Postprint of: Niedźwiecki M., On the lower smoothing bound in identification of time-varying systems, Automatica, Vol. 44, Iss. 2 (2008), pp. 459-464, DOI: 10.1016/j.automatica.2007.05.020

On the Lower Smoothing Bound in Identification of Time-varying Systems *

Maciej Niedźwiecki [a]

[a] *Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control Gdańsk University of Technology, ul. Narutowicza 11/12, Gdańsk , Poland Tel: +48 58 3472519, Fax: +48 58 3415821, e-mail: maciekn@eti.pg.gda.pl*

Abstract

In certain applications of nonstationary system identification the model-based decisions can be postponed, i.e. executed with a delay. This allows one to incorporate in the identification process not only the currently available information, but also a number of "future" data points. The resulting estimation schemes, which involve smoothing, are not causal. Assuming that the infinite observation history is available, the paper establishes the lower steady-state estimation bound for any noncausal estimator applied to a linear system with randomly drifting coefficients (under Gaussian assumptions). This lower bound complements the currently available one, which is restricted to causal estimators.

Key words: system identification, time-varying processes

1 Introduction

Consider the problem of identification of a linear time-varying system governed by

$$y(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t) + v(t) \tag{1}$$

$$\boldsymbol{\theta}(t) = \boldsymbol{\theta}(t-1) + \mathbf{w}(t) \tag{2}$$

where $y(t)$ denotes the system output, $\boldsymbol{\varphi}(t) = [\varphi_1(t), \ldots, \varphi_n(t)]^{\mathrm{T}}$ is a known regression vector, $v(t)$ denotes white measurement noise, $\boldsymbol{\theta}(t) = [\theta_1(t), \ldots, \theta_n(t)]^{\mathrm{T}}$ is the vector of unknown and time-varying system coefficients and $\mathbf{w}(t)$ stands for the one-step parameter change.

In this paper we will focus on systems with randomly drifting parameters, namely we will assume that $\{\mathbf{w}(t)\}$ is a white noise sequance. Under such assumption (2) becomes the so-called random-walk (RW) model. Systems with RW parameter changes have been extensively studied in the literature on identification of nonstationary processes as they allow one to arrive at analytical results. Therefore, even though from the practical viewpoint the RW model may be criticized as "unrealistic", it is widely used as a benchmark for evaluation and comparison of tracking performance of different identification schemes, such as Kalman filter based (KF) algorithms, exponentially weighted least squares (EWLS) algorithms and least mean squares (LMS) algorithms – see e.g. (Guo & Ljung, 1995),(Macchi, 1995), (Haykin, 1996) and (Niedźwiecki, 2000). Naturally, evaluation of universal estimation bounds for such a benchmark problem is interesting from the theoretical viewpoint.

Under Gaussian assumptions the steady-state value of the Cramér-Rao type lower estimation bound, for the system (1) - (2) with parameters evolving according to the RW model, was established in the paper of (Ravikanth & Meyn, 1999). Since the bound derived there was restricted to causal estimation schemes, used for parameter tracking, it will be further referred to as lower tracking bound (LTB). By causal we mean estimators that specify $\widehat{\boldsymbol{\theta}}(t)$ in terms of the current and past data only: $y(s), \boldsymbol{\varphi}(s), s \leq t$. While in the adaptive prediction and control problems, studied in (Ravikanth & Meyn, 1999), causality is an obvious requirement, there are some other important applications, such as adaptive noise canceling or adaptive channel equalization, where the parameter-based decisions can be executed with a delay of a certain number of sampling intervals. In cases like this, the estimate of $\boldsymbol{\theta}(t)$ can be based not only on all past, but also on a number of *future* data points: $y(s), \boldsymbol{\varphi}(s), s > t$. Since estimation accuracy of such noncausal estimation schemes, which incorporate smoothing, exceeds accuracy of their causal counterparts, it is important to know what is the possible margin of improvement. To address this problem, we will derive expression for the steady-state value of the lower smoothing bound (LSB). Since LSB specifies the best achievable accuracy of *any* parameter estimation scheme (whether causal or not) for a time-varying system at hand, in some sense it may be considered a more fundamental limitation than LTB.

The paper is organized as follows. Section 2 summarizes the current state of knowledge about estimation bounds applicable to time-varying systems. Section 3 presents the optimal noncausal estimation scheme for identification of linear systems with parameters drifting according to the random walk model. The lower smoothing bound for such systems is established in Section 4. Section 5 describes a computationally inexpensive smoothing procedure with sub-LTB performance. Finally, Section 6 concludes.

2 Estimation bounds in identification of time-varying systems

When the system is time-invariant, i.e. $\boldsymbol{\theta}(t) = \boldsymbol{\theta}, \forall t$ (or equivalently $\mathbf{w}(t) \equiv 0$) and when the probability density function of $v(t)$ obeys some regularity conditions, the best achievable accuracy of any unbiased estimator $\widehat{\boldsymbol{\theta}}(N)$ of $\boldsymbol{\theta}$, based on the data set $\mathcal{Z} = \{y(1), \boldsymbol{\varphi}(1), \ldots, y(N), \boldsymbol{\varphi}(N)\}$, is determined by the celebrated Cramér-Rao inequality:

$$\operatorname{cov}[\widehat{\boldsymbol{\theta}}(N)|\mathcal{Z}] = \mathrm{E}\left[(\widehat{\boldsymbol{\theta}}(N) - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}(N) - \boldsymbol{\theta})^{\mathrm{T}}|\mathcal{Z}\right] \geq \mathbf{F}_N^{-1} \tag{3}$$

where

$$\mathbf{F}_N = -\mathrm{E}\left[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}} L(\boldsymbol{\theta}; \mathcal{Z})\right]$$

is the Fisher information matrix (assumed to be nonsingular), $L(\boldsymbol{\theta}; \mathcal{Z}) = \log p(\mathcal{Z}|\boldsymbol{\theta})$ denotes the log-likelihood function and averaging is carried over all realizations of the measurement noise sequence $\mathcal{V} = \{v(t), 1 \leq t \leq N\}$. Finally, $\geq$ denotes partial ordering among nonnegative-definite matrices. The matrix $\mathbf{F}_N^{-1}$ is referred to as the Cramér-Rao (lower) bound (CRB).

