb{R}^{n_z}$. It is also assumed that $\text{trace}(S_k) \geq n_z, \forall k$, and the matrices $F_k \in \mathbb{R}^{n_x \times n_x}$ and $G_k \in \mathbb{R}^{n_x \times n_u}$ are known matrices of the system and control, respectively, and $h_k(x_k)$ is a known measurement function as well as the selection matrix $S_k$. The measurement function is supposed to be a $C^1$ function i.e., the first order derivatives exist and are continuous. Further, it is assumed that the Jacobian$^2$ of the measurement function has full-rank i.e., its rank is equal to $n_z$. The state and measurement noises are assumed to be mutually independent and independent of the system initial condition. The moments of the initial state are not supposed to be known. The measurement noise is also supposed to be zero-mean and white with unknown covariance matrix $R \in \mathbb{R}^{n_z \times n_z}$ considering of $R_{i,j}$.

The goal of this paper is to estimate the unknown noise covariance matrices $Q$ and $R$ on the basis of the known system and control matrices $F_k$, $G_k \forall k$, system measurement function $h_k$, selection matrix $S_k \forall k$, and control sequences.

Note that such system description typically appears in tracking applications (e.g., global navigation satellite system based positioning), where the set of sensors (e.g., visible satellites) is varying over the time [23, 24].

### 3. Derivation of measurement difference autocovariance method

The MDA method, is based on a statistical analysis of the difference between the measurement and a predicted measurement, i.e., on the analysis of the measurement prediction error. The measurement prediction is computed directly on the basis of the past measurements.

In this section, the MDA method for estimation of the noise covariance matrices $Q$ and $R$ is proposed for the model (1), (9) with a linear or nonlinear measurement equation with a time-varying set of available sensors.

#### 3.1. MDA for linear system with linear measurement

In this section, the MDA method is introduced for the linear model described by (1) and (9) with $h_k(x_k) = H_k x_k$.

$^2$ The Jacobian $H_k(x_k) = \frac{\partial h_k(x_k)}{\partial x_k}|_{x_k=x_k}$ is the first partial derivative with respect to $x_k$ of function $h_k(x_k)$ evaluated at $x_k$. 

3.1.1. Measurement prediction error definition

The measurement prediction error (MPE), is defined as

$$r_k^{(q)} \triangleq z_k - \hat{z}_k^{(q)}, \quad k = q, \ldots, T,$$

(10)

where $z_k$ is the measurement (9) and $\hat{z}_k^{(q)}$ is the $q$-step measurement prediction of $z_k$ based on the measurement $z_{k-q}$. The prediction is computed as

$$\hat{z}_k^{(q)} = S_k H_k \left( \prod_{i=1}^{q} F_{k-i} \right) (S_{k-q} H_{k-q})^T z_{k-q} = \mathcal{H}_k^{(q)} z_{k-q},$$

(11)

where $(S_k H_k)^\dagger = (H_k^T S_k^T S_k H_k)^{-1} H_k^T S_k$ is the pseudoinverse of the matrix $(S_k H_k)$ and $\mathcal{H}_k^{(q)} = S_k H_k \left( \prod_{i=1}^{q} F_{k-i} \right) (S_{k-q} H_{k-q})^\dagger$. Compared to other correlation methods it can be seen that the measurement prediction is computed from the past measurements only without an explicit need of a state prediction.

For further analysis, the MPE (10) can be rewritten by the use of (1) and (9) to the form

$$r_k^{(q)} = S_k (H_k x_k + v_k) - S_k H_k \left( \prod_{i=1}^{q} F_{k-i} \right) (S_{k-q} H_{k-q})^T z_{k-q} = S_k h_k v_k - S_k H_k \left( \prod_{i=1}^{q} F_{k-i} \right) x_{k-q} - \mathcal{H}_k^{(q)} S_{k-q} v_{k-q},$$

(12)

where $A_k^{(i)} = S_k H_k \prod_{j=1}^{i} F_{k-j}$.

