Methods for Estimating State and Measurement Noise Covariance Matrices: Aspects and Comparison

Jindřich Duník, Miroslav Šimandl and Ondřej Straka

Research Centre Data-Algorithms-Decision Making
Department of Cybernetics
Faculty of Applied Sciences
University of West Bohemia
Czech Republic

July 6, 2009

Introduction

- State estimation, signal processing, and optimal control design methods require complete knowledge of the functions in the system equations and of the statistics of the noises affecting the system and the measurement.
- However, from practical viewpoint of usage of the methods the knowledge of the noise statistics is questionable in many cases.
- Incorrect description of the noise statistics can cause the significant worsening of estimation or control quality.
- Therefore, a great deal of attention has been paid to design of noise covariance matrices estimation methods.

Description of system

Stochastic dynamic time-invariant linear system

$$\mathbf{x}_{k+1} = \mathbf{F}\mathbf{x}_k + \mathbf{w}_k, \quad k = 0, 1, 2, \ldots$$
$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k, \quad k = 0, 1, 2, \ldots$$

is considered, where

- $\mathbf{F}$, $\mathbf{H}$ are known time-invariant matrices,
- $\mathbf{w}_k$ and $\mathbf{v}_k$ are the state and measurement noises with zero mean and unknown covariance matrices $\mathbf{Q}$ and $\mathbf{R}$, respectively,
- the noises are mutually independent and independent of the initial state $\mathbf{x}_0$.

The aim of noise covariance estimation methods

The aim is to find an estimate of $\mathbf{Q}$ and $\mathbf{R}$ on the basis of measured data $\mathbf{z}_k$, $\forall k$.

Standard noise covariance matrix estimation methods

Standard methods can be divided into several main categories:

- correlation methods (Mehra, 1970, Bélanger, 1974),
- prediction error methods (Ljung, 1999),
- maximum likelihood estimation (Kashyap, 1971),
- minimax methods (Verdú and Poor, 1984),
- subspace methods (van Overshee and de Moor, 1996),
- local nonlinear filters (Simon, 2006) - not fully correct.

Novel methods preferred in the paper:

- third-order moment parameter estimator
  (Wiberg et al., 2000, 2008),
- measurement prediction error based methods
  (Odelson, et al. 2006, Duník and Šimandl, 2008).

Goals of the paper

The goals of the paper are

- to provide a regular substantiation why the standard local filters (e.g. the extended Kalman filter, the derivative-free Kalman filters) cannot be used for joint state and noise covariance matrices elements estimation and
- to summarise, analyse, and compare the novel methods for noise covariance matrices estimation, namely
  - third-order moment parameter estimator,
  - measurement prediction error based methods.

Estimation of noise cov. matrix by standard local filters

- First group of the novel noise covariance estimation methods is based on the idea of simultaneous (joint) estimation of the state and parameters, i.e.

$$\mathcal{X} = [\mathbf{x}^T, \boldsymbol{\theta}^T]^T,$$

  where $\boldsymbol{\theta}$ is the vector of unknown parameters in the state space model.
- Such concatenation leads to a generally nonlinear state space model and to nonlinear state estimation techniques.
- However, the standard local filters cannot be used for estimation of elements of $\mathbf{Q} = \text{cov}[\mathbf{w}_k]$, which is sometimes overlooked.

Extended Kalman filter (EKF) and identifiability of $\mathbf{Q}$

- The problem of estimating concatenated state $\mathcal{X}$, where $\boldsymbol{\theta} = [Q_{1,1}, \ldots, Q_{n_x,n_x}]^T$, by the plain EKF is analysed using the Bayesian rec. relations and moment generating functions.
- The filtering pdf is then given by

$$p(\mathcal{X}_0|\mathbf{z}^0) = \mathcal{N}\left\{\mathcal{X}_0 : \begin{bmatrix} \hat{\mathbf{x}}_{0|0} \\ \hat{\boldsymbol{\theta}}_{0|0} \end{bmatrix}, \begin{bmatrix} \mathbf{P}_{x,0|0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{\theta,0|0} \end{bmatrix}\right\},$$

  where $\hat{\mathbf{x}}_{0|0} = \hat{\mathbf{x}}_{0|-1} + \mathbf{K}_0(\mathbf{z}_0 - \mathbf{H}\hat{\mathbf{x}}_{0|-1})$ and $\hat{\boldsymbol{\theta}}_{0|0} = \hat{\boldsymbol{\theta}}_{0|-1}$.
- If $\mathbf{P}_{x,\theta,0|-1} = \mathrm{cov}[\mathbf{x}_0, \boldsymbol{\theta}_0|\mathbf{z}^{-1}] = \mathbf{0}$, the filtering estimates of $\mathbf{x}_0$ and $\boldsymbol{\theta}_0$ remain independent.

