

# Cramér-Rao Bounds for Estimation of Linear System Noise Covariances

Peter Matisko\*, Vladimír Havlena

Czech Technical University in Prague, Faculty of Electrical Engineering Department of Control Engineering

**Abstract** The performance of Kalman filter depends directly on the noise covariances, which are usually not known and need to be estimated. Several estimation algorithms have been published in past decades, but the measure of estimation quality is missing. The Cramér-Rao bounds represent limitation of quality of parameter estimation that can be obtained from given data. In this article, The Cramér-Rao bounds for noise covariance estimation of linear time-invariant stochastic system will be derived. Two different linear system models will be considered. Further, the performance of earlier published methods will be discussed according to the Cramér-Rao bounds. The analogy between the Cramér-Rao bounds and the Riccati equation will be pointed out.

**Keywords** Cramér-Rao bounds, Kalman Filter, Noise covariance estimation

## 1. Introduction<sup>1</sup>

The Cramér-Rao (CR) bounds represent a limitation of the estimation quality of unknown parameters from the given data. If the variance of the estimates reaches CR bounds, it can be stated that the estimation algorithm works optimally. In this paper, we will derive CR bounds for the estimation of noise covariances of a linear stochastic system. The noise covariances are tuning parameters of Kalman filter, and the filter performance depends on them. However, they are not known in general and hence tuning of a Kalman filter remains a challenging task. The very first methods for noise covariance estimation were published in 70s' by Mehra,[1,2]. Recently, new methods were proposed by Odelson et al. [3] and further modified by Rajamani et al.[4,5] and Akesson et al.[6]. A maximum likelihood approach was described in[7]. In Section 5, these two approaches will be compared to the CR bounds and their performance will be discussed. In the last part of the paper, relationship between the Riccati equation and the CR bounds will be pointed out.

Deriving of the noise covariance CR bounds is based on[8] and[9]. Additional information about CR bounds and Bayesian estimation can be found in[10].

## 2. Cramér-Rao Bounds

\* Corresponding author:

peter.matisko@fel.cvut.cz (Peter Matisko)

Published online at <http://journal.sapub.org/jmea>

Copyright © 2012 Scientific & Academic Publishing. All Rights Reserved

The task is to estimate a vector of parameters  $\theta = [\theta_1, \theta_2, \dots, \theta_N]^T$  from a set of measured data  $y^N = [y_1^T, y_2^T, \dots, y_N^T]^T$ .

The upper index  $N$  emphasizes that vector  $y^N$  contains  $N$  data samples. The conditioned probability density function (pdf)  $p(y^N | \theta)$  is assumed to be known. Further, the nabla operator is defined as

$$\nabla_{\theta} = \left[ \frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \dots, \frac{\partial}{\partial \theta_N} \right]. \quad (1)$$

The Fisher information matrix (FIM)  $J(\theta)$  for parameters  $\theta$  is defined as

$$J(\theta) = -\mathcal{E}_y \left\{ \nabla_{\theta} \left[ \nabla_{\theta} \ln p(y^N | \theta) \right]^T \right\}, \quad (2)$$

where operator  $\mathcal{E}_y \{ \cdot | \cdot \}$  is the conditioned expected value defined as

$$\mathcal{E}_y \{ f(\theta) | \theta \} = \int_{-\infty}^{\infty} f(\theta) p(y | \theta) dy. \quad (3)$$

Alternatively, we can define a posterior FIM (and posterior CR bounds) using joint probability  $p(y^N, \theta)$  instead of the conditioned one, where  $p(y^N, \theta) = p(y^N | \theta) p(\theta)$ . Probability function  $p(\theta)$  represents prior information about the parameter vector  $\theta$ .

Assume having an unbiased estimate of parameter vector  $\hat{\theta}$  obtained by an arbitrary method. When the true value of parameter vector is compared to the unbiased estimated set, the following inequality holds, [8].

$$\mathcal{E}_{\theta} \left\{ (\theta - \hat{\theta})(\theta - \hat{\theta})^T \right\} \geq J^{-1}(\theta), \quad (4)$$

where  $J^{-1}(\theta)$  represents the Cramér-Rao bound. Inequality  $A \geq B$  means that  $A - B$  is a positive semidefinite matrix. Based on (4), the following statement can be

<sup>1</sup> This paper was presented at the 18<sup>th</sup> IFAC World Congress, Italy.  
[online] <http://www.ifac-papersonline.net/Detailed/50045.html>

concluded. Assuming that CR bound is reachable, any estimation algorithm working optimally (in the sense of the smallest covariance of obtained estimates), must give estimates whose variance is equal to the Cramér-Rao bound. If the CR bound is reachable, then the optimal result can be achieved by a maximum likelihood approach.