Suppose that

- **(A1)** The noise process $\{v(t)\}$ is an independent and identically distributed (i.i.d.) sequence with Gaussian probability density function $\mathcal{N}(0, \sigma_v^2)$.

Under (A1) the Cramér-Rao inequality for the time-invariant system takes the form

$$\operatorname{cov}[\widehat{\boldsymbol{\theta}}(N)|\mathcal{Z}] \geq \sigma_v^2 \left[\sum_{t=1}^{N} \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\right]^{-1} \tag{4}$$

The derivation of CRB (4) is based on the assumption that $\boldsymbol{\theta}$ is an unknown *deterministic* variable and that $\{\boldsymbol{\varphi}(1), \ldots, \boldsymbol{\varphi}(N)\}$ is a known *deterministic* sequence (e.g. a particular realization of a stochastic process). We will further assume that

- **(A2)** The regression vector process $\{\boldsymbol{\varphi}(t)\}$, independent of $\{v(t)\}$, is stationary and ergodic with covariance matrix $\mathrm{E}[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)] = \boldsymbol{\Phi} > 0$.

When $\{\boldsymbol{\varphi}(t)\}$ is a stochastic process and when averaging is extended to all possible realizations of $\phi = \{\boldsymbol{\varphi}(t), 1 \leq t \leq N\}$, one obtains the following result

$$\operatorname{cov}[\widehat{\boldsymbol{\theta}}(N)] \geq \mathrm{E}\left[\mathbf{F}_N^{-1}\right] \geq \left[\mathrm{E}[\mathbf{F}_N]\right]^{-1} = \mathbf{B}_{\mathrm{MCRB}} \tag{5}$$

which stems from the Jensen's inequality for matrices – see (Olkin & Pratt, 1958).

For the time-invariant system obeying (A1) and (A2) one obtains

$$\operatorname{cov}[\widehat{\boldsymbol{\theta}}(N)] \geq \sigma_v^2 \left(\mathrm{E}\left[\sum_{t=1}^{N} \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\right]\right)^{-1} = \frac{\sigma_v^2 \boldsymbol{\Phi}^{-1}}{N} \tag{6}$$

Remark

The matrix $\mathbf{B}_{\mathrm{MCRB}}$ in (5) *minorizes*, i.e. bounds from below, the mean Cramér-Rao bound $\mathrm{E}\left[\mathbf{F}_N^{-1}\right]$. Evaluation of the exact expression for the mean CRB is usually not possible. As an interesting special case, where this can be done, consider the situation where $\{\boldsymbol{\varphi}(t)\}$ is an i.i.d. Gaussian sequence. Since under such assumption the regression matrix $\sum_{t=1}^{N} \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)$ is Wishart-distributed with N degrees of freedom, from the properties of the inverted Wishart distribution it stems that

$$\sigma_v^2 \mathrm{E}\left[\left(\sum_{t=1}^{N} \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\right)^{-1}\right] = \frac{\sigma_v^2 \boldsymbol{\Phi}^{-1}}{N - n - 1}$$

Note that in this case the actual mean Cramér-Rao bound $\mathrm{E}\left[\mathbf{F}_N^{-1}\right]$ differs from $\left[\mathrm{E}[\mathbf{F}_N]\right]^{-1}$ by terms of order $o(1/N)$, i.e. the approximation is tight.

For time-invariant systems CRB decays to zero as the number of data points N increases to infinity. This is understandable since with growing N the amount of information about $\boldsymbol{\theta}$, gathered in the data set $\mathcal{Z}$, increases. For time-varying systems the situation is more complicated. First, one should realize that certain classes of time-varying systems can be reduced to time-invariant ones, i.e. they can be described, in the entire time domain, by models with constant coefficients. Consider, for example, the system (1) - (2) with constant one-step parameter changes: $\mathbf{w}(t) \equiv \mathbf{w}_o$. Note that in this case the vector $\boldsymbol{\theta}(t)$ can be written down in the form

$$\boldsymbol{\theta}(t) = \mathbf{w}_o t + \mathbf{w}_1$$

i.e. it can be parameterized in terms of time-invariant coefficients $\mathbf{w}_o$ and $\mathbf{w}_1$. Defining the generalized regression vector $\boldsymbol{\psi}(t) = [t\,\boldsymbol{\varphi}^{\mathrm{T}}(t), \boldsymbol{\varphi}^{\mathrm{T}}(t)]^{\mathrm{T}}$ and a new (time-invariant) parameter vector $\boldsymbol{\alpha} = [\mathbf{w}_o^{\mathrm{T}}, \mathbf{w}_1^{\mathrm{T}}]^{\mathrm{T}}$ one can express system equation (1) in the form

