Automatica 49 (2013) 3282–3291

Contents lists available at ScienceDirect

Automatica

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

Input design as a tool to improve the convergence of PEM☆

Diego Eckhard [a,1], Alexandre S. Bazanella [b], Cristian R. Rojas [c], Håkan Hjalmarsson [c]

[a] Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul, Brazil

[b] Department of Electrical Engineering, Universidade Federal do Rio Grande do Sul, Brazil

[c] Automatic Control Lab and ACCESS Linnaeus Centre, School of Electrical Engineering, KTH Royal Institute of Technology, Sweden

ARTICLE INFO

Article history:
Received 13 August 2012
Received in revised form
15 April 2013
Accepted 7 August 2013
Available online 18 September 2013

Keywords:
Identification methods
Experiment design

ABSTRACT

The Prediction Error Method (PEM) is related to an optimization problem built on input/output data collected from the system to be identified. It is often hard to find the global solution of this optimization problem because the corresponding objective function presents local minima and/or the search space is constrained to a nonconvex set. The shape of the cost function, and hence the difficulty in solving the optimization problem, depends directly on the experimental conditions, more specifically on the spectrum of the input/output data collected from the system. Therefore, it seems plausible to improve the convergence to the global minimum by properly choosing the spectrum of the input; in this paper, we address this problem. We present a condition for convergence to the global minimum of the cost function and propose its inclusion in the input design. We present the application of the proposed approach to case studies where the algorithms tend to get trapped in nonglobal minima.

1. Introduction

The Prediction Error Method (PEM) for parameter identification uses input–output data collected from the process to form a cost function. The parameters are then estimated as the solution of the optimization of this cost function. Under mild assumptions, the global minimum of the cost function is a consistent estimate of the model parameters, and the asymptotic variance equals the limit of the Cramér–Rao Bound. Therefore, identification by means of PEM provides a consistent and otherwise statistically appealing estimate of the system parameters and transfer function, provided that the global minimum of the cost function is obtained by the optimization procedure (Ljung, 1999).

One difficulty in applying the PEM method is that in many cases achieving the global minimum may prove difficult (Ljung, 2010; Zou, Tang, & Ding, 2012), for two main reasons: the cost function is usually not convex and the problem is constrained to a nonconvex set — namely, the set of parameters which yields stable predictors. Currently adopted solutions to this problem consist mainly in searching for good initial conditions to initialize the optimization, which is performed with some standard algorithm — steepest descent, Newton–Raphson, Levenberg–Marquardt, and the like. A "good" initial condition is one that is close enough to the global minimum to begin with, that is, closer to it than any local maxima or minima that would prevent convergence to the global minimum if they happened to be in the optimization path. Although there seem to be no firmly established guarantees that these solutions yield the global minimum, they have been successfully applied for many years. They are able to achieve the global minimum of the objective function in most cases, but failure to do so is not such an uncommon occurrence either. As the model order becomes larger, the trend to get trapped in local minima or to "converge" to the boundary of the search space, thus providing a useless model, seems to grow; we provide a couple of such examples in this paper. This problem has received renewed interest in the past few years, as pointed out in Ljung (2010), and different approaches have emerged to cope with it, such as the use of resampling schemes to the input–output data (Garatti & Bitmead, 2010) and the approximation of the process model by (potentially high order) linearly parametrized model structures to obtain convex cost functions (Grossmann, Jones, & Morari, 2009; Hjalmarsson, Welsh, & Rojas, 2012).

In this work, we present a different approach to the convergence problem. We focus on the cost function itself and its shape, trying to avoid the very existence of local minima and maxima. If the cost function has a "good" shape, meaning that it does not have

☆ This work was supported by CAPES and CNPq/Brasil, by the European Research Council under the advanced grant LEARN, contract 267381, and by the Swedish Research Council under contract 621-2009-4017. The material in this paper was partially presented at the 16th IFAC Symposium on System Identification (SYSID 2012), July 11–13, 2012, Brussels, Belgium. This paper was recommended for publication in revised form by Associate Editor Wolfgang Scherrer under the direction of Editor Torsten Söderström.

E-mail addresses: diego@eckhard.com.br (D. Eckhard), bazanela@ece.ufrgs.br (A.S. Bazanella), cristian.rojas@ee.kth.se (C.R. Rojas), hakan.hjalmarsson@ee.kth.se (H. Hjalmarsson).

[1] Tel.: +55 51 9444 6300; fax: +55 51 3308 3293.

0005-1098/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.automatica.2013.08.027

local extrema other than the global minimum or inflection points, then it is easier for any gradient based optimization algorithm to converge to the global minimum, even if initialized far from the global minimum. We analyze what are the conditions under which the cost function has a "good" shape. We show that this property depends on the experimental conditions, more specifically on the spectrum of the input/output data collected from the system. Therefore, it is possible to improve convergence to the global minimum by properly choosing the spectrum of the input signal.

As a next step, we apply this concept to experiment design: we present the use of *input design* as a tool to ensure that the cost function's shape will be amenable to optimization. Typically, input design is formulated as an optimization problem on the input, where several different constraints may be applied. These constraints usually refer to the cost of the identification procedure and/or the quality of the model obtained at the global minimum of the cost function. We propose the addition of a *convergence constraint* to the input design, aiming to obtain an input spectrum such that the cost function will not present local extrema within a given set, which can be made as large as desired. In doing so, we will improve the convergence of iterative algorithms to the global minimum of the PE criterion.

The paper is organized as follows. Section 2 presents basic definitions and the problem formulation. Section 3 presents desired properties of the optimization problem. Conditions about the convergence of the methods are described in Section 4. Input design as a tool to improve the convergence to the global minimum is shown in Section 5. Section 6 presents case studies, and concluding remarks are given in Section 7.