Consider a scalar linear stochastic system

$$\begin{aligned} x_{k+1} &= ax_k + v_k, \\ y_k &= cx_k + e_k, \end{aligned} \quad (5)$$

where  $x_k$  and  $y_k$  represent the state and the measured output, respectively. Stochastic inputs  $v_k$  and  $e_k$  are described by normal distributions

$$v(t) \sim \mathcal{N}(0, Q), \quad e(t) \sim \mathcal{N}(0, R). \quad (6)$$

Using the notation of [9], the logarithms of the conditioned probability of the state, measurements and unknown parameters can be expressed as

$$\begin{aligned} \ln p(\mathbf{x}^k, \mathbf{y}^k | \boldsymbol{\theta}) &= \sum_{i=0}^k \ln p(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}) \\ &+ \sum_{i=0}^k \ln p(\mathbf{x}_{i+1} | \mathbf{x}_i, \boldsymbol{\theta}) + \ln p(\mathbf{x}_0). \end{aligned} \quad (7)$$

The following matrices are used for FIM calculation

$$\mathbf{K}_{k+1}^k = \mathcal{E}_{x_{k+1}} \left\{ -\nabla_{\mathbf{x}_k} \left[ \nabla_{\mathbf{x}_k} \ln p(\mathbf{x}_{k+1} | \mathbf{x}_k) \right]^T | \mathbf{x}_k \right\}, \quad (8)$$

$$\mathbf{K}_{k+1}^{k,k+1} = \mathcal{E}_{x_{k+1}} \left\{ -\nabla_{\mathbf{x}_{k+1}} \left[ \nabla_{\mathbf{x}_k} \ln p(\mathbf{x}_{k+1} | \mathbf{x}_k) \right]^T | \mathbf{x}_k \right\}, \quad (9)$$

$$\mathbf{K}_{k+1}^{k+1} = \mathcal{E}_{x_{k+1}} \left\{ -\nabla_{\mathbf{x}_{k+1}} \left[ \nabla_{\mathbf{x}_{k+1}} \ln p(\mathbf{x}_{k+1} | \mathbf{x}_k) \right]^T | \mathbf{x}_k \right\}, \quad (10)$$

$$\mathbf{K}_{k+1}^{\theta,k} = \mathcal{E}_{x_{k+1}} \left\{ -\nabla_{\mathbf{x}_k} \left[ \nabla_{\boldsymbol{\theta}} \ln p(\mathbf{x}_{k+1} | \mathbf{x}_k, \boldsymbol{\theta}) \right]^T | \mathbf{x}_k, \boldsymbol{\theta} \right\}, \quad (11)$$

$$\mathbf{K}_{k+1}^{\theta,k+1} = \mathcal{E}_{x_{k+1}} \left\{ -\nabla_{\mathbf{x}_{k+1}} \left[ \nabla_{\boldsymbol{\theta}} \ln p(\mathbf{x}_{k+1} | \mathbf{x}_k, \boldsymbol{\theta}) \right]^T | \mathbf{x}_k, \boldsymbol{\theta} \right\}, \quad (12)$$

$$\mathbf{K}_{k+1}^{\theta} = \mathcal{E}_{x_{k+1}} \left\{ -\nabla_{\boldsymbol{\theta}} \left[ \nabla_{\boldsymbol{\theta}} \ln p(\mathbf{x}_{k+1} | \mathbf{x}_k, \boldsymbol{\theta}) \right]^T | \mathbf{x}_k, \boldsymbol{\theta} \right\}, \quad (13)$$

$$\mathbf{L}_k^k = \mathcal{E}_{y_k} \left\{ -\nabla_{\mathbf{x}_k} \left[ \nabla_{\mathbf{x}_k} \ln p(\mathbf{y}_k | \mathbf{x}_k) \right]^T | \mathbf{x}_k \right\}, \quad (14)$$

$$\mathbf{L}_k^{\theta,k} = \mathcal{E}_{y_k} \left\{ -\nabla_{\mathbf{x}_k} \left[ \nabla_{\boldsymbol{\theta}} \ln p(\mathbf{y}_k | \mathbf{x}_k, \boldsymbol{\theta}) \right]^T | \mathbf{x}_k, \boldsymbol{\theta} \right\}, \quad (15)$$

$$\mathbf{L}_k^{\theta} = \mathcal{E}_{y_k} \left\{ -\nabla_{\boldsymbol{\theta}} \left[ \nabla_{\boldsymbol{\theta}} \ln p(\mathbf{y}_k | \mathbf{x}_k, \boldsymbol{\theta}) \right]^T | \mathbf{x}_k, \boldsymbol{\theta} \right\}, \quad (16)$$