$$y(t) = \boldsymbol{\psi}^{\mathrm{T}}(t)\boldsymbol{\alpha} + v(t)$$

which falls into the analysis framework described earlier. After determining the lower bound for $\widehat{\boldsymbol{\alpha}}(N) = [\widehat{\mathbf{w}}_o^{\mathrm{T}}(N), \widehat{\mathbf{w}}_1^{\mathrm{T}}(N)]^{\mathrm{T}}$

$$\operatorname{cov}[\widehat{\boldsymbol{\alpha}}(N)] \geq \sigma_v^2 \left(\mathrm{E}\left[\sum_{t=1}^{N} \boldsymbol{\psi}(t)\boldsymbol{\psi}^{\mathrm{T}}(t)\right]\right)^{-1}$$

$$= \frac{2\sigma_v^2}{N(N-1)} \begin{bmatrix} 6\mathbf{\Phi}^{-1}/(N+1) & -3\mathbf{\Phi}^{-1} \\ -3\mathbf{\Phi}^{-1} & (2N+1)\mathbf{\Phi}^{-1} \end{bmatrix} = \mathbf{A}_N$$

one can easily compute the analogous bound for $\widehat{\boldsymbol{\theta}}(t|N) = \widehat{\mathbf{w}}_o(N)t + \widehat{\mathbf{w}}_1(N)$, $1 \leq t \leq N$:

$$\mathrm{cov}[\widehat{\boldsymbol{\theta}}(t|N)] \geq \mathbf{G}_t^{\mathrm{T}} \mathbf{A}_N \mathbf{G}_t = \mathbf{H}_{t|N}$$

where $\mathbf{G}_t = [t\mathbf{I}_n|\mathbf{I}_n]^{\mathrm{T}}$, $\mathbf{I}_n$ denotes the $n \times n$ identity matrix and

$$\mathbf{H}_{t|N} = \frac{2[6t^2 - 6t(N+1) + (N+1)(2N+1)]\sigma_v^2}{N(N-1)(N+1)} \, \mathbf{\Phi}^{-1}$$

Denote by $\mathbf{O}_n$ the $n \times n$ zero matrix. Observe that, as in the constant parameter case, it holds that $\lim_{N\mapsto\infty} \mathbf{H}_{t|N} = \mathbf{O}_n$, $\forall t \in [1, N]$, i.e. the more data is available, the more accurate the parameter estimates.

Time-varying systems that cannot be "reduced" to time-invariant ones will be called *irreducible*. Quite obviously, all systems with random coefficients fall into this category. When the parameter vector $\boldsymbol{\theta}(t)$ in (2) is a stochastic variable, the classical Cramér-Rao inequality does not apply. A bound that is similar to CRB, but can be applied to signals/systems with random coefficients, was derived by van Trees (van Trees, 1968); later on it was called posterior Cramér-Rao (lower) bound (PCRB) in (Tichavský, 1995) and (Tichavský, Muravchik & Nehorai, 1998).

Not getting into details, we will remark that PCRB can be expressed as an inverse of a posterior Fisher matrix – the sum of the expected value of the standard Fisher information matrix (averaging being carried over different realizations of $\Theta = \{\boldsymbol{\theta}(t), 1 \leq t \leq N\}$, i.e. over different realizations of $\mathcal{W} = \{\mathbf{w}(t), 1 \leq t \leq N\}$ and $\boldsymbol{\theta}(0)$), and another matrix, which represents the *a priori* information about the evolution of $\boldsymbol{\theta}(t)$. While CRB deals with unbiased estimates and depends on the likelihood function $p(\mathcal{Z}|\boldsymbol{\theta})$, PCRB applies to biased estimates (under some mild regularity conditions imposed on the bias) and depends on the joint probability density function $p(\mathcal{Z}, \boldsymbol{\theta})$.

Unlike the time-invariant and reducible time-varying system cases, discussed earlier, when $\boldsymbol{\theta}(t)$ is a nondeterministic stochastic process, the posterior Cramér-Rao bound does not decay to zero as the number of data points grows to infinity. This means that no matter how large the data size and no matter the identification technique, the parameter estimates derived for an irreducible nonstationary process will never actually converge to their true values; instead they will follow parameter changes with some finite accuracy limited by PCRB.

The stochastic version of the Cramér-Rao inequality, derived by van Trees, was successfully used to determine lower bounds for some nonlinear estimation problems, such as tracking of quasi-periodically varying signals (Tichavský, 1995), (Tichavský, Muravchik & Nehorai, 1998) and systems (Niedźwiecki & Kaczmarek, 2006). For a linear system (1) - (2) subject to Gaussian assumptions, the minimum mean-square causal estimator of $\boldsymbol{\theta}(t)$, i.e. the one that actually attains the posterior Cramér-Rao bound, can be expressed in a recursive form known as the Kalman filter. This fact was taken advantage of in (Ravikanth & Meyn, 1999), where the lower bound on the steady-state value of the PCRB was derived by means of bounding the error covariance matrices, recursively updated by the Kalman filtering algorithm. We will extend these results to the Kalman smoother which, under the same assumptions, is known to be the optimal noncausal estimator of $\boldsymbol{\theta}(t)$.

Since the measurement matrix in the state-space system description (1) – (2), equal to $\boldsymbol{\varphi}(t)$, is time-varying and data-dependent, the corresponding Cramér-Rao filtering/smoothing bounds cannot be expressed in a closed form – they have to be computed recursively using the Kalman filter/smoother covariance relationships (see (Šimandl, Královec & Tichavský, 2001) for their extension of Kalman covariance recursions to nonlinear systems). In contrast with this, the lower estimation bounds presented in (Ravikanth & Meyn, 1999) and derived below, are time-invariant and data-independent and hence they allow one to explicitly relate the limiting estimation accuracy to the second-order statistics of $\{v(t)\}$, $\{\boldsymbol{\varphi}(t)\}$ and $\{\mathbf{w}(t)\}$.

