Automatica 58 (2015) 22–27

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

On the end-performance metric estimator selection☆

Dimitrios Katselis[a], Cristian R. Rojas[b], Boris I. Godoy[c], Juan C. Agüero[d], Carolyn L. Beck[a]

[a] Coordinated Science Laboratory & Industrial and Enterprise Systems Engineering, University of Illinois at Urbana–Champaign, IL, 61801, Urbana, USA

[b] Department of Automatic Control and ACCESS Linnaeus Center, KTH Royal Institute of Technology, SE-100 44, Stockholm, Sweden

[c] School of Electrical Engineering and Telecommunications, The University of New South Wales, Kensington, NSW 2052, Australia

[d] Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile

ARTICLE INFO

Article history:
Received 13 June 2014
Received in revised form
28 November 2014
Accepted 13 April 2015
Available online 18 May 2015

Keywords:
Bias selection
End-performance metric
Estimation
Maximum likelihood
Least squares

ABSTRACT

It is well known that appropriately biasing an estimator can potentially lead to a lower mean square error (MSE) than the achievable MSE within the class of unbiased estimators. Nevertheless, the choice of an appropriate bias is generally unclear and only recently there have been attempts to systematize such a selection. These systematic approaches aim at introducing MSE bounds that are lower than the unbiased Cramér–Rao bound (CRB) for all values of the unknown parameters and at choosing biased estimators that beat the standard maximum-likelihood (ML) and/or least squares (LS) estimators in the finite sample case. In this paper, we take these approaches one step further and investigate the same problem from the aspect of an end-performance metric different than the classical MSE. This study is motivated by recent advances in the area of system identification indicating that the optimal experiment design should be done by taking into account the end-performance metric of interest and not by quantifying a quadratic distance of the unknown model from the true one.

1. Introduction

In the context of application-oriented experiment design, experiments are designed to optimize a performance metric associated with a particular application for which the estimated model will be used; see, e.g., Barenthin, Bombois, Hjalmarsson, and Scorletti (2008), Forssell and Ljung (2000), Gevers and Ljung (1986), Katselis, Rojas, Hjalmarsson, and Bengtsson (2012) and the references therein. A conceptual framework for application-oriented experiment design was outlined in Hjalmarsson (2009). Here, a function J quantifying the end-performance metric of the system is used to select the unknown system model. The generic description of such an experiment design is given by the following formulation:

$$\begin{aligned} \min_{\text{Experiment}} \quad & J \\ \text{s.t.} \quad & \hat{G} \in \mathcal{E}_{adm}, \end{aligned} \tag{1}$$

where J is a performance metric of interest dependent on a model G. The set of admissible models is denoted as $\mathcal{E}_{adm} = \{G : J \leq 1/\gamma\}$. The parameter γ is usually referred to in this context as *accuracy*. For the experimental effort, different measures commonly used are input or output power, and experimental length. For $\hat{G}$, standard maximum likelihood (ML) and Bayesian estimation methods are usually employed.

Assuming that the unknown model is parameterized by the vector $\boldsymbol{\theta}_0$ and that its corresponding estimate is $\hat{\boldsymbol{\theta}}$, the achieved end-performance metric is $J(\hat{\boldsymbol{\theta}})$. Clearly, this quantity is a random variable and a natural deterministic version of it, meaningful in measuring the system's performance, is its expected value $\mathbb{E}[J(\hat{\boldsymbol{\theta}})]$. As the Cramér–Rao bound (CRB) is a bound of the mean square error (MSE) within the class of unbiased estimators, an immediate problem of interest is to examine the possibility of devising lower bounds for $\mathbb{E}[J(\hat{\boldsymbol{\theta}})]$ and (possibly biased) estimators that achieve these bounds. This paper is a first effort to investigate this problem and to shed light into the gains in terms of end performance that

☆ Juan C. Agüero was partially supported by the Chilean Research Council (CONICYT) through Basal Project FB0008, and FONDECYT (grant no. 1150954). The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Martin Enqvist under the direction of Editor Torsten Söderström.

E-mail addresses: katselis@illinois.edu (D. Katselis), cristian.rojas@ee.kth.se (C.R. Rojas), b.godoy@unsw.edu.au (B.I. Godoy), juan.aguero@usm.cl (J.C. Agüero), beck3@illinois.edu (C.L. Beck).

http://dx.doi.org/10.1016/j.automatica.2015.04.026
0005-1098/

can be achieved by such estimator designs. To this end, we follow a similar line of reasoning developed in Eldar (2008) for the classical MSE metric. For the linear Gaussian model of measurements, biased estimators can be developed that outperform the classical ML estimator in the sense of yielding a lower value for $\mathbb{E}[J(\hat{\boldsymbol{\theta}})]$. For simplicity, we focus in this paper on linear and quadratic performance metrics or performance metrics that can be well approximated by first and second order Taylor series expansions.