Further, recursive formulas for FIM of the state and unknown parameters are defined as

$$\begin{aligned} \mathbf{J}_{k+1|k}^{\theta,\theta} &= \mathbf{J}_{k|k-1}^{\theta,\theta} + \mathbf{I}_k^{\theta} + \mathbf{K}_{k+1}^{\theta} - \Delta_{kk}^{\theta,x} \Delta_{kk}^{-1} \Delta_{kk}^{x,\theta}, \\ \mathbf{J}_{k+1|k}^{x,\theta} &= \mathbf{K}_{k+1}^{k+1,\theta} - \mathbf{K}_{k+1}^{k+1,k} \Delta_{kk}^{-1} \Delta_{kk}^{x,\theta}, \\ \mathbf{J}_{k+1|k}^{x,x} &= \mathbf{K}_{k+1}^{k+1} - \mathbf{K}_{k+1}^{k+1,k} \Delta_{kk}^{-1} \mathbf{K}_{k+1}^{k,k+1}, \end{aligned} \quad (17)$$

where

$$\begin{aligned} \Delta_{kk} &= \mathbf{J}_{k|k}^{x,x} + \mathbf{K}_{k+1}^k, \\ \Delta_{kk}^{\theta,x} &= \mathbf{J}_{k|k}^{\theta,x} + \mathbf{K}_{k+1}^{\theta,k}, \end{aligned} \quad (18)$$

$$\begin{aligned} \Delta_{kk}^{x,\theta} &= \left[ \Delta_{kk}^{\theta,x} \right]^T, \\ \mathbf{K}_{k+1}^{k+1,k} &= \left[ \mathbf{K}_{k+1}^{k,k+1} \right]^T, \end{aligned} \quad (19)$$

and

$$\begin{aligned} \mathbf{J}_{k|k}^{x,\theta} &= \mathbf{J}_{k|k-1}^{x,\theta} + \mathbf{I}_k^{k,\theta}, \\ \mathbf{J}_{k|k}^{x,x} &= \mathbf{J}_{k|k-1}^{x,x} + \mathbf{I}_k^k. \end{aligned} \quad (20)$$

Initial conditions of the recursive algorithm are set as

$$\begin{aligned} \mathbf{J}_{0|-1}^{x,\theta} &= \mathbf{K}_0^{0,\theta} = 0, \\ \mathbf{J}_{0|-1}^{x,x} &= \mathbf{K}_0^0 = \mathcal{E}_{\mathbf{x}_0} \left\{ -\nabla_{\mathbf{x}_0} \left[ \nabla_{\mathbf{x}_0} \ln p(\mathbf{x}_0) \right]^T \right\}, \\ \mathbf{J}_{0|-1}^{\theta,\theta} &= \mathbf{A}_0 = \mathcal{E}_{\boldsymbol{\theta}} \left\{ -\nabla_{\boldsymbol{\theta}} \left[ \nabla_{\boldsymbol{\theta}} \ln p(\boldsymbol{\theta}) \right]^T \right\}, \end{aligned} \quad (21)$$

where  $\mathbf{K}_0^0$  and  $\mathbf{A}_0$  represent prior information. If no prior probability function is known, then  $\mathbf{A}_0 = 0$ ,  $\mathbf{K}_0^0 = 0$ . The final form of FIM of the state and parameters is

$$\mathbf{J}_{k|k} = \begin{bmatrix} \mathbf{J}_{k|k}^{x,x} & \mathbf{J}_{k|k}^{x,\theta} \\ \mathbf{J}_{k|k}^{\theta,x} & \mathbf{J}_{k|k}^{\theta,\theta} \end{bmatrix}. \quad (22)$$

### 3. Cramér-Rao Bounds for Noise Covariances Estimation

Consider a scalar stochastic system given by (5), where  $|a| \leq 1$ . The noise sources are not correlated and are defined by (6). The probabilities used in (7) can be defined as

$$\begin{aligned} p(x_{k+1} | x_k, Q) &= \mathcal{N}(ax_k, Q), \\ p(y_k | x_k, R) &= \mathcal{N}(cx_k, R). \end{aligned} \quad (23)$$

Now, matrices (11)-(16) can be calculated using the following general integral formulas, assuming  $\alpha > 0$

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-\alpha x^2} dx &= \sqrt{\frac{\pi}{\alpha}}, \\ \int_{-\infty}^{\infty} x e^{-\alpha(x-\mu)^2} dx &= \mu \sqrt{\frac{\pi}{\alpha}}, \\ \int_{-\infty}^{\infty} (x-\mu) e^{-\alpha(x-\mu)^2} dx &= 0, \\ \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\alpha(x-\mu)^2} dx &= \sqrt{\frac{\pi}{4\alpha^3}}. \end{aligned} \quad (24)$$

Further, logarithm of the Gaussian function is given by

$$\ln \mathcal{N}(\mu, P) = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln P - \frac{1}{2P} (x - \mu)^2. \quad (25)$$