3 Optimal noncausal estimator

Consider any instant $t \in [1, N]$ and denote by $\mathcal{Z}_-(t) = \{y(1), \boldsymbol{\varphi}(1), \ldots, y(t), \boldsymbol{\varphi}(t)\} \subset \mathcal{Z}$ and $\mathcal{Z}_+(t) = \{y(t), \boldsymbol{\varphi}(t), \ldots, y(N), \boldsymbol{\varphi}(N)\} \subset \mathcal{Z}$ the sets of "past and current" and "current and future" measurements, respectively. It is well known, cf. (Lewis, 1986), that the optimal, in the mean-square sense, estimator of $\boldsymbol{\theta}(t)$ has the form

$$\widehat{\boldsymbol{\theta}}(t) = \mathrm{E}[\boldsymbol{\theta}(t)|\mathcal{Z}] \tag{7}$$

where averaging is carried over different realizations of $\mathcal{V}$, $\mathcal{W}$ and $\boldsymbol{\theta}(0)$.

Suppose that

- **(A3)** The process of one-step parameter changes $\{\mathbf{w}(t)\}$, independent of $\{v(t)\}$ and $\{\boldsymbol{\varphi}(t)\}$, is an i.i.d. sequence with Gaussian probability function $\mathcal{N}(\mathbf{0}, \mathbf{W})$; the initial parameter vector $\boldsymbol{\theta}(0)$ is a random variable, independent of $\{v(t)\}$, $\{\mathbf{w}(t)\}$ and $\{\boldsymbol{\varphi}(t)\}$, with Gaussian prior density function $\mathcal{N}(\boldsymbol{\theta}_o, \mathbf{P}_o)$.

Under (A1) - (A3) the conditional mean estimator (7) can be expressed in the form (Lewis, 1986)

$$\begin{aligned} \widehat{\boldsymbol{\theta}}(t) &= \mathbf{P}(t) \left[\mathbf{P}_-^{-1}(t|t)\widehat{\boldsymbol{\theta}}_-(t|t) + \mathbf{P}_+^{-1}(t|t+1)\widehat{\boldsymbol{\theta}}_+(t|t+1)\right] \\ &= \mathbf{P}(t) \left[\mathbf{P}_-^{-1}(t|t-1)\widehat{\boldsymbol{\theta}}_-(t|t-1) + \mathbf{P}_+^{-1}(t|t)\widehat{\boldsymbol{\theta}}_+(t|t)\right] \end{aligned} \tag{8}$$

$$\begin{aligned} \mathbf{P}(t) = \mathrm{cov}[\widehat{\boldsymbol{\theta}}(t)|\mathcal{Z}] &= \left[\mathbf{P}_-^{-1}(t|t) + \mathbf{P}_+^{-1}(t|t+1)\right]^{-1} \\ &= \left[\mathbf{P}_-^{-1}(t|t-1) + \mathbf{P}_+^{-1}(t|t)\right]^{-1} \end{aligned} \tag{9}$$

where $\widehat{\boldsymbol{\theta}}_{\pm}(t|t) = \mathrm{E}[\boldsymbol{\theta}(t)|\mathcal{Z}_{\pm}(t)]$ and $\widehat{\boldsymbol{\theta}}_{\pm}(t|t\pm 1) = \mathrm{E}[\boldsymbol{\theta}(t)|\mathcal{Z}_{\pm}(t\pm 1)]$ are the parameter estimates/predictions based on the past $(-)$ and future $(+)$ measurements, respectively, and $\mathbf{P}_{\pm}(t|t)$, $\mathbf{P}_{\pm}(t|t\pm 1)$ denote the corresponding error covariance matrices.
All quantities needed to evaluate (8) can be computed recursively using two Kalman filters: one running forward in time $(-)$ and another one, designed for the reverse-time system model, running backward in time $(+)$. Both filters can be compactly written down in the form

$$\begin{aligned}
\mathbf{P}_{\pm}(t|t\pm 1) &= \mathbf{P}_{\pm}(t\pm 1|t\pm 1) + \mathbf{W} \\
\widehat{\boldsymbol{\theta}}_{\pm}(t|t\pm 1) &= \widehat{\boldsymbol{\theta}}_{\pm}(t\pm 1|t\pm 1) \\
\mathbf{k}_{\pm}(t) &= \frac{\mathbf{P}_{\pm}(t|t\pm 1)\boldsymbol{\varphi}(t)}{\sigma_v^2 + \boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{P}_{\pm}(t|t\pm 1)\boldsymbol{\varphi}(t)} \\
\mathbf{P}_{\pm}(t|t) &= \mathbf{P}_{\pm}(t|t\pm 1) \\
&\quad - \frac{\mathbf{P}_{\pm}(t|t\pm 1)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{P}_{\pm}(t|t\pm 1)}{\sigma_v^2 + \boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{P}_{\pm}(t|t\pm 1)\boldsymbol{\varphi}(t)} \\
\varepsilon_{\pm}(t) &= y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_{\pm}(t|t\pm 1) \\
\widehat{\boldsymbol{\theta}}_{\pm}(t|t) &= \widehat{\boldsymbol{\theta}}_{\pm}(t|t\pm 1) + \mathbf{k}_{\pm}(t)\varepsilon_{\pm}(t)
\end{aligned} \tag{10}$$

