Automatica 47 (2011) 1938–1948

Contents lists available at SciVerse ScienceDirect

Automatica

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

The cost of complexity in system identification: The Output Error case☆

Cristian R. Rojas [a,*], Märta Barenthin [a], James S. Welsh [b], Håkan Hjalmarsson [a]

[a] ACCESS Linnaeus Center, Electrical Engineering, KTH—Royal Institute of Technology, SE-100 44 Stockholm, Sweden

[b] School of Electrical Engineering and Computer Science, The University of Newcastle, NSW 2308, Australia

ARTICLE INFO

Article history:
Received 15 April 2010
Received in revised form
5 November 2010
Accepted 23 March 2011
Available online 20 July 2011

Keywords:
Experiment design
System identification
Prediction error method
LMI optimization
Asymptotic variance

ABSTRACT

In this paper we investigate the cost of complexity, which is defined as the minimum amount of input power required to estimate the frequency response of a given linear time invariant system of order n with a prescribed degree of accuracy. In particular we require that the asymptotic (in the data length) variance is less or equal to γ over a prespecified frequency range $[0, \omega_B]$. The models considered here are Output Error models, with an emphasis on fixed denominator and Laguerre models. Several properties of the cost are derived. For instance, we present an expression which shows how the pole of the Laguerre model affects the cost. These results quantify how the cost of the system identification experiment depends on n and on the model structure. Also, they show the relation between the cost and the amount of information we would like to extract from the system (in terms of ω_B and γ). For simplicity we assume that there is no undermodelling.

0. Notation

The complex conjugate of the complex number z is denoted $\bar{z}$. The complex conjugate transpose for the matrix X is denoted X^*, its transpose is denoted X^T, its adjoint is Adj X, and its column range space is written Range$\{X\}$. The Kronecker delta function is defined by δ_k. The time shift operator is denoted q, *i.e.* $qu(t) = u(t+1)$. $\mathbb{R}$ denotes the field of real numbers. $\mathbb{C}$ denotes the field of complex numbers. The unit circle is defined by $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$. The interior of the unit disc is denoted $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$ and its exterior $\mathbb{E} := \{z \in \mathbb{C} : |z| > 1\}$. ln is considered to be the principal branch of the logarithm, *i.e.* $\mathrm{Im}\{\ln(z)\} \in (-\pi, \pi]$ for all $z \in \mathbb{C} \setminus \{0\}$. Denote by $G(q)$ a transfer function. The Hardy space of analytic functions f on $\mathbb{E}$ taking values on $\mathbb{C}^n$ such that $\lim_{r\to 1_+} \int_{-\pi}^{\pi} \|f(re^{j\omega})\|_2^2 \mathrm{d}\omega < \infty$ is denoted as $\mathcal{H}_2$ (Duren, 1970; Koosis, 1998).

1. Introduction

Real world systems are often of large order. Unfortunately a large model complexity leads to poor model accuracy. For example, consider the well-known approximate variance expression

$$\frac{n}{N}\frac{\Phi_v(\omega)}{\Phi_u(\omega)} \tag{1}$$

for parametric frequency function estimates, presented in Ljung (1985) where Φ_u and Φ_v denote the power spectrum of the input signal and the noise which corrupts the output measurements, respectively. Furthermore, n denotes the model order and N is number of observations of input–output measurements. This formula indicates a 'curse of complexity' since the variance grows linearly with the model order.

This paper is a piece of the puzzle to understanding why system identification, despite this rather pessimistic observation, is widely used in industrial practice as a reliable system modeling tool.

Example 1. Consider a Finite Impulse Response (FIR) system described by

$$y(t) = \sum_{k=0}^{n} \theta_k q^{-k} u(t) + e(t) = G(q, \theta^o)u(t) + e(t),$$

where $e(t)$ is white noise of variance σ_e^2. For a white noise input of variance σ_u^2, it can be established from non-asymptotic (in model order) variance expressions (Ljung, 1999) that the normalized covariance of an efficient estimator of θ is

$$\lim_{N\to\infty} N\, Cov\, \hat{\theta} = \frac{\sigma_e^2}{\sigma_u^2} I,$$

which does not depend on the model complexity. Furthermore, we also obtain consistent estimators of the impulse response

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Martin Enqvist under the direction of Editor Torsten Söderström.