CR bounds of  $Q, R$  estimates are to be found, considering a scalar system with two unknown parameters. The nabla operator, in this case, is of the form

$$\nabla_{\boldsymbol{\theta}} = \left[ \frac{\partial}{\partial Q}, \frac{\partial}{\partial R} \right], \quad (26)$$

where  $\boldsymbol{\theta} = [Q, R]^T$ . At the first step, the arguments of the expectation operators in (11)-(16) are calculated as follows

$$\nabla_{\mathbf{x}_k} [\nabla_{\boldsymbol{\theta}} \ln p(\mathbf{x}_{k+1} | \mathbf{x}_k, \boldsymbol{\theta})]^T = \begin{bmatrix} -\frac{a}{Q^2}(x_{k+1} - ax_k) \\ 0 \end{bmatrix}, \quad (27)$$

$$\nabla_{\mathbf{x}_{k+1}} [\nabla_{\boldsymbol{\theta}} \ln p(\mathbf{x}_{k+1} | \mathbf{x}_k, \boldsymbol{\theta})]^T = \begin{bmatrix} \frac{1}{Q^2}(x_{k+1} - ax_k) \\ 0 \end{bmatrix}, \quad (28)$$

$$\nabla_{\boldsymbol{\theta}} [\nabla_{\boldsymbol{\theta}} \ln p(\mathbf{x}_{k+1} | \mathbf{x}_k, \boldsymbol{\theta})]^T = \begin{bmatrix} \frac{1}{2Q^2} - \frac{1}{Q^3}(x_{k+1} - ax_k)^2 & 0 \\ 0 & 0 \end{bmatrix}, \quad (29)$$

$$\nabla_{\boldsymbol{\theta}} [\nabla_{\boldsymbol{\theta}} \ln p(\mathbf{y} | \mathbf{x}_k, \boldsymbol{\theta})]^T = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2R^2} - \frac{1}{R^3}(y_k - cx_k)^2 \end{bmatrix}. \quad (30)$$

Next, matrices (11)-(16) are obtained using integral formulas (24) as

$$\mathbf{K}_{k+1}^{\theta,k} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad (31)$$

$$\mathbf{K}_{i+1}^{\theta,i+1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad (32)$$

$$\mathbf{K}_{k+1}^{\theta} = \begin{bmatrix} \frac{1}{2Q^2} & 0 \\ 0 & 0 \end{bmatrix}, \quad (33)$$

$$\mathbf{L}_k^{\theta,k} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad (34)$$

$$\mathbf{I}_k^{\theta} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2R^2} \end{bmatrix}, \quad (35)$$

$$\mathbf{A}_0 = \mathcal{E}_0 \left\{ -\nabla_{\boldsymbol{\theta}} [\nabla_{\boldsymbol{\theta}} \ln p(\boldsymbol{\theta})]^T \right\}. \quad (36)$$

It can be seen from (31)-(35) that several of the matrices are zero which significantly simplifies formula (17). In particular matrix  $\mathbf{J}_{k|k}^{\theta,x}$  in (22) is zero and this fact allows us to calculate FIM for unknown covariances independently on the states. The resulting FIM can be written as

$$\mathbf{J}_{k+1|k}^{\theta,\theta} = \mathbf{J}_{k|k-1}^{\theta,\theta} + \mathbf{L}_k^{\theta} + \mathbf{K}_{k+1}^{\theta}, \quad (37)$$

where  $\mathbf{J}_{0|-1}^{\theta,\theta}$  represents the prior information about unknown parameters  $Q, R$ . If no prior information is available, it can be set to zero. From the above equations and the form of matrices in (31)-(35), it can be seen that FIM is calculated for each covariance independently. Formula for FIM is of the form

$$\mathbf{J}_{k+1|k}^{\theta,\theta} = \mathbf{J}_{k|k-1}^{\theta,\theta} + \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2R^2} \end{bmatrix} + \begin{bmatrix} \frac{1}{2Q^2} & 0 \\ 0 & 0 \end{bmatrix}, \quad (38)$$

where it can be assumed that  $\mathbf{J}_{0|-1}^{\theta,\theta} = 0$ . A particular closed form solution to recursive formula (38) is given by

$$\mathbf{J}_{k|k-1}^{\theta,\theta} = \begin{bmatrix} \frac{k}{2Q^2} & 0 \\ 0 & \frac{k}{2R^2} \end{bmatrix}. \quad (39)$$

