Automatica 60 (2015) 100–114

Contents lists available at ScienceDirect

Automatica

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

Identification of linear continuous-time systems under irregular and random output sampling☆

Biqiang Mu [a], Jin Guo [b], Le Yi Wang [c], George Yin [d], Lijian Xu [e], Wei Xing Zheng [f]

[a] Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

[b] School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

[c] Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI 48202, USA

[d] Department of Mathematics, Wayne State University, Detroit, MI 48202, USA

[e] Department of Electrical and Computer Engineering Technology, Farmingdale State College, State University of New York, Farmingdale, NY 11735, USA

[f] School of Computing, Engineering and Mathematics, University of Western Sydney, Sydney, NSW 2751, Australia

ARTICLE INFO

Article history:
Received 25 August 2014
Received in revised form
9 April 2015
Accepted 21 June 2015

Keywords:
Linear continuous-time system
Irregular and random sampling
Identifiability
Iterative algorithm
Recursive algorithm
Strong consistency
Asymptotical normality

ABSTRACT

This paper considers the problem of identifiability and parameter estimation of single-input–single-output, linear, time-invariant, stable, continuous-time systems under irregular and random sampling schemes. Conditions for system identifiability are established under inputs of exponential polynomial types and a tight bound on sampling density. Identification algorithms of Gauss–Newton iterative types are developed to generate convergent estimates. When the sampled output is corrupted by observation noises, input design, sampling times, and convergent algorithms are intertwined. Persistent excitation (PE) conditions for strongly convergent algorithms are derived. Unlike the traditional identification, the PE conditions under irregular and random sampling involve both sampling times and input values. Under the given PE conditions, iterative and recursive algorithms are developed to estimate the original continuous-time system parameters. The corresponding convergence results are obtained. Several simulation examples are provided to verify the theoretical results.

1. Introduction

System identification for continuous-time systems via sampling is a classical field (Åström & Wittenmark, 1997; Chen & Francis, 1995; Ding, Qiu, & Chen, 2009; Phillips & Nagle, 2007). It is well understood that to identify a time-invariant continuous-time system, one may derive its time-invariant discrete-time sampled system with periodic sampling and the zero-order hold; and hence identification of the original continuous-time system is converted to that of its sampled system (Åström & Wittenmark, 1997; Garnier & Wang, 2008; Ljung, 1999; Marelli & Fu, 2010). A sufficient condition to guarantee the one-to-one mapping from the coefficients of the sampled system to the original system is that the sampling period is less than an upper bound related to the imaginary parts of the poles (Ding et al., 2009). This equivalence implies that the existing algorithms for discrete-time systems suffice for identification of the original continuous-time system. Furthermore, it was shown in Ding et al. (2009) that multi-rate sampling schemes can be used to create such a one-to-one mapping when the sampling rate is slower than this bound. Under such a multi-rate sampling system, the sampled system of a linear time-invariant system remains linear and time invariant with a higher order.

In practical systems, especially networked systems, periodic sampling is no longer valid. Examples are abundant, such as communication channels with packet loss and unpredictable round-trip times. Irregular sampling time sequences may be generated passively due to event-triggered sampling (Åström & Bernhardsson, 1999), low-resolution signal quantization (Wang, Yin, Zhang, & Zhao, 2010), activities by input control or threshold adaptation

☆ This work was supported in part by the National Key Basic Research Program of China (973 program) under Grant No. 2014CB845301 and the NSF of China under Grants Nos. 61273193, 61120106011, 61134013 and 61403027, in part by the Army Research Office under Grant W911NF-15-1-0218, and in part by the Australian Research Council under Grant DP120104986. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Juan I. Yuz under the direction of Editor Torsten Söderström.

E-mail addresses: bqmu@amss.ac.cn (B. Mu), guojin@amss.ac.cn (J. Guo), lywang@wayne.edu (L.Y. Wang), gyin@math.wayne.edu (G. Yin), dy0747@wayne.edu (L. Xu), w.zheng@uws.edu.au (W.X. Zheng).

http://dx.doi.org/10.1016/j.automatica.2015.07.009
0005-1098/

under binary-valued sensors (Wang, Li, Yin, Guo, & Xu, 2011), or PWM (Pulse Width Modulation)-based sampling (Wang, Feng, & Yin, 2013). Under irregular or random sampling, the sampled system of a linear time-invariant system becomes time-varying, for which system conversion is complicated and computational complexity is much higher. When sampling is slower, irregular, or random, system identification formulation, identifiability, algorithms, accuracy, and convergence will be fundamentally impacted. This paper will explore related issues in this paradigm.

In Johansson (2010), the original differential equation is first converted to an algebra equation with respect to time by using filtered input and output signals. Then the parameters of the algebra equation are estimated at the irregular sampling points, and the original system parameters are recovered by a one-to-one mapping from the estimated parameters. One possible methodology is to identify the original system parameters without conversion to its sampled system (Gillberg & Ljung, 2010; Larsson & Söderström, 2002; Marelli & Fu, 2010; Vajda, Valko, & Godfrey, 1987). The parameters are directly identified by using a continuous-time frequency domain identification method in Gillberg and Ljung (2010). Similar to its discrete-time counterpart, the continuous-time system can also be expressed by a linear regression equation in a differential operator form (Larsson & Söderström, 2002), in which the regressor involves input and output signals and their derivatives. Since the derivatives are unavailable under sampled data, higher-order derivatives of the input and output signals are approximated by their related differences (Larsson & Söderström, 2002), introducing errors as a consequence. The resulting discrete-time system is then identified by batch or recursive algorithms (Ljung, 1999; Ljung & Vicino, 2005). To reduce approximation errors, fast sampling is required. An instrumental variable approach is used to enhance estimation accuracy for continuous-time autoregressive processes in Mossberg (2008), which demonstrates improved computational efficiency in comparison to the least squares approach (Larsson, Mossberg, & Soderstrom, 2007).

Synchronization between the input sampling and output sampling is also a significant factor. Typical schemes for the indirect method assume that the input and output are sampled at the same sampling points (Larsson & Söderström, 2002; Yuz, Alfaro, Agüero, & Goodwin, 2011). The estimation method in Yuz et al. (2011) represents the original continuous-time state space model by an incremental approximation under nonuniform but fast sampling and employs the maximum likelihood approach. Zhu, Telkamp, Wang, and Fu (2009) propose a two-time scale sampling scheme: fast uniform input sampling, but slow and irregular output sampling, with assumption that the output sampling time is a multiple of the input sampling time. Under an output error structure, the system parameters are then estimated by minimizing a suitable loss function. In contrast, in Gillberg and Ljung (2010) the input is a uniformly spaced piece-wise constant function (zero-order hold), while the output is sampled irregularly. The main technique is to use B-spline approximation to achieve uniformly distributed knots from the non-uniformly sampled output. The method in Gillberg and Ljung (2010) is restricted to the noise-free sampled output and its estimation accuracy enhancement requires fast sampling. Despite extensive research effort in this area, some fundamental questions remain un-answered: (1) How fast and under what types of sampling schemes, is the continuous-time system identifiable? (2) What types of inputs will imply system identifiability? (3) What modifications must be made to identification algorithms? (4) To achieve convergence, how should the input be designed?

This paper investigates these questions from a new angle. Instead of focusing on parameter mappings between the continuous-time system and its sampled system, we view irregularly or randomly sampled values as the available information set and study identifiability, identification algorithms, and input design directly on the original parameters. The main contributions of this paper are in the following aspects. (1) We show that under any inputs of exponential polynomial types, the continuous-time system is identifiable if the sampling points are sufficiently dense in a given time interval. The bound on the density of the sampling points is tight, revealing an interesting connection, in terms of identification information complexity, to Shannon's sampling theorem for signal reconstruction (Proakis & Manolakis, 2007) and our recent results on state estimation (Wang et al., 2011). Note that the input used in this paper is continuous, while the existing literature (Ding et al., 2009; Gillberg & Ljung, 2010) commonly uses piecewise constant inputs by zero-order hold. (2) Robustness of system identifiability under a given sampling density is established. (3) Under noise-free observations, a convergent iterative algorithm is introduced, which is valid for any input signals satisfying certain gradient conditions. (4) Under noisy observations, suitable identification algorithms are proposed, which are shown to converge strongly and carry properties of the CLT (central limit theorem) types if certain ergodicity conditions are satisfied. (5) Persistent excitation (PE) conditions are derived that ensure convergence of the developed algorithms. Departing from the traditional PE conditions that rely only on input values, it is shown that under irregular or random sampling, both sampling time sequences and the input values impact on convergence. These results provide guidance for input design in identification experiments. (6) Consistency of the algorithms is proved without requiring fast sampling.