2. Problem formulation

We consider prediction error identification of a linear time-invariant discrete-time single-input–single-output "true system":

$$\mathcal{S}: \quad y(t) = G_0(z)u(t) + H_0(z)e(t) \tag{1}$$

where $G_0(z)$ and $H_0(z)$ are the process transfer functions, $u(t)$ is the input and $e(t)$ is white noise with variance σ_e^2. Both transfer functions are rational and proper; furthermore, $H_0(z)$ is monic, i.e. $H_0(\infty) = 1$. To be precise, we shall define $\mathcal{S} \triangleq [G_0(z)\ \ H_0(z)]$. The signal $u(t)$ is assumed to be quasistationary (Ljung, 1999). We also assume that the data is collected in open loop such that $\bar{E}[u(t)e(s)] = 0\ \forall t, s$ where

$$\bar{E}[f(t)] \triangleq \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} E[f(t)]$$

with $E[\cdot]$ denoting expectation (Ljung, 1999).

In this paper we consider the identification of linear models

$$y(t) = G(z,\theta)u(t) + H(z,\theta)e(t) \tag{2}$$

where $G(z,\theta)$ and $H(z,\theta)$ are rational and proper transfer functions and θ represents the parameters to be identified. The set $\mathcal{M} \triangleq \{[G(z,\theta)\ H(z,\theta)], \forall\theta \in D_\theta\}$ is called the model set. The search space D_θ is usually constrained to be such that the predictors are BIBO-stable $\forall\theta \in D_\theta$.

We assume that the numerator and denominator of the model are affine functions of the unknown parameters, so that the transfer functions $G(z,\theta)$ and $H(z,\theta)$ have the structure

$$G(z,\theta) = \frac{B^T(z)\theta}{1 + F^T(z)\theta}, \qquad H(z,\theta) = \frac{1 + C^T(z)\theta}{1 + D^T(z)\theta}, \tag{3}$$

where the vectors $B(z), C(z), D(z), F(z) \in \mathbb{R}^n$ are composed of fixed transfer functions and $\theta \in \mathbb{R}^n$ corresponds to the unknown parameter vector of the model. This structure is more general than the structures used in Goodwin, Agüero, and Skelton (2003), Zou and Heath (2009) and Zou and Heath (2012) and encompasses all the classical model structures: Box–Jenkins, Output Error (OE), ARMAX, ARX.

We also consider that the system being identified can be exactly described within the model class considered, as formalized by the following assumption.

Assumption 1. $\mathcal{S} \in \mathcal{M}$.

The real system $\mathcal{S}$ belongs to the model set $\mathcal{M}$, i.e. $\exists\ \theta_0 \in D_\theta$ such that

$$G(z,\theta_0) = G_0(z) \quad \text{and} \quad H(z,\theta_0) = H_0(z). \quad \diamond$$

Prediction error identification based on N input–output data consists in finding, among all the models in the pre-specified model set, one that provides the minimum value for the prediction error criterion, that is, one that solves the following optimization

$$\hat{\theta}_N = \arg\min_\theta V_N(\theta) \tag{4}$$

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}[\hat{y}(t,\theta) - y(t)]^2 \tag{5}$$

where the optimal one-step-ahead predictor is given by

$$\hat{y}(t,\theta) = H^{-1}(z,\theta)G(z,\theta)u(t) + (1 - H^{-1}(z,\theta))y(t). \tag{6}$$

PEM has the property (Ljung, 1999) that under mild conditions the parameter estimate $\hat{\theta}_N$ converges w.p.1, for $N \to \infty$, to a set

$$\Theta^* = \{\theta^* \triangleq \arg\min_{\theta\in D_\theta} V(\theta)\}, \tag{7}$$

with

$$V(\theta) \triangleq \bar{E}[y(t) - \hat{y}(t,\theta)]^2. \tag{8}$$

If $\mathcal{S} \in \mathcal{M}$, then $\theta_0 \in \Theta^*$. Moreover, under appropriate conditions the cardinality of the set Θ^* is one, that is, θ_0 is the unique global minimum of $V(\theta)$. Necessary and sufficient conditions for uniqueness of the global minimum are local identifiability of the model structure and local informativity of the experiment, as shown recently in Bazanella, Bombois, and Gevers (2012). Under these conditions, the parameter error converges to a Gaussian random variable:

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{N\to\infty} \mathcal{N}(0, P_{\theta_0}), \tag{9}$$

where

$$P_{\theta_0} = \left(\frac{1}{\sigma_e^2}\bar{E}\left[\left(\frac{\partial\hat{y}(t,\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\right)\left(\frac{\partial\hat{y}(t,\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\right)^T\right]\right)^{-1}.$$

In this paper, we study the properties of the optimization problem defined in (4), and approximate the properties of $V_N(\theta)$ by those of $V(\theta)$, due to the uniform convergence of the former to the latter (Ljung, 1999; Pintelon & Schoukens, 2001; Söderström & Stoica, 1989). This approximation is very useful because $V_N(\theta)$ is stochastic and $V(\theta)$ is a deterministic function. By the application of Parseval's theorem this function can be written as:

$$V(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|H^{-1}(e^{j\omega},\theta)H_0(e^{j\omega})\right|^2\sigma_e^2 + \left|H^{-1}(e^{j\omega},\theta)(G_0(e^{j\omega}) - G(e^{j\omega},\theta))\right|^2\Phi_u(\omega)d\omega$$

where $\Phi_u(\omega)$ is the spectrum of the input signal $u(t)$.

The solutions of the optimization problems (4) and (7) are the points in the search space D_θ which minimize the respective cost functions, and we will henceforth refer to as the *global minimum*; see the following definition.