The initial conditions should be set to $\widehat{\boldsymbol{\theta}}_{-}(0|0) = \boldsymbol{\theta}_o$, $\mathbf{P}_{-}(0|0) = \mathbf{P}_o$ – for the forward filter, and $\widehat{\boldsymbol{\theta}}_{+}(N|N) = \mathbf{0}$, $\mathbf{P}_{+}^{-1}(N|N) = \mathbf{O}_n$ – for the backward filter. According to (10) the optimal parameter estimate $\widehat{\boldsymbol{\theta}}(t)$ can be obtained as a linear combination of the estimates yielded by the forward Kalman filter $\widehat{\boldsymbol{\theta}}_{-}(t|t)$ and the backward Kalman predictor $\widehat{\boldsymbol{\theta}}_{+}(t|t+1)$, or equivalently, by combining results provided by the forward Kalman predictor $\widehat{\boldsymbol{\theta}}_{-}(t|t-1)$ and the backward Kalman filter $\widehat{\boldsymbol{\theta}}_{+}(t|t)$.

4 Lower smoothing bound

Since $\widehat{\boldsymbol{\theta}}(t)$, given by (8), is the optimal estimator, the minimum attainable error covariance matrix is equal to $\mathrm{E}[\mathbf{P}(t)]$, where averaging is carried over different realizations of ϕ. To arrive at steady-state expressions we will assume that an infinite observation history is available, incorporating all past and all future data samples, i.e. that: $\mathcal{Z} = \{y(s), \boldsymbol{\varphi}(s), -\infty < s < \infty\}$, $\mathcal{Z}_{-}(t) = \{y(s), \boldsymbol{\varphi}(s), -\infty < s < t\}$ and $\mathcal{Z}_{+}(t) = \{y(s), \boldsymbol{\varphi}(s), t < s < \infty\}$. The corresponding steady-state expectation will be denoted by E_∞.
To arrive at the expression for the steady-state lower estimation bound we will exploit results derived in (Ravikanth & Meyn, 1999) for causal estimators. Given that the infinite observation history is available, these results hold for both forward and backward Kalman filters. Following Ravikanth & Meyn we will assume that

(A4) The sequences of covariance matrices $\{\mathbf{P}_{\pm}(t|t)\}$ and $\{\mathbf{P}_{\pm}(t|t\pm 1)\}$ are asymptotically stationary.

As pointed out in (Ravikanth & Meyn, 1999), stationarity of $\{\mathbf{P}_{\pm}(t|t)\}$ and $\{\mathbf{P}_{\pm}(t|t\pm 1)\}$ is warranted if, in addition to (A2), some stochastic persistence of excitation conditions are imposed on $\{\boldsymbol{\varphi}(t)\}$, such as conditions formulated in (Guo, 1990), for example.
Let $\mathbf{P}_\infty = \mathrm{E}_\infty\left[\mathbf{P}_{\pm}(t|t\pm 1)\right]$, $\mathbf{P}^*_\infty = \mathrm{E}_\infty\left[\mathbf{P}_{\pm}(t|t)\right]$, $\mathbf{Q}_\infty = \mathrm{E}_\infty\left[\mathbf{P}_{\pm}^{-1}(t|t\pm 1)\right]$ and $\mathbf{Q}^*_\infty = \mathrm{E}_\infty\left[\mathbf{P}_{\pm}^{-1}(t|t)\right]$. According to (Ravikanth & Meyn, 1999), under (A1) – (A4) it holds that

$$\mathbf{P}_\infty \geq \mathbf{Q}_\infty^{-1} \geq \mathbf{X}_\infty^{-1} \tag{11}$$

where $\mathbf{X}_\infty$ is the positive definite matrix satisfying the equation

$$\mathbf{X}_\infty \mathbf{W} \mathbf{X}_\infty = \frac{1}{\sigma_v^2}\boldsymbol{\Phi} \tag{12}$$

Since $\mathbf{P}_{-}(t|t-1) = \mathbf{P}_{-}(t-1|t-1) + \mathbf{W}$, one arrives at

$$\mathbf{P}^*_\infty = \mathbf{P}_\infty - \mathbf{W} \geq \mathbf{X}_\infty^{-1} - \mathbf{W}$$

This means that for any causal estimator of $\boldsymbol{\theta}(t)$ the matrix that minorizes the steady-state Cramér-Rao tracking (lower) bound can be written down in the form

$$\mathbf{B}_{\mathrm{LTB}} = \mathbf{X}_\infty^{-1} - \mathbf{W} \tag{13}$$

As demonstrated in (Ravikanth & Meyn, 1999), the above bound is tight when system parameters change slowly with time and when $\{\boldsymbol{\varphi}(t)\}$ is a weakly dependent sequence.
To arrive at the analogous smoothing bound note that, according to (9) and Jensen's inequality, it holds that

$$\begin{aligned}
\mathrm{E}_\infty[\mathbf{P}(t)] &= \mathrm{E}_\infty\left\{\left[\ \mathbf{P}_{-}^{-1}(t|t-1) + \mathbf{P}_{+}^{-1}(t|t)\ \right]^{-1}\right\} \\
&\geq \left[\ \mathrm{E}_\infty[\mathbf{P}_{-}^{-1}(t|t-1)] + \mathrm{E}_\infty[\mathbf{P}_{+}^{-1}(t|t)]\ \right]^{-1}
\end{aligned} \tag{14}$$