3.1.2. Measurement prediction error properties

It is easily be shown that the MPE is a stochastic process with the mean

$$E[r_k^{(q)}] = \sum_{i=0}^{q-1} A_k^{(i)} G_{k-i-1} u_{k-i-1}$$

(13)

and with the autocovariance function

$$C_{k,p}^{(q)} = E[r_k^{(q)}(r_{k-p}^{(q)})^T] = \begin{cases} \sum_{i=0}^{q-1} \left( A_k^{(i)} Q \left( A_k^{(i)} \right)^T \right) + S_k R S_k^T + \mathcal{H}_k^{(q)} S_{k-q} R (\mathcal{H}_k^{(q)} S_{k-q})^T, & \text{for } p = 0, \\ \sum_{i=0}^{q-p-1} \left( A_k^{(i+p)} Q \left( A_k^{(i+p)} \right)^T \right), & \text{for } p = 1, \ldots, q - 1, \\ -\mathcal{H}_k^{(q)} S_{k-q} R S_{k-q}^T, & \text{for } p = q, \\ 0_{n_x}, & \text{otherwise}, \end{cases}$$

(14)

where the difference $\tilde{r}_k^{(q)}$ is given by

$$\tilde{r}_k^{(q)} = r_k^{(q)} - E[r_k^{(q)}].$$

(15)

It means that the autocovariance function is fully defined by $q + 1$ (cross-)covariance matrices of $r_k^{(q)}$.  

The equation (14) is a linear function of the unknown matrices \( Q \) and \( R \), and thus it can be rewritten in the following form suitable for the least-square (LS) method formulation

\[
A_k \theta_k = b_k,
\]

where the parameter vector \( \theta_k \), containing the unknown elements of the noise covariance matrices, is given by

\[
\theta_k = \Gamma_k [Q_s^T, R_s^T]^T,
\]

and the matrices \( A_k, b_k \) are defined as

\[
A_k = \left[ \begin{array}{c}
\sum_{i=0}^{q-1} A_k^{(i)} \otimes A_k^{(i)} \\
\sum_{i=0}^{q-2} A_k^{(i)} \otimes A_k^{(i+1)} \\
\vdots \\
A_k^{(q-1)} \otimes A_k^{(q-1)} \otimes A_k^{(q-2)} \\
0_{2n^s \times n_x} \\
- \mathcal{H}_k^{(1)} \otimes \mathcal{H}_k^{(1)} S_{k-q} \\
\end{array} \right],
\]

\[
b_k = \left[ \begin{array}{c}
(C_{k,0}^{(q)})^T \\
(C_{k,1}^{(q)})^T \\
\vdots \\
(C_{k,q}^{(q)})^T \\
\end{array} \right].
\]

The symbol \( \otimes \) denotes the Kronecker product, the notation \( (A)_s \) means the columnwise stacking of the matrix \( A \) into a vector\(^3\) [25] and the notation \( A_k^T \) means \( ((A)_s)^T \). The diagonal matrix \( \Gamma_k \in \mathbb{R}^{n^s_2 + n^s_2 \times n^s_2 + n^s_2} \) with ones and zeros on the diagonal defines which elements \( Q \) and \( R \) are estimable and which are not. The matrix \( Q \) is always estimable but estimability of the matrix \( R \) depends on the matrices \( S_i, i = \{k-q, k\} \) as

\[
\Gamma_k = \text{blkdiag} \left( I_{n^s_2}, \max(S_k \otimes S_k, S_{k-q} \otimes S_{k-q}) \right)
\]

where the element-wise maximum of the matrices is consider.

### 3.1.3. Noise covariance matrices estimation

Estimation of \( \theta_k \) is conditioned by known \( A_k \) and \( b_k \). The matrix \( A_k \) is a function of the known model matrices and therefore, can be computed. The vector \( b_k \) is unknown because it is a function of the unknown matrices \( Q \) and \( R \) due to (14). Nevertheless, the vector \( b_k \) can be estimated on the basis of the known difference \( \hat{r}^{(q)} \) (15) as follows.

