Automatica 68 (2016) 314–328

Contents lists available at ScienceDirect

Automatica

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

The transient impulse response modeling method for non-parametric system identification☆

Per Hägg[a], Johan Schoukens[b], Michel Gevers[c], Håkan Hjalmarsson[a]

[a] Automatic Control Lab and ACCESS, School of Electrical Engineering, KTH, 100 44 Stockholm, Sweden

[b] Vrije Universiteit Brussel, ELEC Department, Pleinlaan 2, 1050 Brussels, Belgium

[c] ICTEAM, Université Catholique de Louvain, B1348 Louvain la Neuve, Belgium

ARTICLE INFO

Article history:
Received 6 November 2014
Received in revised form
29 August 2015
Accepted 15 January 2016
Available online 14 March 2016

Keywords:
Non-parametric identification
Frequency domain
Identification algorithms

ABSTRACT

A method for the nonparametric estimation of the Frequency Response Function (FRF) was introduced in Hägg et al. (2011) and later called Transient Impulse Response Modeling Method (TRIMM). We present here a slightly improved version of the original method and, more importantly, we thoroughly analyze the method in terms of bias and variance errors. This analysis leads to guidelines for the choice of the design parameters in TRIMM. Our theoretical expressions for the bias and variance errors are validated by simulations which, at the same time, highlight the effect of the design parameters on the performance of the method.

1. Introduction

Frequency Response Function (FRF) estimation is a classical problem in system identification. The identification method may be either parametric, where the model of the system is parameterized with a finite number of parameters, often considerably less than the number of data points, or non-parametric where the number of parameters are as many as the number of data points. In this paper we will focus on the latter. The non-parametric estimates are often used in the initial stage of the identification process to get insight into various system properties, such as system order and noise characteristics, and can guide the user in the model selection in order to proceed with a more accurate parametric estimate. Non-parametric frequency response functions are also useful in their own right and are used intensively in many engineering fields, for example in audio applications, power systems and vibration analysis.

The classical approach is to use spectral analysis, see, e.g., textbooks like (Brillinger, 1981; Ljung, 1999; Pintelon & Schoukens, 2001; Stoica & Moses, 2005). The idea is to smooth the raw Discrete Fourier transform (DFT) estimate using information from adjacent frequencies, or, equivalently, to use weighting of the correlation estimates of different time-lags.

All nonparametric methods suffer from transient (or leakage) errors and noise errors. Transient errors occur when using a finite number of data and a non-periodic input signal. This has for a long time been a major deterrent against the use of nonparametric estimates of the FRF in the presence of non-periodic input signals.

However, by assuming the system to be finite dimensional the leakage error can be analyzed in detail, see McKelvey (2000) and Pintelon, Schoukens, and Vandersteen (1997). This analysis indicates that this error is highly structured with a smooth frequency characteristic. In Schoukens, Vandersteen, Barbé, and Pintelon (2009) this property is explored to develop what is known as the Local Polynomial Method (LPM), an alternative to the classical frequency smoothing. The idea is to approximate the smooth leakage term by a Taylor series expansion and to simultaneously estimate the coefficients of this expansion together with the frequency response at one frequency at a time. The method has been demonstrated to provide superior accuracy, as compared to traditional smoothing algorithms, on a number of problems, see for example Pintelon, Schoukens, Vandersteen, and Barbé (2010a,b). The method has been further developed in Gevers, Pintelon, and Schoukens (2011) and McKelvey and Guérin (2012).

☆ The material in this paper was partially presented at the 50th IEEE Conference on Decision and Control and European Control Conference, December 12–15, 2011, Orlando, FL, USA and at the 16th IFAC Symposium on System Identification, July 11–13, 2012, Brussels, Belgium. This paper was recommended for publication in revised form by Associate Editor Antonio Vicino under the direction of Editor Torsten Söderström.

E-mail addresses: per.hagg@ee.kth.se (P. Hägg), johan.schoukens@vub.ac.be (J. Schoukens), michel.gevers@uclouvain.be (M. Gevers), hakan.hjalmarsson@ee.kth.se (H. Hjalmarsson).

http://dx.doi.org/10.1016/j.automatica.2016.01.062

Inspired by the LPM method, the Transient and Impulse Response Modeling Method (TRIMM) was introduced in Hägg, Hjalmarsson, and Wahlberg (2011). The leakage error, or transient, is modeled with a finite impulse response model. To be able to simultaneously estimate both the FIR parameters and the system frequency response over the grid of DFT frequencies, the DFT of the output measurements have to be recycled and used several times. The main difference compared to LPM is that the transient is globally parameterized as opposed to locally in LPM. The global parameters are then estimated using the whole data record. In LPM only the data points in a local window around each frequency are used to estimate the transient.

Some first attempts to analyze the TRIMM method are given in Gevers, Hägg, Hjalmarsson, Pintelon, and Schoukens (2012) and Hägg and Hjalmarsson (2012). The objective of this paper is to give a more detailed variance and bias analysis of the method to be able to guide the user in the choice of design parameters and experimental settings. As the general problem, with arbitrary system, input and noise sources is hard to analyze, we will in this paper consider a few special cases. Although the analysis is restricted, this gives some insight into the inner workings of the method and we will discuss the expected implications from this analysis to more general cases. For the bias we will mainly study the errors for second order systems with low damping. The motivation for this choice is that most systems can be written as sums of first and second order systems and that the part with lowest damping introduces the largest bias error (Schoukens, Vandersteen, Pintelon, Emedi, & Rolain, 2013). We will also give some results for the case when the system is highly damped.