## 4. Cramér-Rao Bounds for a System in Innovation Form

In Section 2, a stochastic system was defined using (5). The system has two independent sources of uncertainty, the process noise and the measurement noise. It is assumed the process and the measurement noises to be normal with zero mean and covariances  $Q, R$ , respectively. In non-scalar case, variables  $\mathbf{Q}, \mathbf{R}$  are matrices and contain redundant information. Noise covariances represent properties of the noise sources. Considering a higher order system allows us to model the same number of noise sources as the sum of numbers of states and outputs. Matrix  $\mathbf{Q}$  would have the same dimension as the number of states and matrix  $\mathbf{R}$  would have a dimension equal to the number of outputs. However, from the observed output data, only the number of noise sources equal to the number of outputs can be recognized. Therefore, we can alternatively use the so-called *innovation form* of stochastic system that represents a minimal form, [11] or [12]. The disadvantage of such formulation comes from the fact that we lose the physical background of the noise sources and their structure. However, in many cases, we only have output data at hand and no other additional information about the stochastic inputs. In such case, the innovation form of the stochastic system can be used

$$\begin{aligned} x_{k+1} &= ax_k + K\varepsilon_k, \\ y_k &= cx_k + \varepsilon_k, \end{aligned} \quad (40)$$

where  $K$  is a steady state Kalman gain and  $\varepsilon(k)$  is the stationary innovation process. Then

$$\varepsilon(k) \sim \mathcal{N}(0, R_{\varepsilon}). \quad (41)$$

The Cramér-Rao bounds for  $K, R_{\varepsilon}$  estimation are derived using the same formulas as for the system (5). Analogously to formulas (27)-(30) we define

$$\nabla_{\mathbf{x}_k} [\nabla_{\boldsymbol{\theta}} \ln p(\mathbf{x}_{k+1} | \mathbf{x}_k, \boldsymbol{\theta})]^T = \begin{bmatrix} -\frac{a}{K^3 R_{\varepsilon}}(x_{k+1} - ax_k) \\ -\frac{a}{K^2 R_{\varepsilon}^2}(x_{k+1} - ax_k) \end{bmatrix}, \quad (42)$$

$$\nabla_{\mathbf{x}_{k+1}} [\nabla_{\boldsymbol{\theta}} \ln p(\mathbf{x}_{k+1} | \mathbf{x}_k, \boldsymbol{\theta})]^T = \begin{bmatrix} \frac{1}{K^2 R_{\varepsilon}}(x_{k+1} - ax_k) \\ \frac{1}{K^2 R_{\varepsilon}^2}(x_{k+1} - ax_k) \end{bmatrix}, \quad (43)$$

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} [\nabla_{\boldsymbol{\theta}} \ln p(\mathbf{x}_{k+1} | \mathbf{x}_k, \boldsymbol{\theta})]^T &= \\ &= \begin{bmatrix} \frac{1}{2K^2} - \frac{1}{R_{\varepsilon} K^4}(x_{k+1} - ax_k)^2 & -\frac{1}{R_{\varepsilon}^2 K^3}(x_{k+1} - ax_k)^2 \\ -\frac{1}{R_{\varepsilon}^2 K^3}(x_{k+1} - ax_k)^2 & \frac{1}{2R_{\varepsilon}^2} - \frac{1}{R_{\varepsilon}^3 K^2}(x_{k+1} - ax_k)^2 \end{bmatrix} \end{aligned} \quad (44)$$

$$\nabla_{\theta} \left[ \nabla_{\theta} \ln p(\mathbf{y} | \mathbf{x}_k, \theta) \right]^T = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2R_{\varepsilon}^2} - \frac{1}{R_{\varepsilon}^3} (y_k - cx_k)^2 \end{bmatrix} \quad (45)$$

The expected values are calculated using integral formulas (24). Matrices (11)-(16) are of the form

$$\mathbf{K}_{k+1}^{\theta,k} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad (46)$$

$$\mathbf{K}_{k+1}^{\theta,k+1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad (47)$$

$$\mathbf{K}_{k+1}^{\theta} = \begin{bmatrix} \frac{2}{K^2} & \frac{1}{KR_{\varepsilon}} \\ \frac{1}{KR_{\varepsilon}} & \frac{1}{2R_{\varepsilon}^2} \end{bmatrix}, \quad (48)$$

$$\mathbf{L}_k^{\theta,k} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad (49)$$

$$\mathbf{L}_k^{\theta} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2R_{\varepsilon}^2} \end{bmatrix}. \quad (50)$$

The formula for FIM calculation is of the form (37). Substituting (48) and (50) to (37) leads to the resulting formula

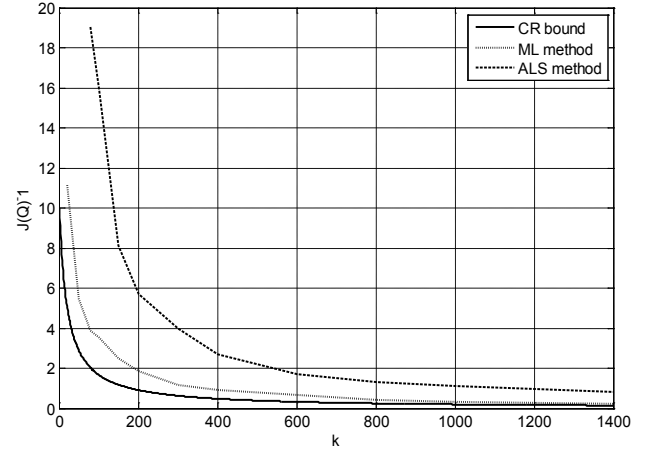
$$\mathbf{J}_k^{\theta,\theta} = \begin{bmatrix} \frac{2k}{K^2} & \frac{k}{KR_{\varepsilon}} \\ \frac{k}{KR_{\varepsilon}} & \frac{k}{R_{\varepsilon}^2} \end{bmatrix}. \quad (51)$$