The estimate \( \hat{C}^{(q)}_{k,i} \) of \( C^{(q)}_{k,i} \) is given by

\[
\hat{C}^{(q)}_{k,i} = \hat{r}^{(q)}_{k-i} (\hat{r}^{(q)}_{k-i})^T, \quad \forall k, \forall i,
\]

and the estimate (21) is an unbiased estimate of \( C^{(q)}_{k,i} \) (14), i.e.,

\[
C^{(q)}_{k,i} = E[\hat{C}^{(q)}_{k,i}] = E[\hat{r}^{(q)}_{k-i} (\hat{r}^{(q)}_{k-i})^T], \quad \forall k, \forall i,
\]

The unbiased estimate of \( b_k \) (19) is then given by

\[
b_k = \left[ \left( \hat{C}^{(q)}_{k,0} \right)^T \left( \hat{C}^{(q)}_{k,1} \right)^T, \ldots, \left( \hat{C}^{(q)}_{k,q} \right)^T \right]^T.
\]
The optimum LS estimate of the parameter $\theta_k$ in (16) is

$$\hat{\theta}_k = A^T_k \hat{b}_k,$$

which is unbiased, i.e.,

$$E[\hat{\theta}_k] = A^T_k E[b_k] = A^T_k A_k \theta_k = \theta_k.$$  

(24)

In the case that $\Gamma_k$ (20) is time-invariant, i.e., the set of sensors is constant over time, the averaging of equation (24) over all time instants, i.e., over $k = 2q, \ldots, T$, leads to the resulting estimate

$$\hat{\theta}^*_k = (\bar{A}(T))^T \bar{b}(T),$$

(26)

with

$$\bar{A}(T) = \Gamma [Q_s^T, R_s^T]^T$$

(27)

$$\bar{b}(T) = \frac{1}{T-2q+1} \sum_{k=2q}^{T} A_k$$

(28)

$$\hat{b}_k = \frac{1}{T-2q+1} \sum_{k=2q}^{T} \hat{b}_k,$$

(29)

where $\bar{A}(T)$ (28) is an average of the known matrices $A_k$ up to the time $T$ and $\bar{b}(T)$ (29) can be interpreted as an average of the unbiased estimates $b_k$ up to the time $T$. It means that the estimate $\hat{\theta}^*_k$ (26) is unbiased as well.

As is derived in Appendix, the estimate (26) can be alternatively computed according to the following recursive relations

$$\hat{\theta}^*_k = \hat{\theta}^*_{k-1} + (\bar{A}(k))^T (\hat{b}_k - A_k \hat{\theta}^*_{k-1}),$$

(30)

$$\bar{A}(k) = \frac{1}{k-2q+1} \sum_{i=2q}^{k} A_i = \bar{A}(k-1) + A_k,$$

(31)

for $k = 2q, \ldots, T$, where $A_{(i)} = 0_{(n^2+iq(x+i)) \times (n^2+i+1)}$ and $\hat{\theta}^*_{(i)} = 0_{(n^2+i+1)}$ for $i = 0, \ldots, 2q - 1$.

In the case that $\Gamma_k$ (20) is varying in the time i.e., the set of sensors is time-varying, the estimate (30) is simply modified to the following equation

$$\hat{\theta}^*_k = \hat{\theta}^*_{k-1} + (\bar{A}(k) \Gamma_k)^T (\hat{b}_k - A_k \hat{\theta}^*_{k-1}),$$

(32)

where the matrix $\bar{A}(k)$ is still computed according to (31). Equation (32) compared to (30) respects availability or unavailability of the elements in the considered interval, i.e., only the available elements are estimated and updated.

### 3.1.4. Illustration of MDA for Linear Systems

To illustrate the basic idea of the proposed MDA method, a simple example is given. Consider a system described by (1) and (9) with $G_k = 0_{n_x \times n_u}$ and $h_k(x_k) = H_k x_k$, where the set of sensors varies in time

$$x_{k+1} = F x_k + w_k = 0.9 x_k + w_k, k = 0, 1, 2, 3, 4,$$

(33)

$$z_k = S_k z_k = S_{k(k+1)} = S_k (H_k x_k + \nu_k) = S_k \left( \begin{bmatrix} 1 \\ 2 \end{bmatrix} x_k + \begin{bmatrix} \nu_{k,1} \\ \nu_{k,2} \end{bmatrix} \right),$$

(34)

where $R = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $Q = 0.5$.