As a remark, we note that there exist in the literature approaches aiming to improve the estimation performance in classical contexts by incorporating into the estimator information available in the form of linear or nonlinear constraints. Significant studies in this direction are provided in Gorman and Hero (1990) and more recently in Mahata and Söderström (2004). In addition, there is some recent work on the use of biased estimators in system identification (Chen, Ohlsson, & Ljung, 2012; Pillonetto & De Nicolao, 2010), where the main focus is on the final MSE (or 'model fit') of the estimates, instead of considering an application-specific metric. The proposed framework in this paper is mostly inspired by the results in Eldar (2008), and capitalizes on the application-oriented designs developed in Barenthin et al. (2008), Forssell and Ljung (2000), Gevers and Ljung (1986), and Katselis et al. (2012).

This paper is organized as follows: Classical end-performance estimator designs within the class of unbiased estimators are discussed in Section 2 and subsections therein. Section 3 examines the biased estimator selection problem within the end-performance framework specializing in the class of linear-bias estimators. Some results on the non-Gaussian data assumption are given in Section 4. Simulations are provided in Section 5 and conclusions in Section 6.

2. Unbiased estimators

In the following, we investigate the problem of bounding the achievable end-performance metric within the class of unbiased estimators. The same study will be performed in a later section for specific cases of biased estimators.

2.1. The classical approach

Here, we will consider the linear Gaussian model to provide intuition on the developed theory. Consider the measurement model:

$$y_i = \boldsymbol{h}_i^T \boldsymbol{\theta}_0 + w_i, \quad i = 1, 2, \ldots, N, \tag{2}$$

where $\boldsymbol{h}_i \in \mathbb{R}^L$ is a known deterministic regression vector, $\boldsymbol{\theta}_0 \in \mathbb{R}^L$ is the unknown parameter vector and $\boldsymbol{w} = [w_1, \ldots, w_N]^T \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ is the measurement noise. We also use the standard assumption that $L \leq N$. The aggregate data is expressed as the usual linear Gaussian regression:

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{\theta}_0 + \boldsymbol{w}, \tag{3}$$

where $\boldsymbol{y} = [y_1, y_2, \ldots, y_N]^T \in \mathbb{R}^N$ and $\boldsymbol{H} = [\boldsymbol{h}_1 \boldsymbol{h}_2 \cdots \boldsymbol{h}_N]^T \in \mathbb{R}^{N\times L}$. A more general case is

$$\boldsymbol{y} = \boldsymbol{x}_{\boldsymbol{\theta}_0} + \boldsymbol{w}, \tag{4}$$

where $\boldsymbol{x}_{\boldsymbol{\theta}_0}$ denotes an $N \times 1$ random vector, which depends on an $L \times 1$ unknown deterministic parameter vector.

If the estimator of $\boldsymbol{\theta}_0$ is denoted by $\hat{\boldsymbol{\theta}}$, then it is characterized by the bias $\boldsymbol{b}(\boldsymbol{\theta}_0) = \mathbb{E}[\hat{\boldsymbol{\theta}}] - \boldsymbol{\theta}_0$ and the covariance matrix $\boldsymbol{C}(\boldsymbol{\theta}_0) = \mathbb{E}\left[(\hat{\boldsymbol{\theta}} - \mathbb{E}[\hat{\boldsymbol{\theta}}])(\hat{\boldsymbol{\theta}} - \mathbb{E}[\hat{\boldsymbol{\theta}}])^T\right]$. The CRB corresponds to a lower bound on the covariance $\boldsymbol{C}(\boldsymbol{\theta}_0)$ within the class of unbiased estimators.

Given our assumptions on the system in (3), i.e., linearity and Gaussianity, it is known that the minimum variance unbiased (MVU), ML and least squares (LS) estimators coincide, and that they are given by the well-known expression

$$\hat{\boldsymbol{\theta}}_{ML} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{y}. \tag{5}$$

Furthermore, it is known that $\hat{\boldsymbol{\theta}}_{ML}$ attains the CRB (Kay, 1993).