Extended Kalman filter and identifiability of $\mathbf{Q}$ (cont'd)

- The predictive pdf is of the form

$$p(\mathcal{X}_1|\mathbf{z}^0) = \mathcal{N}\{\mathbf{x}_1 : \mathbf{F}\hat{\mathbf{x}}_{0|0}, \mathbf{F}\mathbf{P}_{x,0|0}\mathbf{F}^{\mathrm{T}} + \mathbf{Q}(\hat{\boldsymbol{\theta}}_{0|0})\}\times$$
$$\times\,\mathcal{N}\{\boldsymbol{\theta}_1 : \hat{\boldsymbol{\theta}}_{0|0}, \mathbf{P}_{\theta,0|0}\}$$

  and it leads to the following matrices
  - $\mathbf{P}_{x,\theta,1|0} = \mathrm{cov}[\mathbf{x}_1, \boldsymbol{\theta}_1|\mathbf{z}^0] = \mathbf{0}$,
  - $E[(\boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_{1|0}) \otimes (\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})(\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})^{\mathrm{T}}|\mathbf{z}^0] = \begin{bmatrix} \bigoplus_{i=1}^{n_x} P_{\theta,0|0}(1,i) \\ \bigoplus_{i=1}^{n_x} P_{\theta,0|0}(2,i) \\ \vdots \\ \bigoplus_{i=1}^{n_x} P_{\theta,0|0}(n_x,i) \end{bmatrix}$,

    where $\otimes$ is the Kronecker product and $\bigoplus$ the direct sum.

- Predictive estimates of $\mathbf{x}_1$ and $\boldsymbol{\theta}_1$ are dependent. But, considering first two moments only, they seem to be independent.

Local filter as parameter estimator

Local nonlinear filters, computing first two moments only, cannot be used for estimation of elements of the noise covariance.

Third-order moment parameter estimator (3-OM)

- The third-order moment parameter estimator proposed in (Wiberg, 2000, 2008) extends the standard EKF algorithm with the recursive relations for computing third-order cross-conditional moment.
- The extension causes that the cross-correlation matrix $\mathbf{P}_{x,\theta}$ does not converge to zero and the filtering estimate of the noise covariance matrix parameters are updated according to the measurement as

$$\hat{\theta}_{k|k} = \hat{\theta}_{k|k-1} + \mathbf{P}^T_{x,\theta,k|k-1}\mathbf{H}^T\mathbf{P}^{-1}_{z,k|k-1}(\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1}).$$

Outline of second and third-order moment computation

- time update

$$\text{2nd moment}: \mathbf{P}_{x,\theta,k+1|k} = \mathbf{F}\mathbf{P}_{x,\theta,k|k}$$

$$\text{3rd moment}: \mathbf{m}_{k+1|k} = \mathbf{F}\mathbf{m}_{k|k}\mathbf{F}^T + P_{\theta,k|k}$$

- measurement update

$$\mathbf{P}_{x,\theta,k|k} = (\mathbf{I} - \mathbf{K}_{x,k}\mathbf{H})(\mathbf{P}_{x,\theta,k|k-1} + \mathbf{m}_{k|k-1}\mathbf{H}^T\mathbf{P}_{z,k|k-1}^{-1}\mathbf{e}_k)$$

$$\mathbf{m}_{k|k} = (\mathbf{I} - \mathbf{K}_{x,k}\mathbf{H})\mathbf{m}_{k|k-1}(\mathbf{I} - \mathbf{K}_{x,k}\mathbf{H})^T$$

$$\mathbf{e}_k = \mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1}$$

Pros and Cons

+ The 3-OM can also be used for time-variant systems.
+ The convergence of the 3-OM to a local maximum of the likelihood function can be proved.
− The 3-OM is sensitive to initial conditions.

Novel correlation methods (CM)

- Correlation methods are generally based on the reformulating the generally nonlinear estimation problem as a pseudo-linear problem, i.e.
  - second-order statistics of the measurement prediction error sequence (innovation sequence) produced by a linear estimator are computed and then
  - the noise covariance matrices are estimated using ordinary least-squares method.
- Characteristic property - the second-order statistics of the innovation sequence are expressed as a linear function of $\mathbf{Q}$ and $\mathbf{R}$.

Novel correlation methods (cont'd)

Novel correlation methods were proposed in

- (Odelson et al., 2006), where the second-order statistics of innovation sequence produced by suboptimal linear filter are analysed - filter gain has to be specified,
- (Duník and Šimandl, 2008), where the second-order statistics of innovation sequence produced by linear predictor are analysed - no user defined parameters are required,
- (Åkesson, 2008), where the Odelson's method was extended for correlated state and measurement noises.