The availability of the sensors is as follows:
from the past measurement as chosen in accord with the a priori knowledge of the state or the estimate can be obtained as a prediction \( \hat{x}_k \), where

\[
\hat{x}_k = A_k x_{k-1} + B_k u_{k-1},
\]

with

\[
A_k = \begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
40 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]

and

\[
B_k = \begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]

The optimum LS estimation parameter for matrices (36) can be computed by (26), (30), or (32) where \( \hat{\theta}_{2q}^r \) is this case equal to \( [Q \ 0 \ 0 \ 0 \ R_{(2,2)}]^{T} \) where \( A_{(i,j)} \) denotes the element positioned in \( i \)-th row and \( j \)-th column of the matrix \( A \). The elements \( R_{(1,1)} \), \( R_{(2,1)} \), and \( R_{(1,2)} \) are zero because the sensor 1 is not available at time instants \( k = 0, 1, 2 \). The matrices \( A_k \) and \( b_k \) can be computed at \( k = 2 \) because the matrix \( b_2 \) is a function of \( \{z_k\}_{k=0}^2 \), i.e., \( A_2 = A_2 \).

The vector \( \{\hat{\theta}_{2q}^r\}_{k=3}^4 \) (32) is equal to \( [Q \ R_{(1,1)} \ 0 \ 0 \ R_{(2,2)}]^{T} \), because the elements \( Q \) and \( R_{(2,2)} \) are estimable at \( k = 2 \) and the elements \( R_{(1,1)} \) is estimable at time 3 and 4.

Therefore, it can be seen that from the matrix \( \{\Gamma_k\}_{k=2}^4 \) it is possible to estimate the elements \( Q \) and \( R_{(2,2)} \) at time instant \( k = 2 \), the elements \( Q \), \( R_{(1,1)} \), and \( R_{(2,2)} \) at time instant \( k = 3 \), and the elements \( Q \) and \( R_{(1,1)} \) at time instant \( k = 4 \).

### 3.2. MDA for linear system with nonlinear measurement

In this section, the MDA method is proposed for the model with the nonlinear measurement equation which is described by (1) and (9). In this case, the nonlinear function \( h_k(x_k) \) in the measurement equations (2) is linearised by the first order Taylor expansion as

\[
h_k(x_k) \simeq h_k(\bar{x}_k) + \frac{\partial h_k(x_k)}{\partial x_k} |_{x_k = \bar{x}_k} (x_k - \bar{x}_k) = h_k(\bar{x}_k) + H_k(\bar{x}_k)(x_k - \bar{x}_k),
\]

where \( \bar{x}_k \) is an arbitrary reasonable estimate of \( x_k \), which is close to the true state. This estimate may be chosen in accord with the a priori knowledge of the state or the estimate can be obtained as a prediction from the past measurement as

\[
\dot{x}_k = F_k = 1 [S_{k-1} H_{k-1}(\dot{x}_{k-1})]^{T} (z_{k-1} - S_{k-1} h_{k-1}(\dot{x}_{k-1})) + \dot{x}_{k-1} + B_{k-1} u_{k-1},
\]

for \( k = 0, 1, 2, \ldots \).
The MPE is formally the same as in (10), but the measurement prediction (11) is approximate\(^4\) and is of the form

\[
\hat{z}^{(q)}_k = S_k H_k (\hat{x}_k) \prod_{i=1}^{q} F_{k-i} (S_{k-q} H_{k-q} (\hat{x}_{k-q}))^{\dagger} z_{k-q} = \mathcal{H}^{(q)}_k z_{k-q}.
\]  

(41)