2.2. The end-performance metric approach

Within the end-performance metric framework, similar bounds have to be derived with respect to $\mathbb{E}[J(\hat{\boldsymbol{\theta}})]$. To this end, we state the following result:

Theorem 1. *Let $J : \mathcal{B}_{\boldsymbol{\theta}_0}(r) \to \mathbb{R}^+$ be an end-performance metric defined in the ball $\mathcal{B}_{\boldsymbol{\theta}_0}(r) = \{\boldsymbol{\theta} \in \mathbb{R}^L | \, \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < r\}$. J is assumed to have at least twice continuously differentiable partial derivatives. Then, within the class of unbiased estimators $\hat{\boldsymbol{\theta}} \in \mathcal{B}_{\boldsymbol{\theta}_0}(r)$ of $\boldsymbol{\theta}_0$, one has that*

$$\mathbb{E}[J(\hat{\boldsymbol{\theta}})] = J(\boldsymbol{\theta}_0) + \frac{1}{2}\mathbb{E}\left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \nabla^2 J(\boldsymbol{\theta}_0 + \zeta(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0))(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right], \tag{6}$$

where ζ is a random variable in $(0, 1)$. Here, the expectation is with respect to $\hat{\boldsymbol{\theta}}$ and ζ.

Proof. Given $\hat{\boldsymbol{\theta}} \in \mathcal{B}_{\boldsymbol{\theta}_0}(r)$, let $\boldsymbol{\lambda} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$, so $\|\boldsymbol{\lambda}\| < r$, and define $\phi(t) = J(\boldsymbol{\theta}_0 + t\boldsymbol{\lambda})$. Then

$$\phi(0) = J(\boldsymbol{\theta}_0), \qquad \phi(1) = J(\boldsymbol{\theta}_0 + \boldsymbol{\lambda}) = J(\hat{\boldsymbol{\theta}}). \tag{7}$$

From Taylor's theorem,

$$\phi(1) = \phi(0) + \phi'(0) + \frac{1}{2}\phi''(\zeta), \quad \text{for some } \zeta \in (0, 1). \tag{8}$$

Furthermore, we have $\phi'(t) = \nabla J(\boldsymbol{\theta}_0 + t\boldsymbol{\lambda})^T\boldsymbol{\lambda}$ and $\phi''(t) = \boldsymbol{\lambda}^T\nabla^2 J(\boldsymbol{\theta}_0 + t\boldsymbol{\lambda})\boldsymbol{\lambda}$. Then,

$$J(\hat{\boldsymbol{\theta}}) = J(\boldsymbol{\theta}_0) + \nabla J(\boldsymbol{\theta}_0)^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\nabla^2 J(\boldsymbol{\theta}_0 + \zeta(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0))(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0). \tag{9}$$

Notice that $\zeta \in (0, 1)$ is a random variable (as it depends implicitly on $\hat{\boldsymbol{\theta}}$). The measurability of ζ follows from a simple extension of Gourieroux and Monfort (1995, Thm. 24.1). Now, taking the expectation in (9) and using the assumption of unbiasedness of $\hat{\boldsymbol{\theta}}$ (which makes the expectation of the linear term equal to zero), we obtain (6). This completes the proof. □

Remark 2. If J is linear,[1] then $\mathbb{E}[J(\hat{\boldsymbol{\theta}})] = J(\boldsymbol{\theta}_0)$. This performance can be achieved by any unbiased estimator. For J to be a legitimate end-performance metric, we assume that $J \geq 0$ on $\mathcal{B}_{\boldsymbol{\theta}_0}(r)$.

Remark 3. If J is a quadratic function[2] of the form $J(\boldsymbol{\gamma}) = \boldsymbol{\gamma}^T\boldsymbol{A}\boldsymbol{\gamma} + \boldsymbol{b}^T\boldsymbol{\gamma} + c$, $\boldsymbol{A} \succ \boldsymbol{0}$ with $\boldsymbol{A} \in \mathbb{R}^{L\times L}$, $\boldsymbol{b} \in \mathbb{R}^L$ and $c \in \mathbb{R}$ such that $J(\boldsymbol{\gamma}) \geq 0, \forall \boldsymbol{\gamma}$, then $\mathbb{E}[J(\hat{\boldsymbol{\theta}})] = J(\boldsymbol{\theta}_0) + \mathbb{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\boldsymbol{A}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)]$. The requirement that $J(\boldsymbol{\gamma}) \geq 0$ stems from the fact that J is assumed to be a legitimate end-performance metric. The corresponding lower bound is summarized in the next lemma:

[1] E.g., the linearization of any given end-performance metric in control or signal processing literature.

[2] MSE's and weighted MSE's are such end-performance metrics.