The rest of the paper is arranged as follows. The system setting and several key properties are presented in Section 2. System identifiability is investigated in Section 3. The parameter estimation algorithms and their convergence properties under noise-free observation are discussed in Section 4. When observations are noise corrupted, identification algorithms are significantly different from noise-free cases. Two kinds of estimation algorithms (iterative algorithms and recursive algorithms) are introduced and their convergence conditions are established in Section 5. Section 6 is focused on input design problems. The related persistent excitation conditions are obtained. In Section 7 some numerical examples are given to verify the effectiveness of the proposed algorithms of this paper. Section 8 concludes the paper with some further remarks. The main proofs of the assertions in the paper are placed in the Appendix.

2. Preliminaries

This section describes the system setting and establishes several important properties to be used in the subsequent sections.

2.1. Systems

We are concerned with identification of a single-input–single-output, linear, time-invariant, stable, finite dimensional system in the continuous-time domain, represented by a strictly proper transfer function

$$G(s) = \frac{b_1 s^{n-1} + \cdots + b_{n-1}s + b_n}{s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n} \triangleq \frac{b(s)}{a(s)}, \tag{1}$$

where $a(s)$ is stable, i.e., all the roots of $a(s)$ lie on the open left-half complex plane; $a(s)$ and $b(s)$ are coprime, i.e., they do not have common roots. The impulse response of $G(s)$ is denoted by $g(t) = \mathcal{L}^{-1}(G(s))$, where $\mathcal{L}^{-1}$ is the inverse Laplace transform. Let the system parameters be expressed as $\theta = [a_1, \ldots, a_n, b_1, \ldots, b_n]'$. We use $G(s, \theta)$ and $g(t, \theta)$ to indicate their dependence on the parameters. $\mathbb{R}$ and $\mathbb{C}$ are the fields of real and complex numbers, respectively.

Under the zero initial condition, the input–output relationship of the system (1) is

$$y(t) = g(t,\theta) \star u(t) = \int_0^t g(t-\tau,\theta)u(\tau)d\tau \triangleq f(t,\theta), \tag{2}$$

where the symbol $\star$ denotes the continuous-time convolution. In this paper, $u(t)$ is uniformly bounded by $\|u\|_\infty \leq u_{\max} < \infty$, where $\|\cdot\|_\infty$ is the standard L^∞ norm. In the frequency domain, the relationship (2) has the expression

$$Y(s) = G(s,\theta)U(s), \tag{3}$$

where $U(s)$ and $Y(s)$ are the Laplace transforms of the input and output, respectively. In this paper, $u(t)$ is selected in identification experimental design and always assumed to be known. The output $y(t)$ is sampled at the nonuniform sampling times $t_1, t_2, \ldots$. System identification aims to identify θ from the data set $(t_1, y(t_1)), (t_2, y(t_2)), \ldots$.

For convenience of statement, we define the set $\mathcal{A}$ on θ: (1) The elements in $\mathcal{A}$ consist of column vectors $\theta \in \mathbb{R}^{2n}$; (2) For any $\theta = [a_1, \ldots, a_n, b_1, \ldots, b_n]' \in \mathcal{A}$, the polynomials $a(s) = s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n$ and $b(s) = b_1 s^{n-1} + \cdots + b_{n-1}s + b_n$ generated by θ are coprime. Since the roots of a polynomial depend continuously on its coefficients, $\mathcal{A}$ is an open set.

Since $u(t)$ is uniformly bounded, the order of partial derivatives (with respect to θ) and the integral in the following equation is exchangeable if $u(t)$ is continuous almost everywhere, namely,

$$\begin{aligned} h(t,\theta) &\triangleq \frac{\partial f(t,\theta)}{\partial\theta} = \frac{\partial \int_0^t g(t-\tau,\theta)u(\tau)d\tau}{\partial\theta} \\ &= \int_0^t \frac{\partial g(t-\tau,\theta)}{\partial\theta}u(\tau)d\tau = \frac{\partial g(t,\theta)}{\partial\theta} \star u(t). \end{aligned} \tag{4}$$

2.2. Basic properties

For a clear presentation, most proofs are postponed to Appendix.

Lemma 1. *If $u(t) \not\equiv 0$, then the elements of the gradient vector $h(t,\theta)$ of $f(t,\theta)$ are linearly independent.*

Consider an exponential polynomial

$$w(t) \triangleq \sum_{i=1}^{l_w}\sum_{j=1}^{n_i} w_{i,j}\frac{t^{j-1}}{(j-1)!}\exp(\lambda_i t), \tag{5}$$

where $\{\lambda_i \in \mathbb{C}, i = 1, \ldots, l_w\}$ are different from each other, and $\sum_{i=1}^{l_w} n_i = n_w$. The following lemma is fundamental.

Lemma 2 (*Berenstein & Gay, 1995*, Corollary 3.2.45)**.** *Let $T > 0$. The number N_T of zeros in $[0 \quad T]$ of a nontrivial exponential polynomial $w(t)$ defined in (5) is bounded by $\mu_T = (n_w - 1) + \frac{\delta T}{2\pi}$, where $\delta \triangleq \max_{1\leq i,j\leq l_w}\{|\Im(\lambda_i - \lambda_j)|\}$ and the symbol $\Im(c)$ is the imaginary part of a complex number $c \in \mathbb{C}$.*

3. System identifiability

Identifiability concerns the fundamental question of inverse mapping: Under noise-free but sampled output observations, can the system parameters be uniquely determined? Answers to this question depend on model structures, input signals, sampling types, and sample sizes. This paper characterizes sufficient conditions for system identifiability.

3.1. Identifiability

The simplest input is perhaps the step input $U(s) = u_0/s$. The following results allow more general inputs of rational types. The resulting identifiability condition implies that the system is identifiable as long as the number of the sampling points is greater than some positive integer in a given interval, regardless how the sampling times are generated. Consequently, the results are applicable to different sampling schemes, such as the traditional periodic sampling (Åström & Wittenmark, 1997), nonuniform sampling (Ding et al., 2009), event-triggered sampling (Åström & Bernhardsson, 1999; Persson & Gustafsson, 2001), or the PWM-based sampling introduced in Wang et al. (2013).

Assumption 1. $U(s)$ is a nonzero proper rational coprime function, i.e., $U(s) = d(s)/c(s)$, where $c(s)$ and $d(s)$ are coprime polynomials, $\deg d(s) \leq \deg c(s) = m$. This implies that $u(t)$ is an exponential polynomial and hence the output $f(t,\theta)$ is also an exponential polynomial.

Denote $\mu_T^* = (2n+m-1) + \frac{\delta^* T}{2\pi}$, $\delta^* \triangleq \max_{1\leq i,j\leq 2n+m}\{|\Im(\lambda_i - \lambda_j)|\}$ where $\{\lambda_i, i = 1, \ldots, 2n+m\}$ are the roots of $a^2(s)c(s)$.

Lemma 3. *Under Assumption 1 and for $\theta \in \mathcal{A}$, the $N \times 2n$ Jacobian matrix of $f(t,\theta)$ at the sampling points $0 < t_1 < t_2 < \cdots < t_N \leq T$*

$$J_N(\theta) = \left[\frac{\partial f(t_1,\theta)}{\partial\theta}, \frac{\partial f(t_2,\theta)}{\partial\theta}, \ldots, \frac{\partial f(t_N,\theta)}{\partial\theta}\right]'$$

has full column rank if $N > \mu_T^*$.