The approximate MPE (10) with respect to (1) and (9), and the approximation (39) can be rewritten to the form

\[
\begin{align*}
\mathbf{r}^{(q)}_k &= S_k (h_k (\hat{x}_k) + H_k (\hat{x}_k) (x_k - \hat{x}_k) + v_k) - \\
&- h_k S_{k-q} (h_{k-q} (\hat{x}_{k-q}) + H_{k-q} (\hat{x}_{k-q}) (x_{k-q} - \hat{x}_{k-q}) + v_{k-q}) \\
&= S_k h_k (\hat{x}_k) + S_k H_k (\hat{x}_k) x_k - S_k H_k (\hat{x}_k) \hat{x}_k + S_k v_k - h_k S_{k-q} h_{k-q} (\hat{x}_{k-q}) - \\
&- S_k H_k (\hat{x}_k) \prod_{i=1}^{q} F_{k-i} x_{k-q} + S_k H_k (\hat{x}_k) \prod_{i=1}^{q} F_{k-i} \hat{x}_{k-q} - h_k S_{k-q} v_{k-q}, \\
&= S_k v_k + \sum_{i=0}^{q-1} A^{(i)}_k w_{k-i-1} - h_k S_{k-q} v_{k-q} + \varepsilon^{(q)}_k,
\end{align*}
\]  

(42)

where

\[
\varepsilon^{(q)}_k = S_k h_k (\hat{x}_k) - S_k H_k (\hat{x}_k) x_k - h_k S_{k-q} h_{k-q} (\hat{x}_{k-q}) + A^{(q)}_k \hat{x}_{k-q} + \\
+ \sum_{i=0}^{q-1} A^{(i)}_k G_{k-i-1} u_{k-i-1},
\]  

(43)

and

\[
A^{(i)}_k = S_k H_k (\hat{x}_k) \prod_{j=1}^{i} F_{k-j}.
\]  

(44)

Note that MPE is a stochastic process (14) with the mean

\[
\begin{align*}
\mathbb{E}[\mathbf{r}^{(q)}_k] &= \varepsilon^{(q)}_k.
\end{align*}
\]  

(45)

It means that the estimate of the noise covariance matrices can still be computed according (14)-(32), where \(A^{(i)}_k\) is given by (44), \(h_k^{(q)}\) (41), and \(\mathbb{E}[\mathbf{r}^{(q)}_k]\) by (43). In this case, the unbiasedness of the estimate cannot generally be ensured.

4. Discussion and implementation notes

The estimation of noise covariance matrices by the MDA method can be summarised in three steps

i. Computation of the MPE defined by the \(q\)-step measurement prediction error according to (10).

ii. Calculation of the matrix \(A_k\) (18) from the known matrices \(F_k, H_k, S_k\) and estimation of the vector \(b_k\) according to (23).

iii. Estimation of the noise covariance matrices (26) (or (30),(31)) in the case that the set of sensors is constant in time and (31),(32) in the case that the set of sensors is time-varying.

\(\text{If the state estimate is obtained as the prediction from the past measurement, the estimate depends on the estimated noise covariance matrices via the linearisation point. This causes a correlation between} w, v, \text{and} \hat{x}_k \text{which is neglected in computation of autocovariance function of the MPE (42).}\)
Note 1: The main assumption used by the MDA method is that the matrices $F_k$, $G_k$, $S_k$ are known and the known function $h_k(x_k)$ is differentiable with "invertible" Jacobian. However, this assumption is not sufficient for time-varying systems, because the matrix $A_k(T)$ in (26) is an average of matrices $A_k$, for $k = 2q, \ldots, T$. Even if each matrix $A_k$ is of sufficient rank, the average matrix $A_k(T)$ need not be of sufficient rank. Therefore, the system matrices and function are supposed to be of a reasonable structure to ensure the MDA estimate exists and is unique.

Note 2: The problem with availability of sensors over the time have been solved by formulation the problem as covariance matrix estimation with a time-varying measurement equation in (9).

Note 3: The MDA estimate is unbiased only if the equation (39) is not an approximation of the measurement function. In such case, the estimate is unbiased for an arbitrary number of measurements $T$, for $T \to \infty$ the estimate covariance matrices $\text{cov}([\theta])$ go to zero. The estimate is consistent.