The main contribution of this paper is first to present the idea behind, and to summarize the previous work related to the TRIMM method. The second contribution is to analyze the bias and variance errors of the estimated frequency response function with TRIMM. This will allow us, in future work, to compare different estimation methods and to give some user guidance on when and how the different methods should be used.

The outline of the rest of the paper is as follows. In Section 2 the frequency domain input–output relation is shown that is then used in Section 3 to derive the TRIMM method. In Section 4, bias and variance expressions for TRIMM are derived. The results and their implications in terms of the design choices of the TRIMM method are discussed in Section 5. The bias and variance expressions are then verified in Monte-Carlo simulations in Section 6 and applied to a vibrating steel beam experimental system in Section 7. Finally, Section 8 concludes the paper.

2. The input–output relation

Consider a linear discrete-time single-input single-output (SISO) system, $G(q)$. The system is excited by an input signal $u(t)$ and the output $y(t)$ is assumed to be disturbed by additive measurement noise $v(t)$. The input–output relation can then be written as

$$y(t) = G(q)u(t) + v(t) \tag{1}$$

where q is the forward shift operator and $G(q)$ is a causal rational function of q.

Equivalently, it can also be represented in state-space form as

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + v(t) \end{aligned} \tag{2}$$

where $x(t)$ is the state vector and $x(0) = x_0$.

Taking the N-point DFT

$$Z(k) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} z(t) e^{-j\omega_k t} \tag{3}$$

of the finite record of measured input and output data $\{u(t)\}$ and $\{y(t)\}$, $t = 0, \ldots, N-1$ gives the following input–output DFT relation (McKelvey, 2000) for $k = 0, \ldots, N-1$

$$Y(k) = G(e^{j\omega_k})U(k) + T(e^{j\omega_k}) + V(k), \tag{4}$$

where $\omega_k \triangleq \frac{2\pi k}{N}$ are the DFT frequencies. The leakage term T is due to the non-zero initial condition and the finite data record length. It is important to understand that (4) is an exact relation between the finite input and output data records (McKelvey, 2000; Pintelon et al., 1997).

To simplify the notation we write the frequency domain expression (4) as

$$Y_k = G_k U_k + T_k + V_k, \quad k = 0, \ldots, N-1$$

where $X_k = X(e^{j\omega_k})$.

It has been shown that the transient term, evaluated at $e^{j\omega_k}$ can be expressed as McKelvey (2000):

$$\begin{aligned} T_k &= \frac{1}{\sqrt{N}} C(I - e^{-j\omega_k}A)^{-1}(x_0 - x_N) \\ &= \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} CA^t (I - A^N)^{-1}(x_0 - x_N) e^{-j\omega_k t} \end{aligned} \tag{5}$$

where x_N is the state at time $t = N$ of the state space realization (2). This special structure of the transient is utilized in TRIMM.

Although the method presented in this paper can be applied to both random and deterministic input signals, we will assume that both the input, $u(t)$, and the disturbing noise, $v(t)$, can be described as filtered zero mean noise with existing moments of any order. The DFT of the input and the noise are then asymptotically ($N \to \infty$) independent over the frequencies, circular complex normally distributed (Pintelon & Schoukens, 2001). Furthermore we assume that the system operates in open loop and thus $u(t)$ is independent of $v(t)$.

3. The TRIMM method

The objective is to estimate the Frequency Response Function $G(e^{j\omega_k})$ for the whole frequency grid, $k = 0, \ldots, N-1$. To perform this estimation we utilize the exact relations (4) and also identify the transient term, T. However, since there are only $N/2$ complex equations, it is impossible to directly estimate the N complex unknown parameters $\{G_k, T_k, k = 0, \ldots, N-1\}$. To generate more equations we approximate G_k and T_k in a local window of size $2L+1$ around each frequency ω_k.

To relate the frequency response at frequency ω_k with the frequency response at the neighboring frequencies ω_{k+r}, $r = -L, \ldots, L$ we write

$$\begin{aligned} Y_{k+r} &= G_{k+r}U_{k+r} + T_{k+r} + V_{k+r} \\ &= G_k U_{k+r} + [G_{k+r} - G_k]\, U_{k+r} + T_{k+r} + V_{k+r}. \end{aligned} \tag{6}$$

Using the definition of G_k we can now express the difference $G_{k+r} - G_k$ in (6) as

$$\begin{aligned} G_{k+r} - G_k &= \sum_{t=1}^{\infty} g(t) \left(e^{-j\omega_{k+r}t} - e^{-j\omega_k t}\right) \\ &= \sum_{t=1}^{N-1} \sum_{p=0}^{\infty} g(t+pN) \left(e^{-j\omega_{k+r}t} - e^{-j\omega_k t}\right) \\ &\triangleq \sum_{t=1}^{N-1} \tilde{g}(t) \left(e^{-j\omega_{k+r}t} - e^{-j\omega_k t}\right) \end{aligned} \tag{7}$$

where $g(t) \triangleq CA^{t-1}B$ is the impulse response of the system (2) and $\tilde{g}(t) = \sum_{p=0}^{\infty} g(t+pN)$.