* Corresponding author. Tel.: +46 8 790 7427; fax: +46 8 790 7329.
E-mail addresses: cristian.rojas@ee.kth.se (C.R. Rojas), marta.barenthin@ee.kth.se (M. Barenthin), james.welsh@newcastle.edu.au (J.S. Welsh), hakan.hjalmarsson@ee.kth.se (H. Hjalmarsson).

0005-1098/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2011.06.021

coefficients in case of undermodelling (Hjalmarsson, Mårtensson, & Wahlberg, 2006). However, if we were interested in the frequency response of G, we have by (1) (which gives an exact result in this case) that

$$\lim_{N\to\infty} N\, Var\hat{G}(e^{j\omega}) = n\frac{\sigma_e^2}{\sigma_u^2},$$

which increases with n. Furthermore, it holds that $\hat{G}$ is not consistent in the case of undermodelling.

Now consider a constant input with power σ_u^2. In this case, by a non-asymptotic (in model order) analysis (Hjalmarsson, 2005), we obtain at $\omega = 0$ that

$$\lim_{N\to\infty} N\, Var\hat{G}(e^{j0}) = \frac{\sigma_e^2}{\sigma_u^2},$$

and $\hat{G}(1)$ gives a consistent estimator of the static gain of G. For this input, however, the impulse response coefficients are not even identifiable! □

From the previous example, we see that we cannot expect to be able to accurately identify complex systems with flexible models. However, certain properties can be extracted accurately. Here, the input excitation plays a key role. The input signal can force interesting properties to become visible in the output and it can hide properties of little or no interest. This is in fact what optimal input design is all about. Let us give a formal result presented in Mårtensson and Hjalmarsson (in press) (see also Hjalmarsson et al. (2006)). Let $J(g)$ denote the quantity of interest, where $g = (g_0, g_1, \ldots)$ is the impulse response. Suppose that $\Phi(\omega) = \sum_{k=-\infty}^{\infty} \frac{\partial J(g)}{\partial g_{|k|}} e^{-j\omega k}$ is a spectrum (*i.e.* non-negative for all $\omega \in [-\pi, \pi]$). Taking as input spectrum $\Phi_u(\omega) \sim \Phi(\omega)$ results in an estimate of J that is optimal in the sense that minimal energy is used to achieve the resulting accuracy of J. The variance of the estimate of J is independent of how many non-zero impulse response coefficients the system has (as long as the model structure contains the true system). Furthermore, the estimate of J is consistent for any model order for this input signal. It can be shown that if the property of interest is for example a minimum phase zero ($J = z_o$) or the static gain of the system ($J = \sum_{k=0}^{\infty} g_k$), then the optimal input is independent of the model order! This result is encouraging in the light of the 'curse of complexity' earlier mentioned. This paper is an extension of those previous results to include the entire frequency range in J.

The conceptual problem treated in this paper is the following. Let c be a system complexity measure. Let η be a measure of how much system information is to be extracted from data, for example, system parameters, poles, zeros, the static gain or the frequency response over some bandwidth. The identification purpose is to estimate the system information corresponding to η to within an accuracy γ. The objective is to quantify the experimental cost as a function of c, η and γ (and the noise properties). The cost of complexity is formally denoted $Q(c, \eta, \gamma)$. In this paper we consider Output Error (OE) models with an emphasis on fixed denominator models and Laguerre models. In relation to the authors' previous work, this paper extends the results in Rojas, Barenthin, Welsh, and Hjalmarsson (2008b, 2010), which treat Finite Impulse Response (FIR) models, to more general model structures. Furthermore, in this paper the model complexity is represented by the number of parameters, *i.e.* $c = n$. Let $\hat{\theta}_{N,n}$ denote the prediction error (PE) parameter estimate (Ljung, 1999). The system information of interest is the frequency response estimate $G(e^{j\omega}, \hat{\theta}_{N,n})$ over the frequency range $[0, \omega_B]$. The accuracy is given by

$$\lim_{N\to\infty} N \operatorname{Var}\{G(e^{j\omega}, \hat{\theta}_{N,n})\} \le \frac{1}{\gamma}.$$