Note 4: In the case that (39) is only an approximation and the model has a reasonable structure and nonlinearity of measurements function the MDA estimate is close to the true noise covariance matrices $Q$ and $R$.

Note 5: The MDA estimate (24) for the LTV model is not the best linear unbiased estimate (BLUE) as the system of linear equations (16) in not optimally weighted in (24). The optimum weighting matrix depends on the estimated covariance matrices $Q$ and $R$ and thus, it is unknown [18, 26].

Note 6: The estimate (24) does not respect the symmetry of the estimated noise covariance matrices. The symmetry can be ensured if the matrix $A_k$ is multiplied by the matrix $P$ defined in such a way that each column of the matrix $A_k$ depends only on one element of the matrices $Q$ or $R$. Therefore, the columns of $A_k$ which are related to the same element are summed by the matrix $P$. Besides the symmetrisation, the matrix $P$ reduces the complexity of the calculation (matrix $A_kP$ has only $(n_x+1)n_z + (n_z+1)n_z$ columns instead of $n_x^2 + n_z^2$ in the matrix $A_k$) [18].

Note 7: The recursive forms (30)-(31) or (31)-(32) are advantageous for the on-line use of the algorithm. They may also be useful for a formulation of the MDA method with the exponential forgetting factor if the noise covariance matrices $Q$ and $R$ are slowly varying in time.

Note 8: If there exist some knowledge about the estimated covariance matrices, then it can be used as the initial condition in the recursive form (30)-(31) or (31)-(32).

5. Numerical Illustrations

In this section two examples illustrating performance of the proposed MDA estimation are given.

5.1. Example 1 - MDA for linear system

The first example of the MDA method is for LTV model described by (1) and (9) with $h_k(x_k) = Hx_k$ and $G_k = 0_{m_x \times n_w}$. The model is considered with the following matrices

$$F_k = \begin{bmatrix} \gamma_k & 0 \\ 0.3 & 0.8 \end{bmatrix}, \quad Q = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 2 \end{bmatrix}, \quad S_k = \begin{cases} \text{blkdiag}([1 1 0]) & \text{for } k = 0, \ldots, T, \\ \text{blkdiag}([1 1 0]) & \text{for } k = T + 1, \ldots, 2T, \\ \text{blkdiag}([1 1 1]) & \text{for } k = 2T + 1, \ldots, 3T, \end{cases}$$

$$H = \begin{bmatrix} -1 & 2 \\ 0.5 & 1 \\ 2 & 0 \end{bmatrix}, \quad R = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 4 & 0 \\ 1 & 0 & 2 \end{bmatrix}, \quad \gamma_k = 0.7 + 0.32 \sin(0.0005k), \text{ for } k = 0, \ldots, T.$$

The method is analysed using $10^4$ Monte-Carlo (MC) simulations with $T = 10^4$ time instants per a MC simulation and with $q = 2$. The estimates for each MC simulation are given in Figure 1 together with the confidence interval of three standard deviations (3STD). The evolution of the MDA estimates for one selected MC simulation is plotted in Figure 2 and the availability of the measurement elements together with three zones that denote the availability of the respective measurements and thus estimability of the noise covariance elements are illustrated in Figure 3.
5.2. Example 2 - MDA for system with nonlinear measurement function

The second example is for a model with the nonlinear measurement equation. The system is considered with the following system matrices and the measurement functions

\[
F_k = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \forall k,
\]

\[
u_k = \begin{bmatrix} \cos\left(\frac{2\pi k}{T}\right) \\ \sin\left(\frac{2\pi k}{T}\right) \end{bmatrix}, \quad k = 0, \ldots, T,
\]

\[
G_k = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix}, \forall k,
\]

\[
h_k(x_k) = \begin{bmatrix} \sqrt{(x_{k,1} - 100)^2 + (x_{k,2} - 300)^2} \\ \sqrt{(x_{k,1} + 400)^2 + (x_{k,2} - 600)^2} \\ \sqrt{(x_{k,1} + 200)^2 + (x_{k,2})^2} \end{bmatrix}, \forall k,
\]

\[
S_k = \begin{cases} \text{blkdiag}([0 1 1]), & \text{for } k = 0, \ldots, \frac{T}{2}, \\
\text{blkdiag}([1 1 1]), & \text{for } k = \frac{T}{2} + 1, \ldots, T, \end{cases}
\]

with \(Q\) and \(R\) selected as in Example 1. The model is motivated by navigation with distance measurements by three different radio transmitters such as the distance measuring equipment (DME).