From (11) it follows that

$$\mathrm{E}_\infty[\mathbf{P}_{-}^{-1}(t|t-1)] = \mathbf{Q}_\infty \leq \mathbf{X}_\infty \tag{15}$$

Using the matrix inversion lemma (Söderström & Stoica, 1988) one can rewrite the covariance update equation

$$\begin{aligned}
\mathbf{P}_{+}(t|t) &= \mathbf{P}_{+}(t|t+1) \\
&\quad - \frac{\mathbf{P}_{+}(t|t+1)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{P}_{+}(t|t+1)}{\sigma_v^2 + \boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{P}_{+}(t|t+1)\boldsymbol{\varphi}(t)}
\end{aligned}$$

in the form $\mathbf{P}_{+}^{-1}(t|t) = \mathbf{P}_{+}^{-1}(t|t+1) + \frac{1}{\sigma_v^2}\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)$, leading to

$$\mathrm{E}_\infty[\mathbf{P}_{+}^{-1}(t|t)] = \mathbf{Q}^*_\infty = \mathbf{Q}_\infty + \frac{1}{\sigma_v^2}\boldsymbol{\Phi} \leq \mathbf{X}_\infty + \frac{1}{\sigma_v^2}\boldsymbol{\Phi} \tag{16}$$

Combining (9), (14), (15) and (16) one arrives at the following inequality, which establishes the lower smoothing bound for all estimators of $\boldsymbol{\theta}(t)$, including noncausal ones

$$\begin{aligned}
&\mathrm{E}\left[(\widehat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t))(\widehat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t))^{\mathrm{T}}\right] \\
&\geq \left[2\mathbf{X}_\infty + \frac{1}{\sigma_v^2}\boldsymbol{\Phi}\right]^{-1} = \mathbf{B}_{\mathrm{LSB}}
\end{aligned} \tag{17}$$

When system parameters change sufficiently slowly with time, namely

$$\mathbf{W} \ll \mathbf{X}_\infty^{-1} \tag{18}$$

it holds that $\mathbf{B}_{\mathrm{LTB}} \cong \mathbf{X}_\infty^{-1}$. Similarly, since the condition (18) entails $\mathbf{X}_\infty \mathbf{W} \mathbf{X}_\infty \ll \mathbf{X}_\infty$ and consequently (cf. (12)) $(1/\sigma_v^2)\boldsymbol{\Phi} \ll \mathbf{X}_\infty$, one obtains $\mathbf{B}_{\mathrm{LSB}} \cong (1/2)\mathbf{X}_\infty^{-1}$. Hence, under the slow variation condition one arrives at

$$\mathbf{B}_{\mathrm{LSB}} \cong \frac{1}{2}\mathbf{B}_{\mathrm{LTB}} \tag{19}$$

Remark 1

The model of parameter variation adopted in (Ravikanth & Meyn, 1999) has the form $\boldsymbol{\theta}(t) = f\boldsymbol{\theta}(t-1) + \mathbf{w}(t)$, $0 < f \leq 1$, i.e. it is more general than (2). Since the reverse-time model for such a Markov process is $\boldsymbol{\theta}(t-1) = f_\star\boldsymbol{\theta}(t) + \mathbf{w}_\star(t)$, where $\mathbf{w}_\star(t) = \mathbf{w}(t)/f$ and $f_\star = 1/f \geq 1$, only the RW case ($f = f_\star = 1$) can be analyzed in the way described above. According to (19), when $f = 1$, incorporation of future measurements allows one to halve the lower estimation bound. This means that inclusion of future measurements effectively doubles information content of the analyzed data, i.e. there is as much *new* information about $\boldsymbol{\theta}(t)$ in the future observation history as in the past. Generally, one can expect that the margin of improvement should decrease along with f – when the coefficient f is small, i.e. when the sequence $\{\boldsymbol{\theta}(t)\}$ is rapidly mixing, the improvement achievable by means of smoothing may be negligible.

Remark 2

Let $\Delta y(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)(\boldsymbol{\theta}(t) - \boldsymbol{\theta}(t-1)) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{w}(t)$. The scalar coefficient

$$\eta(t) = \sqrt{\frac{\mathrm{E}[(\Delta y(t))^2]}{\sigma_v^2}}$$

was proposed in (Macchi, 1995) as a measure of nonstationarity of the system governed by (1) - (2). According to Macchi, a system can be regarded as slowly time-varying if it obeys the condition $\eta(t) \ll 1, \forall t$. Note that in the case considered this condition is equivalent to

$$\eta = \sqrt{\frac{\mathrm{tr}\{\boldsymbol{\Phi}\mathbf{W}\}}{\sigma_v^2}} \ll 1 \tag{20}$$