Theorem 1. *Under Assumption 1 and the sampling points $0 < t_1 < t_2 < \cdots < t_N \leq T$, the true system parameter $\theta^* \in \mathcal{A}$ can be uniquely determined by the sampling data $\{y(t_1), \ldots, y(t_N)\}$ whenever the number of the sampling points $N > \mu_T^*$.*

Proof. By (2), the sampled outputs $\{y(t_i), 1 \leq i \leq N\}$ satisfy the following set of equations:

$$y(t_i) = f(t_i, \theta^*), \quad i = 1, \ldots, N \tag{6}$$

which is a vector-valued function F_N from $\mathcal{A}$ onto $\mathbb{R}^N$. Under the hypothesis, the Jacobian matrix $J_N(\theta^*)$ of the set of Eqs. (6) has full column rank by Lemma 3. Hence the system parameter θ^* can be uniquely determined by $(y(t_1), \ldots, y(t_N))$ by the inverse function theorem (Rudin, 1973, p. 252). □

Remark 1. Theorem 1 does not require $a(s)$ to be stable. As a result, unstable systems can also be uniquely determined by a finite number of noise-free sampled outputs with sufficient data density. This will be verified by a numerical example in Section 7. The bound μ_T^* depends on both the system dynamics and the order of the input. This is understandable since sampling is on the output. A "low-frequency" system under a "high-frequency" input will still generate a high-frequency output. We shall emphasize that the exponential polynomials starting at $t = 0$ always have infinite bandwidths, and as such Shannon's sampling theorem cannot be applied on the output signal. It can also be verified that the bound μ_T^* is tight in the sense that for large interval T, there exist systems and inputs for which if $N < \mu_T^*$, then the system is not identifiable. For some counter examples for a related but different problem, see Wang et al. (2011).

3.2. Identifiability robustness with respect to sampling size

Theorem 1 gives a required sampling size N for identifiability for θ^*. Identifiability robustness pertains to whether the size N is also sufficient to identify systems in a neighborhood of θ^*. This is important since in identification experimental design, N needs to be pre-designed before θ^* can be identified. We use two useful tools in this study. The issue here is related to global invertibility problems. There are two kinds of conditions on the global invertibility for vector-valued functions.

The first kind of conditions (Sanderberg, 1980; Wu & Desoer, 1972) can be stated as follows. Let $\psi : \mathbb{R}^l \to \mathbb{R}^l$ be a C^1 map. Then ψ is a homeomorphism of $\mathbb{R}^l$ onto $\mathbb{R}^l$, i.e., ψ is a continuous bijective map and ψ^{-1} is also continuous, if and only if ψ satisfies: (1) The mapping ψ is a local homeomorphism, i.e., whenever $\bar{x} \in \overline{X}$ and $\bar{y} \in \overline{Y}$ satisfy $\psi(\bar{x}) = \bar{y}$, there exist open neighborhoods $\overline{U}$ of $\bar{x}$ and $\overline{V}$ of $\bar{y}$ such that ψ restricted to $\overline{U}$ is a homeomorphism of $\overline{U}$ onto $\overline{V}$; (2) ψ is a proper mapping, i.e., the inverse image $\psi^{-1}(\mathfrak{M})$ of any compact set $\mathfrak{M} \in \mathbb{R}^l$ is compact. The second kind of conditions (Mas-Colell, 1979) involves matrix properties. Let $\psi : \mathfrak{M} \to \mathbb{R}^l$ be a C^1 map, where $\mathfrak{M} \subset \mathbb{R}^l$ is a compact and convex set. (1) Let $\mathfrak{M}$ be a rectangle. If for every $x \in \mathfrak{M}$, the Jacobian matrix $J(x)$ of ψ at the point x is a P matrix (i.e., every principal minor of $J(x)$ has positive sign), then ψ is a homeomorphism. (2) If for every $x \in \mathfrak{M}$, the Jacobian matrix $J(x)$ of ψ at the point x is positive quasi-definite (i.e., $v'J(x)v > 0$ for all $x \in \mathbb{R}^l$, $v \neq 0$), then ψ is a homeomorphism. However, both kinds of conditions can only deal with the mappings from $\mathbb{R}^l$ or a compact and convex subset of $\mathbb{R}^l$ onto $\mathbb{R}^l$. Our results on identifiability robustness are related to these conditions, but are different in subtle and important ways.

Let θ^* be any vector in $\mathcal{A}$. Then $J_N(\theta^*)$ has full column rank under the conditions of Theorem 1. It follows that there exists a nonsingular $\mathbb{R}^{2n\times 2n}$ submatrix $J_{i_1,\ldots,i_{2n}}(\theta^*)$ of $J_N(\theta^*)$ composed of the $i_1, \ldots, i_{2n}$ rows of $J_N(\theta^*)$. Since $J_{i_1,\ldots,i_{2n}}(\theta)$ is continuous over θ, there is a compact and convex set $K \subset \mathcal{A}$ containing θ^*. If $F_{2n}(\theta) = [f(t_{i_1}, \theta), \ldots, f(t_{i_{2n}}, \theta)]'$ is still a convex set, then F_{2n} is injective on K by the same derivations as those used in Wu and Desoer (1972).

Theorem 2. *Let the conditions of Theorem 1 hold and $\mathfrak{M}$ be a convex subset of $\mathcal{A}$. The mapping F_N on the domain $\mathfrak{M}$ defined in (6) is injective if and only if for any distinct points x_1 and x_2 in $\mathfrak{M}$, the integral along the line segment joining x_1 to x_2 satisfies $\int_{x_1}^{x_2} J_N(\bar{\xi})\mathrm{d}\bar{\xi} \neq 0$.*

Proof. *Sufficiency*: Suppose that F_N is not injective. Then there exist two distinct points x_1 and x_2 such that $F_N(x_1) = F_N(x_2)$. Since the map F_N is differentiable on the convex domain $\mathfrak{M}$, we have

$$F_N(x_2) - F_N(x_1) = \int_0^1 J_N\big(x_1 + t(x_2 - x_1)\big)(x_2 - x_1)\mathrm{d}t = \int_{x_1}^{x_2} J_N(\bar{\xi})\mathrm{d}\bar{\xi} = 0,$$

which contradicts the assumption. Hence, F_N on the domain $\mathfrak{M}$ is injective.

Necessity: If there exist two distinct points x_1 and x_2 such that $\int_{x_1}^{x_2} J_N(\bar{\xi})d\bar{\xi} = 0$, then

$$F_N(x_2) - F_N(x_1) = \int_0^1 J_N\big(x_1 + t(x_2 - x_1)\big)(x_2 - x_1)\mathrm{d}t = \int_{x_1}^{x_2} J_N(\bar{\xi})\mathrm{d}\bar{\xi} = 0,$$

which means that $F_N(x_1) = F_N(x_2)$. It follows that F_N is not injective on the domain $\mathfrak{M}$, which violates the condition that F_N is injective on the domain $\mathfrak{M}$. As a result, $\int_{x_1}^{x_2} J_N(\bar{\xi})d\bar{\xi} \neq 0$ for any two different points x_1 and x_2 in $\mathfrak{M}$. Therefore, the proof is complete. □

3.3. Examples

The following two examples illustrate several key aspects of the identifiability conditions in Theorem 1.