## 5. Comparison of the Methods for Noise Covariance Estimation

In the previous sections, we derived the CR bounds for noise covariances estimation. In this chapter, a comparison of ALS (Autoregressive least squares) method and maximum likelihood method will be presented. The ALS method, [3-5], uses the assumption that estimated autocorrelation function calculated from the finite set of data converges to the true autocorrelation function. In [7], it is shown that the assumption of fast convergence is not correct and this fact can significantly limit the quality of the estimates. In the same paper, a maximum likelihood approach is demonstrated using a simple maximum seeking method. We can compare performance of both approaches according to the CR bounds. The ALS algorithm prior setting contains the optimal Kalman filter gain and the maximum lag for autocorrelation calculation  $N_d = 15$ . The ML method searches the maximum in the interval 0.1 to 10. The grid is logarithmic and contains  $80 \times 80$  points.

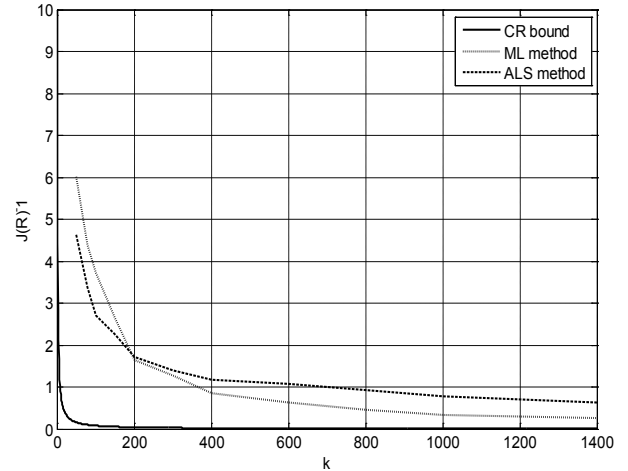
Consider a system of the form (5) where  $a = -0.5$ ,  $c = 1$ ,  $Q = 10$  and  $R = 2$ . The estimation algorithm uses  $k$  time samples and is repeated 300 times. Then, the resulting

variance of the estimates is calculated. The dependence of the variance of the estimated  $Q$ -parameters on the number of used data samples is shown in Figure 1.



**Figure 1.** Cramér-Rao bound for  $Q$  estimation and variance of estimated  $Q$  using ALS and Maximum likelihood method

In Figure 2, an analogous result for  $R$  estimation is shown. We can state that ML method converges faster to the CR bound than ALS. ML algorithm gives better estimates in case of covariance  $Q$  than ALS even with very small sets of data (around 50 samples).



**Figure 2.** Cramér-Rao bound for  $R$  estimation and variance of estimated  $R$  using ALS and Maximum likelihood method

## 6. Relationship between Cramér-Rao Bounds and Riccati Equation

In this section, we will point out the relationship between Riccati equation (RE) and CR bounds. It is well known, that under some conditions, Kalman filter is optimal. It was stated in the previous section, that CR bounds represent a limit of estimation quality. Therefore, it can be expected that RE should converge to CR bounds for state estimation of a linear system. Using formulas (8)-(10) and (14), the CR bounds for state estimation can be easily obtained, Tichavský et al. (1998). We have

$$\begin{aligned} \mathbf{K}_{k+1}^k &= \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A}, \quad \mathbf{K}_{k+1}^{k,k+1} = -\mathbf{A}^T \mathbf{Q}^{-1}, \\ \mathbf{K}_{k+1}^{k+1} &= \mathbf{Q}^{-1}, \quad \mathbf{L}_k^k = \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C}. \end{aligned} \quad (52)$$

Now, the recursive formula for FIM calculation, Šimandl et al. (2001), is of the form

$$\mathbf{J}_{k+1} = \mathbf{K}_{k+1}^{k+1} - \mathbf{K}_{k+1}^{k+1,k} (\mathbf{J}_k + \mathbf{K}_{k+1}^k + \mathbf{L}_k^k) \mathbf{K}_{k+1}^{k,k+1}. \quad (53)$$

Using (52) we obtain resulting formula

$$\mathbf{J}_{k+1} = \mathbf{Q}^{-1} - \mathbf{Q}^{-1} \mathbf{A} (\mathbf{J}_k + \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A} + \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C})^{-1} \mathbf{A}^T \mathbf{Q}^{-1}. \quad (54)$$