It is straightforward to show that when either $\boldsymbol{\Phi}$ or $\mathbf{W}$ are similar to identity matrices, the condition (20) entails (18). Suppose, for example, that $\mathbf{W} = \sigma_w^2\mathbf{I}_n$ which means that system coefficients evolve independently of each other, with the same mean-square rate of change. Note that in this special case equation (12) can be easily solved for $\mathbf{X}_\infty$

$$\mathbf{X}_\infty = \frac{1}{\sigma_w\sigma_v}\boldsymbol{\Phi}^{1/2}$$

where the positive definite matrix $\boldsymbol{\Phi}^{1/2}$, obeying $\boldsymbol{\Phi}^{1/2}\boldsymbol{\Phi}^{1/2}=\boldsymbol{\Phi}$ is the (unique) square-root of $\boldsymbol{\Phi}$. Condition (18) is then equivalent to $\sigma_w^2\mathbf{I}_n \ll \sigma_w\sigma_v\boldsymbol{\Phi}^{-1/2}$, i.e. to $(\sigma_w/\sigma_v)\boldsymbol{\Phi}^{1/2} \ll \mathbf{I}_n$. Denote by $\lambda_{\max}(\boldsymbol{\Phi})$ the maximum eigenvalue of $\boldsymbol{\Phi}$. Since $\boldsymbol{\Phi}^{1/2} \leq \sqrt{\lambda_{\max}(\boldsymbol{\Phi})}\ \mathbf{I}_n < \sqrt{\mathrm{tr}\{\boldsymbol{\Phi}\}}\ \mathbf{I}_n$, a sufficient condition for (18) to hold is $(\sigma_w/\sigma_v)\sqrt{\mathrm{tr}\{\boldsymbol{\Phi}\}} \ll 1$, which in the case considered is identical with (20).
The case where $\boldsymbol{\Phi} = \sigma_\varphi^2\mathbf{I}_n$, i.e.

$$\mathbf{X}_\infty = \frac{\sigma_\varphi}{\sigma_v}\mathbf{W}^{-1/2}$$

can be handled in an analogous way.

5 Some practical issues

The lower smoothing bound (17) was derived for an infinite-lag smoother which is not realizable. In practice, instead of $\widehat{\boldsymbol{\theta}}(t) = \mathrm{E}[\boldsymbol{\theta}(t)|\mathcal{Z}]$, one can use a fixed-lag smoother $\widehat{\boldsymbol{\theta}}(t-\tau|t) = \mathrm{E}[\boldsymbol{\theta}(t-\tau)|\mathcal{Z}_-(t)]$, where τ is the permissible decision delay. It can be shown that

$$\mathrm{cov}[\widehat{\boldsymbol{\theta}}(t-\tau|t)] \leq \mathrm{cov}[\widehat{\boldsymbol{\theta}}(t-\tau|t-\tau)],\ \forall\tau \geq 0$$

Fixed-lag smoothing can be realized using a Kalman filtering algorithm designed for an augmented state space model of parameter variation (Lewis, 1986). Denote by $\boldsymbol{\theta}_a(t) = [\boldsymbol{\theta}^{\mathrm{T}}(t), \ldots, \boldsymbol{\theta}^{\mathrm{T}}(t-\tau)]^{\mathrm{T}}$ the augmented state vector. Note that

$$\widehat{\boldsymbol{\theta}}_a(t|t) = \mathrm{E}[\theta_a(t)|\mathcal{Z}_-(t)] = [\widehat{\boldsymbol{\theta}}^{\mathrm{T}}(t|t), \ldots, \widehat{\boldsymbol{\theta}}^{\mathrm{T}}(t-\tau|t)]^{\mathrm{T}}$$

which means that the smoothed estimate $\widehat{\boldsymbol{\theta}}(t-\tau|t)$ can be “extracted” from the augmented parameter estimate yielded by the Kalman filter designed to track $\boldsymbol{\theta}_a(t)$. Although conceptually very simple, the fixed-lag smoother described above may be computationally very demanding, even for moderate values of τ.
As pointed out in the classical paper of Hedelin (Hedelin, 1977), delaying the state estimates provided by the Kalman filter can be often regarded a suboptimal form of smoothing. We will show that the same “trick” can be used to obtain computationally attractive smoothing procedures for identification of time-varying systems.
As an example consider the following two-tap finite impulse response (FIR) system, inspired by channel equalization applications

$$\begin{aligned} y(t) &= \theta_1(t)u(t) + \theta_2(t)u(t-1) + v(t) \\ \theta_1(t) &= \theta_1(t-1) + w_1(t) \\ \theta_2(t) &= \theta_2(t-1) + w_2(t) \end{aligned}$$

where $v(t) \sim \mathcal{N}(0, \sigma_v^2)$, $[w_1(t), w_2(t)]^{\mathrm{T}} \sim \mathcal{N}(\mathbf{0}, \sigma_w^2\mathbf{I}_2)$ and $\{u(t)\}$ is a pseudo-random binary input sequence (PRBS): $u(t) = \pm 1$, $\sigma_u^2 = 1$. Note that in this case $\boldsymbol{\theta}(t) = [\theta_1(t), \theta_2(t)]^{\mathrm{T}}$ and $\boldsymbol{\varphi}(t) = [u(t), u(t-1)]^{\mathrm{T}}$.

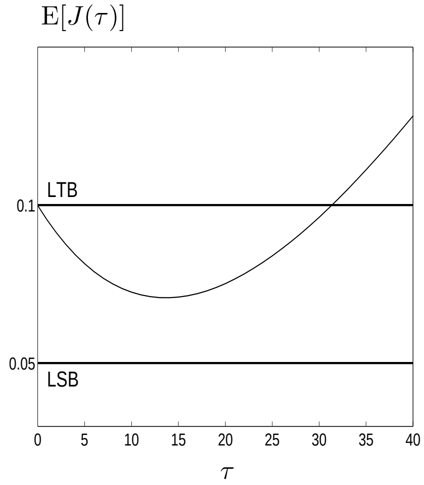

Fig. 1. Dependence of the value of the parameter matching error $\mathrm{E}[J(\tau)]$ on the delay τ. The limiting values of the mean-square parameter estimation errors, which follow from the lower tracking bound (LTB) and from the lower smoothing bound (LSB), are indicated by horizontal lines.