Example 1. This example will reveal that the Jacobian matrix $J_N(\theta^*)$ is not full column rank if the denominator and the numerator of $G(s)$ have a common factor. Suppose that $G(s) = \frac{b_1 s + b_2}{s^2 + a_1 s + a_2}$ and the denominator of $G(s)$ has two distinct real factors $s + \lambda_1$ and $s + \lambda_2$, where $\lambda_1 > 0$, $\lambda_2 > 0$, and $\lambda_1 \neq \lambda_2$. Then $G(s)$ has the following factorization:

$$G(s) = \frac{c_1}{s + \lambda_1} + \frac{c_2}{s + \lambda_2},$$

where $c_1 = \frac{b_2 - b_1\lambda_1}{\lambda_2 - \lambda_1}$ and $c_2 = \frac{b_2 - b_1\lambda_2}{\lambda_1 - \lambda_2}$. Suppose $u(t) = 1, t \geq 0$. Then $U(s) = 1/s$. Hence

$$Y(s) = \frac{c_1}{s(s + \lambda_1)} + \frac{c_2}{s(s + \lambda_2)} = d_1\Big(\frac{1}{s} - \frac{1}{s + \lambda_1}\Big) + d_2\Big(\frac{1}{s} - \frac{1}{s + \lambda_2}\Big),$$

where $d_1 = c_1/\lambda_1$ and $d_2 = c_2/\lambda_2$, and the output in the time domain is

$$y(t) = d_1(1 - \exp(-\lambda_1 t)) + d_2(1 - \exp(-\lambda_2 t)).$$

It is clear that identifying the original parameters $\{a_1, a_2, b_1, b_2\}$ is equivalent to identifying the new parameters $\{\lambda_1, \lambda_2, d_1, d_2\}$. Set $\theta = [\lambda_1, \lambda_2, d_1, d_2]'$. Thus, we have

$$h(t, \theta) = [d_1 t \exp(-\lambda_1 t), d_2 t \exp(-\lambda_2 t), 1 - \exp(-\lambda_1 t), 1 - \exp(-\lambda_2 t)]'. \quad (7)$$

If $G(s)$ is coprime, then $d_1 \neq 0, d_2 \neq 0$ and $\lambda_1 \neq 0, \lambda_2 \neq 0$. For any vector γ satisfying $\gamma' h(t, \theta) = 0$, we have $\gamma = 0$ as the unique solution, which indicates that the elements of $h(t, \theta)$ in (7) are linearly independent. If $G(s)$ is not coprime, then either $c_1 = 0$ or $c_2 = 0$. Without loss of generality, assume that $c_1 = 0$. It follows that $d_1 = 0$. This implies that

$$h(t, \theta) = [0, d_2\lambda_2 t \exp(-\lambda_2 t), 1 - \exp(-\lambda_1 t), 1 - \exp(-\lambda_2 t)]'. \quad (8)$$

It is apparent that the elements of $h(t, \theta)$ in (8) are linearly dependent since $[1, 0, 0, 0]h(t, \theta) = 0$ if $G(s)$ is not coprime.

Example 2. This example demonstrates that we will not lose the identifiability even if a zero of $U(s)$ cancels a pole of the plant $G(s)$. Assume that $G(s) = \frac{b}{s+a}$ and $U(s) = \frac{s+2}{s+1}$, where $a > 0$ and $a \neq 1$. We want to investigate whether the identifiability for a and b will be compromised if $a = 2$. The output is given by

$$Y(s) = \frac{b(s + 2)}{(s + a)(s + 1)} = \frac{c_1}{s + a} + \frac{c_2}{s + 1},$$

where $c_1 = \frac{b(2-a)}{1-a}$ and $c_2 = -\frac{b}{1-a}$. It follows that

$$y(t) = c_1 \exp(-at) + c_2 \exp(-t).$$

Set $\theta = [a, b]'$. Then we have

$$h(t, \theta) = \Big[\frac{\partial c_1}{\partial a}\exp(-at) - c_1 t \exp(-at) + \frac{\partial c_2}{\partial a}\exp(-t), \frac{2 - a}{1 - a}\exp(-at) - \frac{1}{1 - a}\exp(-t)\Big]'. \quad (9)$$

If $a \neq 2$, then $h(t,\theta)$ in (9) is linearly independent. If $a = 2$, then $c_1 = 0$, but $\frac{\partial c_1}{\partial a} \neq 0$ and $\frac{\partial c_2}{\partial a} \neq 0$. This indicates that $h(t,\theta)$ in (9) is still linearly independent even if $a = 2$. In other words, if we use the same Jacobian matrix in our algorithms, it will still give local convergence for the parameter estimation of the plant.

4. Identification algorithms and convergence properties under noise-free observations

If observations are noise-free, then to obtain the true parameters $\theta^* \in \mathcal{A}$ it suffices to solve the nonlinear equation set (6) by Theorem 1 when the number of the sampling points $N > \mu_T^*$ in $[0 \quad T]$.Since the true parameter θ^* is unique in a small neighborhood, solving (6) is equivalent to solving the following optimization problem.

Define the objective function

$$S(\theta) = \sum_{i=1}^{N} \big(y(t_i) - f(t_i,\theta)\big)^2. \tag{10}$$

By Theorem 1, we have $S(\theta^*) = 0$ and θ^* is the unique argument which takes the minimum value of the optimization (10) in a small neighborhood of θ^*.

To find θ^*, the Gauss–Newton algorithm is employed. Choose arbitrarily an initial value θ_0 which is assumed to be in a small neighborhood of θ^*. The iterative sequence $\{\theta_k\}$ is updated by the following algorithm

$$\theta_{k+1} = \theta_k + (J_N'(\theta_k)J_N(\theta_k))^{-1}J_N'(\theta_k)r(\theta_k), \tag{11}$$

where

$$J_N(\theta_k) = \Big[\frac{\partial f(t_1,\theta_k)}{\partial\theta}, \frac{\partial f(t_2,\theta_k)}{\partial\theta}, \ldots, \frac{\partial f(t_N,\theta_k)}{\partial\theta}\Big]',$$
$$r(\theta_k) = [y(t_1) - f(t_1,\theta_k), \ldots, y(t_N) - f(t_N,\theta_k)]'.$$

Remark 2. Under Assumption 1, the Laplace transform $Y(s)$ of the output $y(t)$ is a rational polynomial: $Y(s) = b(s)d(s)/(a(s)c(s))$, and hence the Laplace transform $\mathcal{L}(h(t,\theta))$ of the gradient vector of $y(t)$ is still a rational polynomial vector

$$\frac{\partial Y(s,\theta)}{\partial\theta} = \Big[-\frac{b(s)d(s)[s^{n-1},\ldots,s,1]}{a^2(s)c(s)}, \frac{d(s)[s^{n-1},\ldots,s,1]}{a(s)c(s)}\Big]'.$$

As a result, taking inverse Laplace transform on both sides gives

$$h(t,\theta) = \mathcal{L}^{-1}\Big(\Big[-\frac{b(s)d(s)[s^{n-1},\ldots,s,1]}{a^2(s)c(s)}, \frac{d(s)[s^{n-1},\ldots,s,1]}{a(s)c(s)}\Big]'\Big).$$

Since the inverse Laplace transform of a rational polynomial is an exponential polynomial, each component of $h(t,\theta)$ is still an exponential polynomial, which is uniquely determined by a finite number of coefficients of the polynomials $a(s), b(s), c(s), d(s)$. For example, let $p(s)$ be a proper rational polynomial and its factorization be $p(s) = \sum_{l=1}^{\bar{p}}\sum_{i=1}^{n_l}\frac{v_{li}}{(s-\lambda_l)^i}$. Thus the inverse Laplace transform of $p(s)$ is $\mathcal{L}^{-1}(p(s)) = \sum_{l=1}^{\bar{p}}\sum_{i=1}^{n_l} v_{li}\frac{t^{i-1}}{(i-1)!}\exp(\lambda_l t)$. This is an accurate expression over the variable t. As a result, the Jacobian matrix given in the algorithm (11) can be accurately calculated for any $\theta \in \mathcal{A}$ if the input is available and the computation process will not introduce systematic errors.

To prove the convergence of the algorithm (11), the following lemma is needed, which is an application of the contraction mapping principle (Granas & Dugundji, 2003; Istratescu, 1981).