The method is simulated with \(T = 10^4\) time instants with \(q = 3\). The estimates are plotted in Figure 4 and the state trajectory together with the positions of transmitters are given in Figure 5.
It can be seen that the estimate approaches to the true values in Figure 4. It can also be seen that the estimate $\hat{R}(1, 1)$ is not available only from $k = 0$ to $k = \frac{T}{2}$, because the sensor 1 (which represents the measurement of the first transmitter) is not available in this time interval.

6. Concluding remarks
The paper focused on the estimation of the noise covariance matrices of systems described by the nonlinear time-varying state-space model with the varying set of sensors. For such a model the measurement difference autocovariance least-squares-based method, was proposed. The method belongs to the correlation methods, which are based on the statistical analysis of the measurement prediction error. The properties of the method were thoroughly analysed, and it was shown that the MDA method provides unbiased and consistent estimates of noise covariance matrices for a linear model, even for a time-varying set of sensors. The method was illustrated using two numerical studies covering the linear time-varying model and the nonlinear model with a time-varying set of sensors. The results have shown satisfactory performance of the MDA method for both models.

Acknowledgement: The work was supported by the Czech Science Foundation, project GA15-12068S, and the Czech Ministry of Education, Youth and Sports, project LO1506.

Appendix: Recursive Relations for MDA
In this appendix, a recursive algorithm for the MDA method is derived. Starting from the single-shot version of the MDA (24), the matrix (28) and (29) can be rewritten to the following form

$$\hat{A}(k) = \frac{1}{k-2q+1} \sum_{i=2q}^{k} A_i = \frac{1}{k-2q} \sum_{i=2q}^{k-1} A_i + A_k = \hat{A}_{(k-1)} + A_k,$$

$$\hat{b}(k) = \frac{1}{k-2q+1} \sum_{k=2q}^{k} \hat{b}_k = \frac{1}{k-2q} \sum_{i=2q}^{k-1} \hat{b}_i + \hat{b}_k = \hat{b}_{(k-1)} + \hat{b}_k.$$

It means that (26) by the use of (49) is equal to

$$\hat{\theta}(k) = (\hat{A}(k))^{\dagger}(\hat{b}_{(k-1)} + \hat{b}_k),$$
where with (48) and the fact that $b_{(k-1)} = \bar{A}_{(k-1)}\hat{\theta}^*_{(k-1)}$ the equation (50) can be modified to

$$
\hat{\theta}^*_{(k)} = (\bar{A}_{(k)})^\dagger \left((\bar{A}_{(k)} - A_{(k)})\hat{\theta}^*_{(k-1)} + \hat{b}_k\right) = \\
\hat{\theta}^*_{(k-1)} + (\bar{A}_{(k)})^\dagger \left(\hat{b}_k - A_{(k)}\hat{\theta}^*_{(k-1)}\right).
$$

(51)

References

[1] R. K. Mehra, “On the identification of variances and adaptive filtering,” IEEE Transactions on Automatic Control, vol. 15, no. 2, pp. 175–184, 1970.

[2] P. R. Bélanger, “Estimation of noise covariance matrices for a linear time-varying stochastic process,” Automatica, vol. 10, no. 3, pp. 267–275, 1974.

[3] B. J. Odellson, M. R. Rajamani, and J. B. Rawlings, “A new autocovariance least-squares method for estimating noise covariances,” Automatica, vol. 42, no. 2, pp. 303–308, 2006.