To check how well the estimated parameter trajectory $\{\widehat{\boldsymbol{\theta}}_{-}(t|t)\}$, yielded by the forward Kalman filter, matches the *delayed* true parameter trajectory $\{\boldsymbol{\theta}(t-\tau)\}$, the following index was computed

$$J(\tau) = \frac{1}{1000} \sum_{t=2001}^{3000} ||\, \widehat{\boldsymbol{\theta}}_{-}(t|t) - \boldsymbol{\theta}(t-\tau)\, ||^2 \qquad (21)$$

and averaged over 1000 different realizations of $\{u(t), v(t), w(t)\}$. Note that summation in (21) starts at the instant $t = 2001$, large enough to guarantee that the KF algorithm reaches the steady-state behavior before its performance is evaluated.

Figure 1 shows how the average value of $J(\tau)$ depends on the delay τ in the case where $\sigma_v^2 = 1$ and $\sigma_w^2 = 0.0025$. It is straightforward to check that in the case considered $\mathbf{B}_{\mathrm{LTB}} \cong (\sigma_w \sigma_v / \sigma_u)\mathbf{I}_2 = 0.05\mathbf{I}_2$ and hence, according to (17), it holds that

$$\mathrm{E}[J(0)] = \mathrm{E}[||\, \widehat{\boldsymbol{\theta}}_{-}(t|t) - \boldsymbol{\theta}(t)\, ||^2] \geq \mathrm{tr}\{\mathbf{B}_{\mathrm{LTB}}\} \cong 0.1$$

The analogous bound for noncausal estimators, including all fixed-lag smoothers $\widehat{\boldsymbol{\theta}}(t-\tau|t), \tau > 0$, is equal to $\mathrm{tr}\{\mathbf{B}_{\mathrm{LSB}}\} \cong 0.05$.

According to Figure 1, to achieve the sub-LTB performance in a computationally cheap way it is sufficient to regard $\widehat{\boldsymbol{\theta}}_{-}(t|t)$ as an estimate of the past parameter value $\boldsymbol{\theta}(t-\tau)$, rather than as an estimate of its current value $\boldsymbol{\theta}(t)$, where τ is a judiciously chosen delay which must be incorporated in the decision loop. More sophisticated "cheap smoothing" identification schemes were recently proposed in (Niedźwiecki, 2007).

6 Conclusion

We have considered the problem of identification of a linear time-varying system with randomly drifting coefficients and we have established the Cramér-Rao type lower estimation bound (LSB), which limits accuracy of *any* estimator of system parameters, including noncausal estimators. The obtained results complement those derived earlier for causal estimation schemes (LTB). Additionally, we have shown that the sub-LTB performance can be achieved in a very simple way by means of delaying parameter estimates yielded by the Kalman filter based tracker.

References

Guo, L. (1990). Estimating time-varying parameters by the Kalman filter based algorithm: Stability and convergence. *IEEE Trans. Automat. Contr.*, **vol. 35**, 141–147.

Guo, L. & L. Ljung (1995). Performance analysis of general tracking algorithms. *IEEE Trans. Automat. Contr.*, **vol. 40**, 1388–1402.

Haykin, S. (1996). *Adaptive Filter Theory.* Prentice Hall, Englewood Cliffs NJ. MIT Press. Cambridge, Mass.

Hedelin, P. (1977). Can the zero-lag filter be a good smoother? *IEEE Trans. on Information Theory*, **vol. 23**, 490–499.

Lewis, F.L. (1986). *Optimal Estimation.* Wiley. New York.

Macchi, O. (1995). *Adaptive Processing: The Least Mean Squares Approach with Applications in Transmission.* Wiley. New York.

Niedźwiecki, M. (2000). *Identification of Time-varying Processes.* Wiley. New York.

Niedźwiecki, M. & P. Kaczmarek (2006). Tracking analysis of generalized adaptive notch filters. *IEEE Trans. on Signal Processing*, **vol. 54**, 304–314.

Niedźwiecki, M. (2007). On 'cheap smoothing' opportunities in identification of time-varying systems. Gdańsk University of Technology, Technical Report # ETI-2/2007 (submitted for publication), http://www.eti.pg.gda.pl/katedry/ksa/pracownicy/Maciej Niedzwiecki.

Olkin, I. & J. Pratt (1958). A multivariate Tchebycheff inequality. *Ann. Math. Statist.*, **vol. 29**, 226–234.

Ravikanth, R. & S.P. Meyn (1999). Bounds on achievable performance in the identification and adaptive control of time-varying systems. *IEEE Trans. Automat. Contr.*, **vol. 44**, 670–682.

Šimandl, M., J. Královec & P. Tichavský (2001). Filtering, predictive and smoothing Cramér-Rao bounds for discrete-time nonlinear dynamic systems. *Automatica*, **vol. 37**, 1703–1716.

Söderström, T. & P. Stoica (1988). *System Identification.* Prentice Hall. Englewood Cliffs NJ.

Tichavský, P. (1995). Posterior Cramér-Rao bound for adaptive harmonic retrieval. *IEEE Trans. on Signal Processing*, **vol. 43**, 1299–1302.

Tichavský, P., C.H. Muravchik & A. Nehorai (1998). Posterior Cramér-Rao bounds for discrete-time nonlinear filtering problems. *IEEE Trans. on Signal Processing*, **vol. 46**, 1386–1396.

van Trees H.L. (1968). *Detection, Estimation and Modulation Theory.* Wiley. New York.