Lemma 4. *Suppose that $A, B \in \mathbb{R}^{l\times l}$, A is nonsingular, and $\|A^{-1}\| \leq \varsigma$. If $\|A - B\| \leq \bar{\nu}$ and $\varsigma\bar{\nu} < 1$, then B is nonsingular and $\|B^{-1}\| \leq \frac{\varsigma}{1-\varsigma\bar{\nu}}$.*

Theorem 3. *Under the conditions of Theorem 1, there exists $\epsilon > 0$ such that for any initial value θ_0 in the neighborhood $O(\theta^*, \epsilon)$ of θ^* with radius ϵ, the iterative sequence $\{\theta_k\}$ defined in (11) converges to the true parameter θ^* and has the following convergence rate*

$$\|\theta_{k+1} - \theta^*\| \leq \frac{1}{4}\|\theta_k - \theta^*\|, \tag{12}$$

$$\|\theta_{k+1} - \theta^*\| \leq \varrho\|\theta_k - \theta^*\|^2, \tag{13}$$

for some $\varrho > 0$.

5. Identification algorithms and convergence properties under noisy observations

Under noisy observations, this section uses nonlinear least squares (Jennrich, 1969) and stochastic approximation methods (Kushner & Yin, 2003) to design identification algorithms for estimating the unknown parameters. Iterative and recursive algorithms will be proposed, respectively, and their strong convergence will be established.

Suppose that the sampled output at the sampling time t_k is corrupted by noise

$$y(t_k) = f(t_k,\theta^*) + e_k,$$

where θ^* is the true parameter vector and e_k is the noise. In the case of noisy observations, the plant is naturally required to be stable and it is reasonable to consider asymptotic convergence properties of parameter estimation algorithms as the number of the sampling points grows. The following condition on e_k is assumed throughout the rest of the paper.

Assumption 2. The noise $\{e_k\}$ is a sequence of zero mean i.i.d. random variables with finite variance σ^2.

5.1. Nonlinear least squares estimators and convergence properties

Define the sample objective function

$$Q_N(\theta) = \frac{1}{N}\sum_{k=1}^{N}(y(t_k) - f(t_k,\theta))^2. \tag{14}$$

The vector $\widehat{\theta}_N$ which minimizes (14) on a compact subset Θ of $\mathbb{R}^{2n}$ containing θ^* is called the nonlinear least-squares (NLS) estimate for θ^* based on the sampled outputs $\{y(t_1),\ldots,y(t_N)\}$.

We impose the following conditions as assumptions for now. They will be verified by input design in Section 6.

Assumption 3. (i) The limit $\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}f^2(t_k,\theta)$ exists for any $\theta \in \Theta$ and $Q(\theta) = \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\big(f(t_k,\theta) - f(t_k,\theta^*)\big)^2$ has a unique minimum at $\theta = \theta^*$.

(ii) The gradient vector $h(t_k,\theta)$ and Hessian matrix $\chi(t_k,\theta)$ of $f(t_k,\theta)$ exist and are continuous on Θ, and the limits $\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\rho_k(\theta)\varpi_k(\theta)$, with $\rho_k(\theta), \varpi_k(\theta)$ replaced by $f(t_k,\theta)$, $h(t_k,\theta)$, $\chi(t_k,\theta)$ respectively, exist. This implies that the limit $M(\theta) = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}h(t_k,\theta)h'(t_k,\theta)$ exists for any θ in Θ.

(iii) The true value θ^* is an interior point of Θ and the matrix $M(\theta^*)$ is nonsingular.

The assertions below follow closely from the conditions imposed in Jennrich (1969), that render the NLS estimator $\widehat{\theta}_N$ a consistent estimator of θ^* and asymptotical normality via applying the corresponding convergent results of the NLS estimator in Jennrich (1969). The details are omitted here.

Theorem 4. *Let $\widehat{\theta}_N$ be the NLS estimates of (14). Under Assumptions 2 and 3(i), we have $\widehat{\theta}_N \to \theta^*$ with probability one as N tends to infinity. Further, if Assumptions 3(ii) and (iii) also hold, then*

$$\sqrt{N}(\widehat{\theta}_N - \theta^*) \to \mathcal{N}(0, \sigma^2 M^{-1}(\theta^*)) \quad \text{as } N \to \infty. \tag{15}$$

5.2. Iterative algorithms

Similar to the case without observation noise, the Gauss–Newton algorithm is a reliable iterative method to calculate the NLS estimator $\widehat{\theta}_N$ when the sampling size is large, which is given below.

$$\begin{aligned}\theta_N(k+1) &= \theta_N(k) + \big(J_N'(\theta_N(k))J_N(\theta_N(k))\big)^{-1} \\ &\quad \times J_N'(\theta_N(k))\big(Y_N - F_N(\theta_N(k))\big),\end{aligned} \tag{16}$$

where

$$J_N(\theta_N(k)) = \begin{bmatrix} h'(t_1, \theta_N(k)) \\ h'(t_2, \theta_N(k)) \\ \vdots \\ h'(t_N, \theta_N(k)) \end{bmatrix}, \qquad Y_N = \begin{bmatrix} y(t_1) \\ y(t_2) \\ \vdots \\ y(t_N) \end{bmatrix},$$

$$F_N(\theta_N(k)) = \begin{bmatrix} f(t_1, \theta_N(k)) \\ f(t_2, \theta_N(k)) \\ \vdots \\ f(t_N, \theta_N(k)) \end{bmatrix}.$$

Theorem 5 (*Jennrich, 1969*). *Let $\widehat{\theta}_N$ be the NLS estimator of (14). Under Assumptions 2 and 3, there exists a neighborhood $\mathfrak{G}$ of θ^* such that the Gauss–Newton iteration $\theta_N(k)$ given in (16) converges strongly to $\widehat{\theta}_N$ from any starting value in $\mathfrak{G}$ whenever the number of the sampling points $N \geq N_y$, where N_y is a positive integer.*

5.3. Recursive algorithms and strong convergence

Let θ_k be the estimate of θ^* at the time instance t_k. The parameter algorithm is updated recursively

$$\begin{aligned}\theta_{k+1} &= \theta_k + \varepsilon_k h(t_k, \theta_k)(y(t_k) - f(t_k, \theta_k)) \\ &= \theta_k + \varepsilon_k h(t_k, \theta_k)(f(t_k, \theta^*) - f(t_k, \theta_k) + e_k),\end{aligned} \tag{17}$$

where ε_k is the scalar step size satisfying the typical conditions ($\varepsilon_k > 0$, $\varepsilon_k \to 0$, and $\sum_{k=1}^{\infty} \varepsilon_k = \infty$). This is a stochastic approximation algorithm. For convergence analysis, we will employ the ODE (ordinary differential equation) approach (Kushner & Yin, 2003). The main conditions on the sampling times $\{t_1, t_2, \ldots\}$ and the input signals will be stated here first as assumptions and then elaborated later.

Ergodicity conditions: Suppose that the following limits exist locally. That is, there exists m_0 such that for all $\overline{m} > m_0$, and each θ, there exist continuous functions $M(\theta)$ and $W(\theta)$ such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=\overline{m}}^{\overline{m}+N-1} h(t_k, \theta)h'(t_k, \theta) \triangleq M(\theta), \tag{18}$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=\overline{m}}^{\overline{m}+N-1} h(t_k, \theta)(f(t_k, \theta^*) - f(t_k, \theta)) \triangleq W(\theta). \tag{19}$$

Now, consider the limit ODE

$$\dot{\theta} = W(\theta). \tag{20}$$

We note that $W(\theta^*) = 0$, implying that θ^* is an equilibrium point of (20). In addition, $\frac{\partial W(\theta)}{\partial \theta}\Big|_{\theta=\theta^*} = -M(\theta^*)$. By the ODE method (Kushner & Yin, 2003), we have the following convergence conclusion.

Theorem 6. *If $M(\theta^*)$ is full rank, then*

$$\theta_k \to \theta^* \text{ locally w.p.1 as } k \to \infty.$$

The condition of $M(\theta^*)$ being full rank is of "persistent excitation" types on the input and the sampling time sequences and will be explored in the next section.

6. Persistent excitation (PE) conditions

We now derive sufficient conditions under which the ergodicity conditions (18) and (19) exist, $M(\theta^*)$ is full rank, and Assumption 3 is valid. First, we comment that the condition on $M(\theta^*)$ being full rank is a type of "persistent excitation" conditions that are known to be needed for parameter estimation of discrete-time linear systems. Under uniform sampling schemes, this condition is translated into an input design problem for selecting suitable input probing signals. In irregular sampling schemes, additional complications arise since sampling times are not uniformly spaced.