[4] B. M. Åkesson, J. B. Jørgensen, N. K. Poulsen, and S. B. Jørgensen, “A generalized autocovariance least-squares method for Kalman filter tuning,” Journal of Process Control, vol. 18, no. 7–8, pp. 769–779, 2008.

[5] M. Šimandl and J. Duník, “Estimation of noise covariance matrices for periodic systems,” International Journal of Adaptive Control and Signal Processing, vol. 25, pp. 928–942, 2011.

[6] J. Zhou and R. H. Luecke, “Estimation of the covariances of the process noise and measurement noise for a linear discrete dynamic system,” Computers & Chemical Engineering, vol. 19, no. 2, pp. 187–195, 1995.

[7] D. G. Lainiotis, “Optimal adaptive estimation: Structure and parameters adaptation,” IEEE Transactions On Automatic Control, vol. 16, no. 2, pp. 160–170, 1971.

[8] S. Särkkä and A. Nummenmaa, “Recursive noise adaptive Kalman filtering by variational Bayesian approximations,” IEEE Transactions on Automatic Control, vol. 54, no. 3, pp. 596–600, 2009.

[9] R. L. Kashyap, “Maximum likelihood identification of stochastic linear systems,” IEEE Transactions on Automatic Control, vol. 15, no. 1, pp. 25–34, 1970.

[10] R. H. Shumway and D. S. Stoffer, “Time series analysis and its applications,” 2000.

[11] K. A. Myers and B. D. Tapley, “Adaptive sequential estimation with unknown noise statistics,” IEEE Transactions on Automatic Control, vol. 21, no. 8, pp. 520–523, 1976.

[12] D. M. Wiberg, T. D. Powell, and D. Ljungquist, “An on-line parameter estimator for quick convergence and time-varying linear systems,” IEEE Transactions on Automatic Control, vol. 45, no. 10, pp. 1854–1863, 2000.

[13] B. Carew and P. R. Bélanger, “Identification of optimum steady-state gain for systems with unknown noise covariances,” IEEE Transactions on Automatic Control, vol. 18, no. 6, pp. 582–587, 1973.

[14] 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.

[15] J. Duník, M. Šimandl, and O. Straka, “Methods for estimating state and measurement noise covariance matrices: Aspects and comparison,” in Proceedings of 15th IFAC Symposium on System Identification, (Saint-Malo, France), July 2009.

[16] M. R. Rajamani and J. B. Rawlings, “Estimation of the disturbance structure from data using semidefinite programming and optimal weighting,” Automatica, vol. 45, no. 1, pp. 142–148, 2009.

[17] F. V. Lima, M. R. Rajamani, T. A. Soderstrom, and J. B. Rawlings, “Covariance and state estimation of weakly observable systems: Application to polymerization processes,” IEEE Transactions on Control Systems Technology, vol. 21, no. 4, pp. 1249–1257, 2013.

[18] J. Duník, O. Straka, and M. Šimandl, “Estimation of noise covariance matrices for linear systems with nonlinear measurements,” in Proceedings of the 17th IFAC Symposium on System Identification, (Beijing, China), 2016.

[19] C. Wunsch, Discrete Inverse and State Estimation Problems: With Geophysical Fluid Applications. Cambridge, 2006.

[20] J. Duník, O. Straka, and O. Kost, “Measurement difference autocovariance method for noise covariance matrices estimation,” in Accepted for the 55th IEEE Conference on Decision and Control, Dec. 2016.

[21] R. M. Rogers, Applied Mathematics in Integrated Navigation Systems (2nd Edition). AIAA, 2003.

[22] P. Groves, Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems, Second Edition: GNSS/GPS, Artech House, 2013.

[23] J. W. Brewer, “Kronecker products and matrix calculus in system theory,” IEEE Transactions on Circuits and Systems, vol. 25, no. 9, pp. 772–781, 1978.

[24] J. Duník, O. Straka, and Šimandl, “On autocovariance least-squares method for noise covariance matrices estimation,” accepted for IEEE Transactions on Automatic Control, DOI 10.1109/TAC.2016.2571899, 2017.