The cost of the identification experiment is the input power $1/2\pi \int_{-\pi}^{\pi} \Phi_u(\omega)\mathrm{d}\omega$. The problem is formally posed as

$$\begin{aligned} Q(n, \omega_B, \gamma) &= \min_{\Phi_u(\omega)} \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_u(\omega)\mathrm{d}\omega \\ \text{s.t.} \quad & \Phi_u(\omega) \ge 0, \quad |\omega| \le \pi \\ & \lim_{N\to\infty} N \operatorname{Var}\{G(e^{j\omega}, \hat{\theta}_{N,n})\} \le \frac{1}{\gamma}, \quad |\omega| \le \omega_B. \end{aligned} \tag{2}$$

In fact this problem is an input design problem in the least costly framework (Bombois, Scorletti, Gevers, Van den Hof, & Hildebrand, 2006), which implies that there is a trade-off between the cost and the quality of the system information to be extracted. The contribution of this paper is that we take this one step further by analytically quantifying the cost $Q(n, \omega_B, \gamma)$. For OE models, we examine the monotonicity of the cost with respect to ω_B and n and the cost associated with white input spectra. Also the cost is quantified for the cases $\omega_B = 0$ and $\omega_B = \pi$. For fixed denominator structures a lower bound for the cost is provided. This bound is a function of the (known) system poles. For Laguerre models the cost is compared to FIR models with respect to the pole location in the Laguerre model.

The observation that the input spectrum can be designed to shape the model quality was already made in the 1970s (Goodwin & Payne, 1977; Mehra, 1974) and has been widely applied, see *e.g.* Bombois et al. (2006), Cooley and Lee (2001), Hildebrand and Gevers (2003), Hjalmarsson (2005), Jansson and Hjalmarsson (2005) and Walter and Pronzato (1997). Results on optimal experiment design have also appeared in the statistics literature (Atkinson, Donev, & Tobias, 2007; Fedorov, 1972; Fedorov & Hackl, 1997; Pronzato, 2008; Pukelsheim, 1993). In this paper we exploit the fact that the problem (2) can be recast as a Linear Matrix Inequality (LMI) optimization problem; see Hildebrand and Gevers (2003), Jansson and Hjalmarsson (2005) and Lindqvist (2001). In order to establish properties of the cost of complexity we then use orthonormal basis functions and so-called reproducing kernels (Malmquist, 1926; Ninness & Gustafsson, 1997; Ninness & Hjalmarsson, 2004; Ninness, Hjalmarsson, & Gustafsson, 1998). Such functions have been used for system identification; see *e.g.* Van den Hof, Heuberger, and Bokor (1995) and Wahlberg (1991). In Heuberger, Van den Hof, and Wahlberg (2005), Mårtensson (2007) and Ninness and Hjalmarsson (2004) these functions have been used to quantify the variance of different model properties.

The model structures considered in this paper include those which are nonlinearly parameterized. For those structures, the optimal input associated with the cost Q will typically depend on the parameters of the true system. We are not concerned with this problem here, since our focus is on the theoretical analysis of the optimal cost Q. However, the interested reader may consult references such as (Gerencsér & Hjalmarsson, 2005; Rojas, Welsh, Goodwin, & Feuer, 2007) on robust and adaptive input design, where the problem of the dependence on the true system is explicitly addressed.

It is worth noting that the number of parameters n is not the only possible measure of model complexity. In fact, Zames and collaborators also studied the cost of identifying a linear system (Zames, 1979; Zames & Owen, 1993), but they pursued a different approach, using concepts like ϵ-entropy and ϵ-dimension, from Kolmogorov's theory of complexity. However, in our framework, n seems to be a natural measure of model complexity, as it arises in most of the expressions (and bounds) for the cost of complexity deduced in the paper.

This paper is organized as follows. Section 2 presents the model structure. In Section 3, the problem is mathematically formulated. The main results are presented in Sections 4–6. In Section 4, we