6.1. PE conditions under irregular sampling

6.1.1. Repeated experiments with harmonic inputs

Recall that an input signal is said to be *harmonic of order L and base frequency ω* if $u(t) = \sum_{i=1}^{L} A_i \cos(i\omega t)$. This input is periodic with the period $T_0 = \frac{2\pi}{\omega}$. The system response to such a signal, without noise, is

$$f(t, \theta^*) = \phi(t, \theta^*) + O(\mu^t), \tag{21}$$

where

$$\phi(t, \theta^*) = \sum_{i=1}^{L} A_i |G(j\omega i, \theta^*)| \cos(i\omega t + \angle G(j\omega i, \theta^*))$$

with j being the imaginary unit ($j^2 = -1$) and $O(\mu^t)$ with $0 < \mu < 1$ are the steady-state and transient components of the system (1), respectively. An experiment is said to be *repeated of period T_0* if in the subsequent disjoint intervals $(iT_0, (i+1)T_0]$, $i = 1, 2, \ldots$, the sampling points are repeated as that in $(0, T_0]$, i.e., $t_{iN_0+l} = iT_0 + t_l$ with $t_0 = 0$ for $i \geq 1$, $1 \leq l \leq N_0$, where N_0 is the number of the sampling points in $(0, T_0]$. This is a typical scenario in the TDMA (time division multiple access) protocol of communication systems, in which each frame is of duration T_0 and frames are repeated. Define the $2n \times 2L$ matrix

$$\begin{aligned}\mathcal{J}(\theta) \triangleq \Big[&A_1 \frac{\partial |G(j\omega, \theta)|}{\partial \theta}, -A_1 |G(j\omega, \theta)| \frac{\partial \angle G(j\omega, \theta)}{\partial \theta}, \ldots, \\ &A_L \frac{\partial |G(j\omega L, \theta)|}{\partial \theta}, -A_L |G(j\omega L, \theta)| \frac{\partial \angle G(j\omega L, \theta)}{\partial \theta}\Big].\end{aligned}$$

Theorem 7. *Suppose that the input is harmonic of $L \geq n$, the sampling scheme is repeated with period T_0, $\mathcal{J}(\theta^*)$ is full rank, and $N_0 > \mu_T^* = (2n + 2L - 1) + \frac{\delta^* T_0}{2\pi}$, where $\delta^* = \max\{\bar{\delta}, 2L\omega\}$ and $\bar{\delta} = \max_{1\leq i,j\leq n}\{|\Im(\lambda_i - \lambda_j)|\}$ and $\{\lambda_i, i = 1, \ldots, n\}$ are the roots of $a(s)$. Then the limits (18) and (19) are valid, $M(\theta^*)$ is full rank, and Assumption 3 is valid.*

Theorem 8. *Under the conditions of Theorem 7 and Assumption 2, the iterative algorithm (16) converges to the true value and achieves the asymptotic normality (15) and the recursive algorithm (17) also converges with probability one.*

Proof. The result in Theorem 7 indicates that Assumptions 3(ii) and (iii) are true. It remains to show that Assumption 3(i) also holds. By the similar derivations in Theorem 7, we have

$$Q(\theta) \triangleq \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N}(f(t_k,\theta)-f(t_k,\theta^*))^2 = \frac{1}{N_0}\sum_{k=1}^{N_0}(\phi(t_k,\theta)-\phi(t_k,\theta^*))^2.$$

It follows that the Hessian matrix of $Q(\theta)$ at the point θ^* is positive definite by Theorem 7, which implies that $Q(\theta)$ has a unique minimum at $\theta=\theta^*$ over Θ. □

6.2. PE conditions under random sampling

In this subsection, we are concerned with two different kinds of random sampling rules including i.i.d. sampling processes and independent increment stochastic processes, and the corresponding input signals are designed to satisfy the PE conditions, respectively.

6.2.1. I.I.D. stochastic sampling processes

Assumption 4. The sampling sequence $\{t_k\}$ is independently produced from a non-degenerate continuous distribution F with the density function $p(t)$, where $t\in(p_1,p_2)\subset(0,\infty)$.

Theorem 9. *If the input $u(t)$ satisfies Assumption 1 where all the roots of $c(s)$ lie in the closed left-half complex plane and the sampling sequence $\{t_k\}$ satisfies Assumption 4, then Assumption 3 holds.*

Corollary 1. *Under the conditions of Theorem 9 and Assumption 2, the iterative algorithm (16) converges to the true value and achieves the asymptotic normality (15).*

Remark 3. When the sampling sequence $\{t_k\}$ generated by the way in Assumption 4 is resorted in the ascending order, the ergodicity conditions of (18) and (19) are valid and $M(\theta^*)$ is full rank. Consequently, the recursive algorithm (17) converges to the true values under this setting with probability one.

6.2.2. Independent increment stochastic sampling processes

Assumption 5. The input signal is designed as $u(t)=\sum_{i=1}^{L}A_i\cos(\omega_i t)$ with $L\geq n$, where the frequencies ω_i (with $i=1,\ldots,L$) are distinct.

Assumption 6. (i) The sampling process $\{t_k\}$ with $t_0=0$ is an independent increment stochastic process, that is, the inter-sampling time $\tau_k=t_k-t_{k-1}$ is a sequence of i.i.d. random variables taking positive value.

(ii) Denote the characteristic function of τ_k by $\varphi(\omega)$, i.e., $\varphi(\omega)=E(\exp(j\omega\tau_k))$. Further, $0<|\varphi(\omega)|<1$ and $|\frac{\varphi(2\omega)}{\varphi(\omega)}|<1$ for the frequencies $\omega=\omega_i, 2\omega_i, \omega_i+\omega_l, |\omega_i-\omega_l|$, where $\omega_i, \omega_l,\ 1\leq i,l\leq L$ are the frequencies in Assumption 5.

Remark 4. The inter-sampling time τ_k, a sequence of independent and exponentially distributed of rate $\widetilde{\eta}$, satisfies Assumption 6. Note that the characteristic function of the exponential distribution with rate $\widetilde{\eta}$ is $\varphi(\omega)=(1-j\omega/\widetilde{\eta})^{-1}$. It follows that $|\varphi(\omega)|=\widetilde{\eta}/\sqrt{\widetilde{\eta}^2+\omega^2}$ and $|\frac{\varphi(2\omega)}{\varphi(\omega)}|=\sqrt{\frac{\widetilde{\eta}^2+\omega^2}{\widetilde{\eta}^2+4\omega^2}}$, which indicates that $0<|\varphi(\omega)|<1$ and $|\varphi(2\omega)/\varphi(\omega)|<1$ for any frequency $\omega>0$.

Define

$$\widetilde{\mathcal{G}}(\theta) \triangleq \Big[A_1\frac{\partial|G(j\omega_1,\theta)|}{\partial\theta}, -A_1|G(j\omega_1,\theta)|\frac{\partial\angle G(j\omega_1,\theta)}{\partial\theta},\ldots, A_L\frac{\partial|G(j\omega_L,\theta)|}{\partial\theta}, -A_L|G(j\omega_L,\theta)|\frac{\partial\angle G(j\omega_L,\theta)}{\partial\theta}\Big].$$

Theorem 10. *If the input signal $u(t)$ and the sampling sequence $\{t_k\}$ satisfy Assumptions 5 and 6, respectively, then both Assumption 3 and ergodicity conditions (18) and (19) hold, and $M(\theta^*)$ is positive definite whenever the matrix $\widetilde{\mathcal{G}}(\theta^*)$ is full rank. Therefore, under additional Assumption 2, the iterative algorithm (16) converges to the true value and achieves the asymptotic normality (15); and the recursive algorithm (17) also converges with probability one.*

The number L of sinusoids in the input designs for repeated experiments with harmonic input and independent increment processes requires $L\geq n$, which is a sufficient condition to excite all the parameters of the plant. A benefit from increasing L can reduce the covariance matrix $\sigma^2M^{-1}(\theta^*)$ of the estimation errors in Theorem 4 by the forms of $M(\theta^*)$ for both cases in (30) and (41).