Riccati equation is of the form

$$\mathbf{P}_{k+1} = \mathbf{A} \mathbf{P}_k \mathbf{A}^T - \mathbf{A} \mathbf{P}_k \mathbf{C}^T (\mathbf{C} \mathbf{P}_k \mathbf{C}^T + \mathbf{R})^{-1} \mathbf{C} \mathbf{P}_k \mathbf{A}^T + \mathbf{Q}. \quad (55)$$

It can be stated, that  $\mathbf{P}_k$  and  $\mathbf{J}_k^{-1}$  converge to the same matrix. Moreover, if initial values  $\mathbf{P}_0 = \mathbf{J}_0^{-1}$ , then trajectories of  $\mathbf{P}_k$  and  $\mathbf{J}_k^{-1}$  are the same. This fact proves the optimality of Kalman filter. Equivalency of equations (54) and (55) is proved in Appendix 1.

## 7. Conclusions

In this paper, Cramér-Rao bounds for the noise covariance estimation were derived for two different stochastic models. From (31)-(35), it can be seen that matrices  $\mathbf{K}_{k+1}^{\theta,k}$ ,  $\mathbf{K}_{k+1}^{\theta,k+1}$  and  $\mathbf{L}_k^{\theta,k}$  are zero, that means that the information between state and noise covariances is not correlated. It can be concluded that the covariances and the states can be estimated independently of each other. Another observation can be done on matrices (33) and (35). It can be seen that there is no correlation between information about the covariances. That means that CR bounds can be calculated for each covariance separately. Considering limit cases,  $Q$  converging to zero and  $R$  being arbitrary large, CR bound for  $Q$  estimation will converge to zero either. It means that there exists a method that can estimate  $Q$  up to an arbitrary level of accuracy.

CR bounds can be used as a quality measure for all newly proposed estimating algorithms. We have compared a maximum likelihood approach with ALS method of Odelson et al. (2005) according to the CR bounds. Based on these examples (Figure 1 and Figure 2), it can be observed that the CR bounds for the noise covariances converge to zero relatively fast even for small sets of data (200 to 400 samples).

In Section 6, we pointed out a relationship between Fisher information matrix recursive formula and Riccati equation. Equivalence of the equations is proved in Appendix 1. We concluded that this is another proof of optimality of Kalman filter.

## APPENDIX 1

Consider two recursive equations (56) and (57) that are Fisher information matrix recursive equation and discrete Riccati equation. Both equations are equivalent and  $\mathbf{P}_k = \mathbf{J}_k^{-1}$  for sufficient large  $k$ , not necessarily  $k \rightarrow \infty$ .

$$\mathbf{J}_{k+1} = \mathbf{Q}^{-1} - \mathbf{Q}^{-1} \mathbf{A} (\mathbf{J}_k + \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A} + \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C})^{-1} \mathbf{A}^T \mathbf{Q}^{-1} \quad (56)$$

$$\mathbf{P}_{k+1} = \mathbf{A} \mathbf{P}_k \mathbf{A}^T - \mathbf{A} \mathbf{P}_k \mathbf{C}^T (\mathbf{C} \mathbf{P}_k \mathbf{C}^T + \mathbf{R})^{-1} \mathbf{C} \mathbf{P}_k \mathbf{A}^T + \mathbf{Q}. \quad (57)$$

The right side of equation (56) is modified using substitution  $\mathbf{J}_k^{-1} = \mathbf{P}_k$

$$\begin{aligned} \mathbf{Q}^{-1} - \mathbf{Q}^{-1} \mathbf{A} (\mathbf{J}_k + \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A} + \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C})^{-1} \mathbf{A}^T \mathbf{Q}^{-1} &= \\ = \mathbf{Q}^{-1} \mathbf{A} \left( \mathbf{A}^{-1} \mathbf{Q} \mathbf{A}^{-T} - (\mathbf{P}_k^{-1} + \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A} + \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C})^{-1} \right) \mathbf{A}^T \mathbf{Q}^{-1} &= \\ = \mathbf{Q}^{-1} \mathbf{A} \left( \mathbf{A}^{-1} \mathbf{Q} \mathbf{A}^{-T} (\mathbf{P}_k^{-1} + \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A} + \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C})^{-1} - 1 \right) &= \\ \times (\mathbf{P}_k^{-1} + \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A} + \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C})^{-1} \mathbf{A}^T \mathbf{Q}^{-1} &= \\ = \mathbf{Q}^{-1} \mathbf{A} \left( \mathbf{A}^{-1} \mathbf{Q} \mathbf{A}^{-T} (\mathbf{P}_k^{-1} + \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C}) \right) &= \\ \times (\mathbf{P}_k^{-1} + \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A} + \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C})^{-1} \mathbf{A}^T \mathbf{Q}^{-1} & \end{aligned}$$

