Automatica 47 (2011) 1646–1655

Contents lists available at ScienceDirect

Automatica

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

Parameter estimation with scarce measurements[☆]

Feng Ding[a,b,*], Guangjun Liu[c], Xiaoping P. Liu[d]

[a] Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, China
[b] Control Science and Engineering Research Center, Jiangnan University, Wuxi 214122, China
[c] Department of Aerospace Engineering, Ryerson University, Toronto, Canada M5B 2K3
[d] Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada K1S 5B6

ARTICLE INFO

Article history:
Received 5 October 2009
Received in revised form
5 December 2010
Accepted 19 February 2011
Available online 11 June 2011

Keywords:
Parameter estimation
Recursive identification
Stochastic gradient
Missing data
Multi-innovation identification

ABSTRACT

In this paper, the problems of parameter estimation are addressed for systems with scarce measurements. A gradient-based algorithm is derived to estimate the parameters of the input–output representation with scarce measurements, and the convergence properties of the parameter estimation and unavailable output estimation are established using the Kronecker lemma and the deterministic version of the martingale convergence theorem. Finally, an example is provided to demonstrate the effectiveness of the proposed algorithm.

1. Introduction

The problems of scarce measurements or missing measurements often arise in control systems like chemical process control and network-based control (e.g., Wang, Ho, Liu, & Liu, 2009 and Wei, Wang, & Shu, 2009). The reconstruction of unavailable measurements and modeling and identification for systems with scarce measurements have received much research attention for decades (Albertos, Sanchis, & Sala, 1999; Sanchis & Albertos, 2002; Sanchis, Peñarrocha, & Albertos, 2007; Wallin, Isaksson, & Noréus, 2001) and many methods for the reconstruction of unavailable (or missing) measurements have been reported. One method is using the interpolation values between two available measurements by linear, parabola, cubic, spline or piecewise constant interpolations. Another way is to reconstruct the missing measurements with a dynamical model. Also, Broersen, de Waele, and Bos (2004) proposed a finite interval likelihood algorithm for AR model spectral estimation using the conditional log likelihoods.

Pintelon and Schoukens (2000) used frequency domain modeling in the reconstruction, which can be also applied to continuous-time models for system identification. They treated the missing data as unknown parameters, which is a promising treatment if only few data are missing.

Missing measurements can also be estimated using the expectation maximization (EM) approach. For example, Isaksson (1993), who was one of the pioneers in the field of engineering, compared several reconstruction methods of missing measurements for ARX models, including Kalman filtering, maximum-likelihood estimation and iterative reconstruction. Yet another approach to reconstructing missing measurements is to use the iterative expectation maximum algorithm. A simplified iteration of data reconstruction and ARX parameter estimation were proposed by Wallin, Isaksson, and Ljung (2000). Gibson and Ninness (2005) explored the maximum likelihood (ML) estimates of state space models from multivariable measurements, and employed the EM algorithm as a means of computing ML estimates; Raghavan, Tangirala, Gopaluni, and Shah (2006) studied the EM-based state space model identification problems with irregular output sampling. However, almost all methods resulted in biased estimates if a large number of data are missing. Ljung pointed out that the EM algorithm could be also interpreted as a method to minimize a prediction error criterion (Ljung, 1999).

☆ This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and in part by the Canada Research Chair Program, and also in part by the National Natural Science Foundation of China (60974043). 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. This research was completed while the first author visited Carleton University (Ottawa, Canada) and Ryerson University (Toronto, Canada) from May 2008 to October 2009.

* Corresponding author at: Control Science and Engineering Research Center, Jiangnan University, Wuxi 214122, China. Tel.: +86 15161698763; fax: +86510 85910633.

E-mail addresses: fding@jiangnan.edu.cn (F. Ding), gjliu@ryerson.ca (G. Liu), xpliu@sce.carleton.ca (X.P. Liu).

0005-1098/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2011.05.007

The issue of missing measurements is also tackled by designing a predictor based on the process model to either infer the lacking measurements (inference systems) or estimate the state and all the required process variables, by using a Kalman filter (Sanchis et al., 2007). For the systems with scarce measurements—a class of missing output systems, Albertos et al. (1999) introduced an output predicition/estimation algorithm for systems where some output data are missing and analyzed its stability for the special case that every Nth sample of the output is observed, and Wallin et al. (2001) showed how to analyze the stability for arbitrary periodical missing data patterns. Sanchis et al. (2007) designed a model-based output predictor for systems with scarce irregular measurements with time-varying delays, taking into account the past measured outputs. Moreover, Albertos et al. (1999) discussed how the Kalman prediction can be used to fill in the output data by assuming that the system parameters are known.

Sanchis and Albertos (2002) discussed the problem of recursive identification under scarce-data operation. The control action is assumed to be updated at a fixed rate, while the output is assumed to be measured synchronously with the input update, but with an irregular availability pattern. Under these conditions, Sanchis and Albertos used the pseudo-linear recursive algorithms for the parameter estimation and carried out the convergence analysis for the case of regular but scarce data availability (Sanchis & Albertos, 2002). Recently, an auxiliary model based least squares algorithm with a forgetting factor was proposed for systems with irregularly missing data (Ding & Ding, 2010).

Dual-rate/multirate (non-uniformly) sampled systems can be viewed as a class of missing-data systems (Ding, Liu, & Yang, 2008; Ding, Shi, Wang, & Ding, 2010; Liu & Lu, 2010; Liu, Xie, & Ding, 2009; Xie, Liu, Yang, & Ding, 2010). In this literature, Ding, Qiu, and Chen (2009) studied reconstruction and identification problems of continuous-time systems from their non-uniformly sampled discrete-time systems; Ding, Liu, and Liu (2010a) presented a partially coupled stochastic gradient identification algorithm for non-uniformly sampled systems.