Remark 5. The results, including system identifiability and the PE conditions required by the convergence of the algorithm, are derived under the exponential type of inputs. For open-loop identification problems, the goal is to identify the system parameters by designing proper input signals and corresponding sampling time sequences. Hence the input is generally assumed to be at the user's disposal. The exponential type of inputs is used in this paper since they are simple and widely used in the experimental design. Future work will consider parameter estimation problems of continuous-time systems by the direct methods under other types of inputs (e.g., piece-wise constant inputs, etc.) under irregularly sampled inputs and outputs.

7. Illustrative examples

This section provides some examples to demonstrate the effectiveness of the algorithms developed in Sections 4 and 5.

Example 3. Consider a second-order transfer function $G(s)=\frac{b_1s+b_2}{s^2+a_1s+a_2}$, where $a_1=1.5$, $a_2=1$, $b_1=2.5$, and $b_2=-1$. Under the noise-free circumstance, the performance of the algorithm (11) is illustrated via the following five different exponential polynomial input signals: (1) the step input $1/s$; (2) the decaying input $1/(s+1)$; (3) the zero-pole cancellation input $1/(s-0.4)$; (4) the sinusoidal input $s/(s^2+1^2)$; and (5) a more complex input $(s^2+1.5s+1)/[(s^2+1^2)(s-2)]$ of order 3 canceling the poles of the plant. For all the input signals, the sampling points (0.0349, 0.1360, 0.2729, 0.4138, 0.5744, 0.5772, 0.6009, 0.6896) are generated from the uniform distribution over the interval [0, 1]. The number of the sampling points equals 8, which is greater than 4, 6, and 7 required in Theorem 3 for the first three inputs, the fourth input, and the last input of order 3, respectively. All the initial values of the algorithm (11) for the situations considered in this example are set to be zero vector $[0,0,0,0]'$. The iterative estimates for the true system parameters $\theta^*=[1.5,1,2.5,-1]'$ under the different input exciting signals are displayed in Table 1, which clearly show that the algorithm (11) quickly converges to the true values after 5 iterations.

Example 4. This example illustrates that the algorithm (11) can still find the true parameters of an unstable system by the noise-free input–output data. Let the transfer function be $G(s)=\frac{b_1s+b_2}{s^2+a_1s+a_2}$, where $a_1=-1.5$, $a_2=0$, $b_1=2.5$, and $b_2=-1$.

Table 1
Parameter estimation of a stable system under different input signals.

True values	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$U(s)=1/s$					
$a_1=1.5$	0.0000	0.8747	1.5032	1.5000	1.5000
$a_2=1$	0.0000	−0.7109	0.9931	1.0000	1.0000
$b_1=2.5$	2.3563	2.5000	2.5003	2.5000	2.5000
$b_2=-1$	−3.3994	−2.6904	−1.0010	−1.0000	−1.0000
$U(s)=1/(s+1)$					
$a_1=1.5$	0.0000	0.9891	1.5331	1.5001	1.5000
$a_2=1$	0.0000	−0.8529	1.0788	0.9997	1.0000
$b_1=2.5$	2.1758	2.5000	2.5008	2.5000	2.5000
$b_2=-1$	−2.8185	−2.5986	−0.9323	−0.9998	−1.0000
$U(s)=1/(s-0.4)$					
$a_1=1.5$	0.0000	0.8739	1.5032	1.5000	1.5000
$a_2=1$	0.0000	−0.7101	0.9928	1.0000	1.0000
$b_1=2.5$	2.3582	2.5000	2.5003	2.5000	2.5000
$b_2=-1$	−3.4032	−2.6907	−1.0009	−1.0000	−1.0000
$U(s)=s/(s^2+1^2)$					
$a_1=1.5$	0.0000	0.8746	1.5025	1.5000	1.5000
$a_2=1$	0.0000	−0.7105	0.9919	1.0000	1.0000
$b_1=2.5$	2.3556	2.5000	2.5003	2.5000	2.5000
$b_2=-1$	−3.4014	−2.6912	−1.0027	−1.0000	−1.0000
$U(s)=(s^2+1.5s+1)/((s^2+1^2)(s-2))$					
$a_1=1.5$	0.0000	0.8669	1.5020	1.5000	1.5000
$a_2=1$	0.0000	−0.7024	0.9877	1.0000	1.0000
$b_1=2.5$	2.3730	2.5000	2.5003	2.5000	2.5000
$b_2=-1$	−3.4393	−2.6943	−1.0035	−1.0000	−1.0000

Table 2
Parameter estimation of an unstable system under different input signals.

True values	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$U(s)=1/s$					
$a_1=-1.5$	0.0000	0.3996	−1.9253	−1.4878	−1.5000
$a_2=0$	0.0000	−2.3030	0.5364	−0.0137	0.0000
$b_1=2.5$	2.2762	2.4978	2.4993	2.5000	2.5000
$b_2=-1$	4.5792	3.7181	−2.0400	−0.9690	−1.0000
$U(s)=1/(s+1)$					
$a_1=-1.5$	0.0000	0.3943	−1.9157	−1.4869	−1.5000
$a_2=0$	0.0000	−2.2996	0.5255	−0.0147	0.0000
$b_1=2.5$	2.2687	2.4977	2.4993	2.5000	2.5000
$b_2=-1$	4.5953	3.7040	−2.0155	−0.9668	−1.0000
$U(s)=1/(s-0.4)$					
$a_1=-1.5$	0.0000	0.4017	−1.9303	−1.4882	−1.5000
$a_2=0$	0.0000	−2.3042	0.5420	−0.0132	0.0000
$b_1=2.5$	2.2795	2.4978	2.4993	2.5000	2.5000
$b_2=-1$	4.5732	3.7238	−2.0525	−0.9700	−1.0000
$U(s)=s/(s^2+1^2)$					
$a_1=-1.5$	0.0000	0.4006	−1.9207	−1.4876	−1.5000
$a_2=0$	0.0000	−2.3051	0.5309	−0.0139	0.0000
$b_1=2.5$	2.2752	2.4978	2.4993	2.5000	2.5000
$b_2=-1$	4.5751	3.7199	−2.0283	−0.9684	−1.0000
$U(s)=(s^2-1.5s)/((s^2+1^2)(s-2))$					
$a_1=-1.5$	0.0000	0.4012	−1.9321	−1.4883	−1.5000
$a_2=0$	0.0000	−2.3031	0.5440	−0.0130	0.0000
$b_1=2.5$	2.2801	2.4993	2.4993	2.5000	2.5000
$b_2=-1$	4.5745	3.7227	−2.0570	−0.9703	−1.0000

The sampling points and the input signals are the same as those in Example 3 except for the last input signal $U(s)=(s^2-1.5s)/((s^2+1^2)(s-2))$ canceling the poles of plant. Table 2 gives the iterative results of the algorithm (11) with the initial value $[0, 0, 0, 0]'$.

The following examples involve noisy observations. We introduce the signal-to-noise ratio (SNR)

$$\text{SNR} = \frac{\text{variance of signal}}{\text{variance of noise}} = \frac{\frac{1}{N-1}\sum_{i=1}^{N}\left(f(t_i,\theta^*)-\frac{1}{N}\sum_{i=1}^{N}f(t_i,\theta^*)\right)^2}{\sigma^2}$$

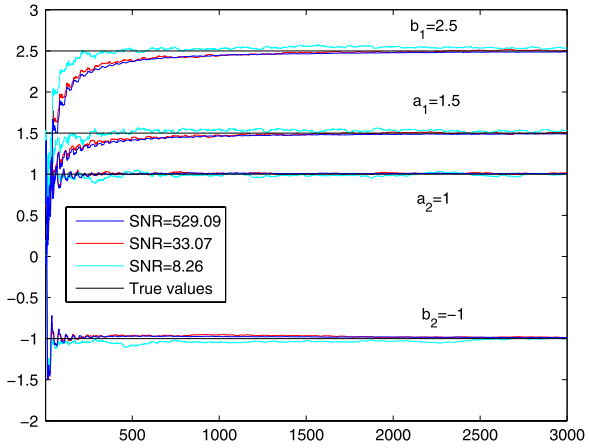


Fig. 1. Recursive estimation under repeatedly periodic sampling for different SNRs.

to demonstrate the impact of the different noise levels to the parameter estimation.