Now, both sides of the equation are inverted, so the left side is  $\mathbf{P}_{k+1}$  and the right side is of the form

$$\begin{aligned} \mathbf{Q}^{-1} \mathbf{A} \left( \mathbf{A}^{-1} \mathbf{Q} \mathbf{A}^{-T} (\mathbf{P}_k^{-1} + \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C}) \right) &= \\ \times (\mathbf{P}_k^{-1} + \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A} + \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C})^{-1} \mathbf{A}^T \mathbf{Q}^{-1} &= \\ = \mathbf{Q} \mathbf{A}^{-T} (\mathbf{P}_k^{-1} + \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A} + \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C}) &= \\ \times \left( (\mathbf{P}_k^{-1} + \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C})^{-1} \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A} \right) \mathbf{A}^{-1} \mathbf{Q} &= \\ = \mathbf{Q} \mathbf{A}^{-T} (\mathbf{P}_k^{-1} + \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A} + \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C}) &= \\ \times \left( (\mathbf{P}_k^{-1} + \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C})^{-1} \mathbf{A}^T \right) &= \\ = \mathbf{A} (\mathbf{P}_k^{-1} + \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C})^{-1} \mathbf{A}^T + \mathbf{Q} & \end{aligned} \quad (58)$$

Further, an inversion lemma

$$(\mathbf{A} - \mathbf{B} \mathbf{D}^{-1} \mathbf{C})^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1} \mathbf{B} (\mathbf{D} - \mathbf{C} \mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{C} \mathbf{A}^{-1} \quad (59)$$

will be applied on the expression  $(\mathbf{P}_k^{-1} + \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C})^{-1}$ .

The final formula is of the form

$$\mathbf{P}_{k+1} = \mathbf{A} \mathbf{P}_k \mathbf{A}^T - \mathbf{A} \mathbf{P}_k \mathbf{C}^T (\mathbf{C} \mathbf{P}_k \mathbf{C}^T + \mathbf{R})^{-1} \mathbf{C} \mathbf{P}_k \mathbf{A}^T + \mathbf{Q},$$

which is exactly Riccati equation.

## ACKNOWLEDGEMENTS

This work was supported by the Czech Science Foundation, grant No. P103/11/1353.

## REFERENCES

- [1] Mehra R.K. (1970). On the identification of variances and adaptive Kalman filtering, *IEEE Trans. on Automat. Control*, AC-15, pp. 175–184.
- [2] Mehra R. K. (1972). Approaches to Adaptive Filtering, *IEEE*

*Trans. on Automat. Control*, AC-17, pp. 693–698.

- [3] Odelson B. J., Rajamani M. R. and Rawlings J. B. (2005). A new autocovariance least-squares method for estimating noise covariances, *Automatica*. Vol. 42., pp. 303–308.
- [4] Rajamani M. R. and Rawlings J. B. (2008). Estimation of the disturbance structure from data using semidefinite programming and optimal weighting, *Automatica*, Vol. 45, pp.1–7.
- [5] Rajamani M. R. and Rawlings J. B. [online] <http://jbrwww.che.wisc.edu/software/als>
- [6] Akesson B. M., Jorgensen J. B., Poulsen N. K. and Jorgensen S. B. (2008). A generalized autocovariance least-squares method for Kalman filter tuning, *Journal of Process control*, Vol. 18, pp. 769–779.
- [7] Matisko P. and Havlena V. (2010). Noise covariances estimation for Kalman filter tuning. *Proceedings of IFAC AL-COSP, Turkey*.
- [8] Tichavský P., Muravchik C. H. and Nehorai A. (1998). Posterior Cramér-Rao Bounds for Discrete-Time Nonlinear Filtering. *IEEE Trans. on Signal Processing*. Vol. 46, pp. 1386-1396.
- [9] Šimandl M., Královec J. and Tichavský P. (2001). Filtering, predictive, and smoothing Cramér-Rao bounds for discrete-time nonlinear dynamic systems. *Automatica*. Vol. 37, pp. 1703 – 1716.
- [10] Candy J. V. (2009). *Bayesian Signal Processing*, John Wiley & Sons, Inc., USA.
- [11] Ljung L. (1987). *System Identification, The Theory for The User*. Prentice Hall, USA.
- [12] Kailath T., Sayed A. H. and Hassibi B. (2000). *Linear Estimation*. Prentice Hall, USA.
- [13] Anderson B. D. O. and Moore J. N. (1979). *Optimal filtering*, Dover Publications, USA
- [14] Matisko P. (2009). *Estimation of covariance matrices of the noise of linear stochastic systems*, Diploma thesis, Czech Technical University in Prague.
- [15] Matisko P. and Havlena V. (2011). Cramér-Rao bounds for estimation of linear system noise covariances, *Proceedings of the 18<sup>th</sup> IFAC World Congress*, Italy.
- [16] Papoulis A. (1991). *Probability, Random Variables, and Stochastic Processes Third edition*, Mc. Graw-Hill, Inc., USA.