Unlike the work (Sanchis & Albertos, 2002; Sanchis et al., 2007) which uses the ARX models, this paper uses the output error models and studies the parameter identification problems for the systems with scarce and irregular measurements. The main contributions of this paper lie in that a gradient-based parameter identification method for input–output representations and its convergence properties are explored using only the available inputs and scarce outputs.

In the fields of modeling, identification and prediction for systems with missing observations or irregular sampling, Kim and Stoffer (2008) studied fitting stochastic volatility modeling in the presence of irregular sampling via particle methods and the EM algorithm; Gopaluni (2008) explored the identification problem of nonlinear processes under missing observations based on the particle filter approach and the EM algorithm; Shumway and Stoffer (1982) used the time series approach to smooth and forecast missing observations using the EM algorithm. Also, it has been reported in Shumway and Stoffer (2000) that the EM approach was applied to filtering, smoothing and forecasting of the state space models with missing data modifications.

The rest of this paper is organized as follows. Section 2 introduces the problem formulation of system identification with scarce measurements. Section 3 discusses the gradient-based parameter identification for input–output representations. Section 4 proves the convergence of the gradient-based algorithm in the stochastic framework. Section 5 gives two illustrative examples to show the effectiveness of the proposed algorithms. Finally, we offer some concluding remarks in Section 6.

2. Problem formulation

Consider the stochastic output error system in Fig. 1, where P_d has the following transfer representation (Ljung, 1999):

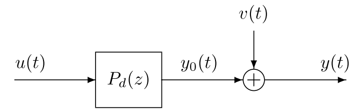

Fig. 1. The stochastic system.

$$P_d(z) = \frac{B(z)}{A(z)}, \tag{1}$$

$u(t) \in \mathbb{R}^1$ and $y(t) \in \mathbb{R}^1$ are the system input and output, respectively, $v(t) \in \mathbb{R}^1$ is a white noise with zero mean and variance σ^2, and $A(z)$ and $B(z)$ are the polynomials in a unit backward shift operator $z^{-1}(z^{-1}u(t) = u(t-1)$ or $zu(t) = u(t+1))$, and defined by

$$A(z) := 1 + a_1z^{-1} + a_2z^{-2} + \cdots + a_nz^{-n},$$
$$B(z) := b_0 + b_1z^{-1} + b_2z^{-2} + b_3z^{-3} + \cdots + b_nz^{-n}.$$

Referring to Fig. 1 and (1), we have

$$y_0(t) = \frac{B(z)}{A(z)}u(t), \tag{2}$$

$$y(t) = y_0(t) + v(t). \tag{3}$$

The inner variables (i.e., the true output or noise-free output) $y_0(t)$ and the noise $v(t)$ are unmeasurable and $y(t)$ is the noisy measurement of $y_0(t)$ corrupted by the noise $v(t)$.

In this paper, we consider such a system with scarce measurements (e.g., Sanchis et al., 2007) that the input $u(t)$ is available at every instant t because the input signals are usually generated by digital computers in practice and are normally available, and only scarce measurement data $y(t_0), y(t_1), y(t_2)$, etc. are available, as shown in Fig. 2, where the integer sequence $\{t_s : s = 0, 1, 2, \ldots\}$ satisfies

$$0 = t_0 < t_1 < t_2 < t_3 < \cdots < t_{s-1} < t_s < \cdots,$$

with $t_s^* := t_{s+1} - t_s \geqslant 1$. Thus $y(t)$ is available only when $t = t_s(s = 0, 1, 2, \ldots)$, or equivalently, the data set $\{y(t_s) : s = 0, 1, 2, \ldots\}$ contains all available outputs, and the unavailable data $y(t_s + 1), y(t_s + 2), \ldots, y(t_{s+1} - 1)$ are all missing for all $s = 0, 1, 2, \ldots$. For instance, for the scarce measurement pattern in Fig. 2, the available measurements are $y(0), y(1), y(3), y(6), y(10), y(15), y(16), y(21), y(28), \ldots,$ namely, $y(t_0), y(t_1), y(t_2), y(t_3), y(t_4), y(t_5), y(t_6), y(t_7), y(t_7), \ldots$, for $t_0 = 0, t_1 = 1, t_2 = 3, t_3 = 6, t_4 = 10, t_5 = 15, t_6 = 16, t_7 = 21, t_8 = 28, \ldots$. This is a general framework in which we assume the patterns with scarce output availability; of course, it includes all output availability as special cases when $t_s^* = 1$ for all s.

Fig. 3 plots the missing output data pattern with *scarce missing data* where +'s stand for missing outputs or bad data (outliers or unbelievable data).

From Figs. 2 and 3, we can see the difference between the scarce measurement pattern and the missing output data pattern. The system with scarce measurements implies that most data are missing and a few data are available over a period of time. Otherwise, the missing-data system implies that most data are available and a small amount of data are missing over a period of time.

It is worth noting that an FIR model may be used as the identification model for such scarce measurements, but the drawback is that the number of parameters is possibly very large if the poles of the system are close to the border of the unit circle.

For such a scarce measurement pattern in Fig. 2, the objective of this paper is to study the gradient-based estimation method to identify the system parameters for input–output representations and its convergence properties, using only the available inputs and scarce outputs $y(t_s)$.