Outline of algorithm of correlation methods

- Compute innovation sequence $\mathbf{e}_k = \mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_k$ using suboptimal linear filter or linear multi-step predictor for $k = 0, 1, \ldots, N$.
- Estimate the autocovariance according to

$$\hat{\mathbf{C}}_{e,j} = \frac{1}{N-j+1} \sum_{i=0}^{N-j} \mathbf{e}_i \mathbf{e}_{i+j}^T.$$

- Determine estimates of the noise covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ as a solution of a system of linear equations of the form

$$\text{vec}(\hat{\mathbf{C}}_{e,0}) = (\mathbf{H} \otimes \mathbf{H})(\mathbf{I} - \mathbf{F} \otimes \mathbf{F})^{-1}\text{vec}(\mathbf{Q}) + \text{vec}(\mathbf{R}),$$

where $\text{vec}(\cdot)$ is vector operator converting an arbitrary matrix into a column vector.

Pros and Cons

+ Estimates of the noise covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ can be proved to converge to the true values with increasing sample size N.
+ The methods are not too sensitive to user defined parameters.
− Generally, these techniques can be used on assumption of time-invariant and completely known matrices $\mathbf{F}$ and $\mathbf{H}$.
− The methods were proposed for off-line data processing.

System definition

The following system is considered

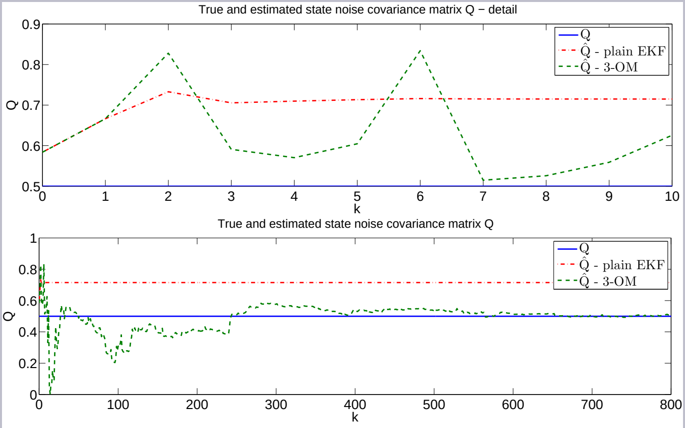
$$x_{k+1} = 0.9x_k + w_k,$$
$$z_k = x_k + v_k,$$

where

- $k = 0, 1, \ldots, 400,$
- the system initial condition is given by the mean $\hat{x}_0 = 20$ and the variance $P_0 = 0.1$, and
- the true variances of the state and measurement noise are $Q = 0.5$ and $R = 1$.

Example of estimation of Q using plain EKF and 3-OM

The EKF and 3-OM were run with initial values $\hat{x}_{0|-1} = 20$, $\hat{\theta}_{0|-1} = 1$, $P_{x,0|-1} = P_{\theta,0|-1} = 1$, and $P_{x,\theta,0|-1} = 0.5$.



Sample means and variances of estimated noise variances

Experiment was repeated $M = 10^4$ times in order to compute sample mean and variance of the noise variance Q estimate

$$\hat{E}[\hat{Q}] = \frac{1}{M}\sum_{j=1}^{M}\hat{Q}_j, \qquad \hat{\text{var}}[\hat{Q}] = \frac{1}{M-1}\sum_{j=1}^{M}\left(\hat{Q}_j - \hat{E}[\hat{Q}]\right)^2.$$

Analogously for R.

Results

	$\hat{E}[\hat{Q}]$	$\hat{E}[\hat{R}]$	$\hat{\text{cov}}[\hat{Q}]$	$\hat{\text{cov}}[\hat{R}]$
3-OM($\hat{Q}_{0\|-1} = 1,\ \hat{R}_{0\|-1} = 0.5$)	0.57	0.96	0.01	0.02
3-OM($\hat{Q}_{0\|-1} = 2,\ \hat{R}_{0\|-1} = 2$)	1.62	0.34	0.04	0.03
CM($K = 0.2$)	0.49	0.99	0.01	0.01
CM($K = 0$)	0.51	0.99	0.02	0.02
true	0.5	1	–	–

Conclusion

- The paper dealt with novel methods for estimation of the state and measurement noise covariance matrices of linear systems.
- It was shown that any of the standard local filter (recursively computing the first two moments only) does not allow to identify the noise covariance matrix by means of joint estimation.
- The novel methods were analysed and the correlation methods were presented in a unified framework.
- Pros and cons of particular novel methods were presented.