Example 5. In the noisy observation scenario, the effectiveness of the iterative algorithm (16) and the recursive algorithm (17) is verified by the second-order system given in Example 3. The input signal is $u(t) = \cos(t) + 2\cos(2t)$ (its Laplace transform is $U(s) = s/(s^2+1^2) + 2s/(s^2+2^2)$) with the period $T_0 = \pi/2$, and the sampling points in the interval $[0, \pi/2]$ are 0.2013, 0.2380, 0.3114, 0.4642, 0.4669, 0.5641, 0.6657, 0.7422, 1.0959, 1.1651, which are uniformly produced from the interval $[0, \pi/2]$. The successive sampling points are periodically repeated with the period $\pi/2$. In order to show the effectiveness of the algorithms under different noise levels, the observation noise $\{e_k\}$ is chosen to be a sequence of Gaussian random variables with zero mean for different variances: $0.1^2, 0.4^2, 0.8^2$, and the resulting SNRs are 529.09, 33.07, 8.26 since the variance of signal is 5.2909. In the following, all the estimates are based on the average of the estimates for ten runs. Table 3 summarizes the estimates and the corresponding squared sums of the estimation errors (SSEE) at the sample size $N = 600,\ 1200, 1800, 2400, 3000$ based on the iterative algorithm (16) with the initial value $[0, 0, 0, 0]'$ under different SNRs, where the values in the parentheses are the resulting standard deviations based on ten runs. While the recursive estimate is displayed in Fig. 1 in terms of the recursive algorithm (17) with the initial value $[0, 0, 0, 0]'$.

Example 6. The estimation effectiveness of the iterative algorithm (16) is illustrated in this example in the case of i.i.d. random sampling scheme. Consider the second-order transfer function given in Example 3. The input signal is $u(t) = 2,\ t \geq 0$ (its Laplace transform is $U(s) = 2/s$.) The sampling sequence $\{t_k\}$ comes from the uniform distribution over the interval $[0, 4]$. The observation noise $\{e_k\}$ is set to be a sequence of the Gaussian random variables with zero mean for different variances: $0.1^2, 0.4^2, 0.8^2$, and the resulting SNRs are 145.51, 9.09, 2.27 since the variance of signal is 1.4551. The initial value of the algorithm is set to be $[0, 0, 0, 0]'$. All the estimates are based on the average of ten runs. The estimates, the corresponding standard deviations in the parentheses, and the SSEE at the sampling sizes $N = 600, 1200, 1800, 2400, 3000$ are shown in Table 4 under different SNRs.

Example 7. The estimation effectiveness of the iterative algorithm (16) and the recursive algorithm (17) will be examined under the independent increment random sampling scheme. Consider the second-order transfer function given in Example 3. The input signal is $u(t) = \cos(t) + 2\cos(1.5t)$ (its Laplace transform is

Table 3
Iterative estimation under repeatedly periodic sampling for different SNRs.

N	600	1200	1800	2400	3000
$SNR = 529.09$					
$a_1 = 1.5$	1.5021 (0.0078)	1.5017 (0.0071)	1.5011 (0.0060)	1.5019 (0.0045)	1.5013 (0.0036)
$a_2 = 1$	0.9999 (0.0088)	1.0009 (0.0052)	0.9998 (0.0055)	0.9993 (0.0048)	0.9992 (0.0049)
$b_1 = 2.5$	2.5052 (0.0110)	2.5036 (0.0083)	2.5020 (0.0066)	2.5028 (0.0055)	2.5019 (0.0048)
$b_2 = -1$	−1.0016 (0.0179)	−0.9994 (0.0097)	−1.0010 (0.0088)	−1.0005 (0.0077)	−1.0007 (0.0073)
SSEE	3.36E−05	1.69E−05	6.20E−06	1.25E−05	6.24E−06
$SNR = 33.07$					
$a_1 = 1.5$	1.4773 (0.0217)	1.4910 (0.0159)	1.4915 (0.0185)	1.4946 (0.0220)	1.4973 (0.0154)
$a_2 = 1$	1.0040 (0.0241)	0.9995 (0.0258)	0.9983 (0.0267)	1.0002 (0.0203)	1.0002 (0.0191)
$b_1 = 2.5$	2.4737 (0.0310)	2.4929 (0.0240)	2.4913 (0.0282)	2.4921 (0.0307)	2.4969 (0.0221)
$b_2 = -1$	−0.9975 (0.0489)	−1.0043 (0.0395)	−1.0029 (0.0431)	−1.0024 (0.0344)	−1.0011 (0.0339)
SSEE	1.23E−03	1.49E−04	1.59E−04	9.65E−05	1.83E−05
$SNR = 8.26$					
$a_1 = 1.5$	1.5397 (0.1008)	1.5073 (0.0667)	1.4941 (0.0415)	1.4917 (0.0378)	1.4987 (0.0318)
$a_2 = 1$	0.9009 (0.3138)	1.0119 (0.0449)	0.9992 (0.0275)	0.9949 (0.0285)	1.0000 (0.0319)
$b_1 = 2.5$	2.5663 (0.1582)	2.5117 (0.0958)	2.4975 (0.0650)	2.4898 (0.0552)	2.5024 (0.0419)
$b_2 = -1$	−1.1488 (0.4724)	−0.9777 (0.0671)	−1.0002 (0.0544)	−1.0017 (0.0478)	−0.9946 (0.0491)
SSEE	3.79E−02	8.28E−04	4.14E−05	2.01E−04	3.64E−05

$U(s) = s/(s^2 + 1^2) + 2s/(s^2 + 1.5^2)$). The inter-sampling sequence $\{\tau_k\}$ comes from the exponential distribution with rate $\widetilde{\eta} = 0.4$. To show the performance of the iterative and recursive algorithms given in the paper under different SNRs, the observation noise $\{e_k\}$ is set to be a sequence of Gaussian random variables with different variances: 0.1^2, 0.4^2, 0.8^2, and the corresponding SNRs are 630.65, 39.42, 9.85 since the variance of the signal is 6.3065, respectively. The estimates given below are based on the average of ten runs. The estimates obtained by the iterative algorithm (16) with the initial value $[0, 0, 0, 0]'$, the resulting SSEE and the standard deviations for different SNRs at the sampling sizes $N = 600, 1200, 1800, 2400, 3000$ are illustrated in Table 5. Under the same setting, the recursive estimation generated by the algorithm (17) with the initial value $[0, 0, 0, 0]'$ is presented in Fig. 2.

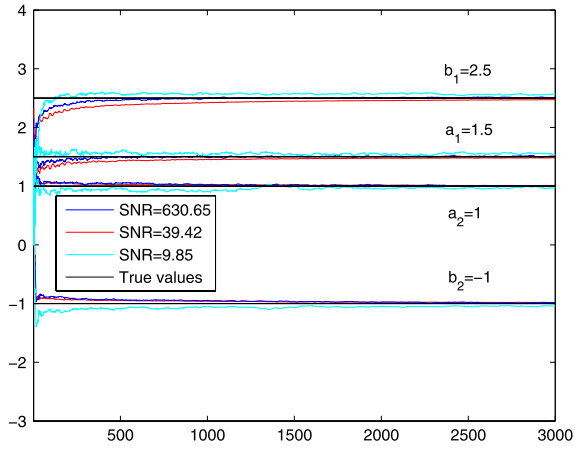


Fig. 2. Recursive estimation under Poisson sampling process for different SNRs.

8. Concluding remarks

Under the framework of irregular and random sampling, identifiability, identification algorithms, convergence properties, and input design of linear time-invariant continuous systems have been investigated in this paper. The identifiability conditions obtained in this paper have established that the parameters of coprime rational systems can be uniquely determined from noise-free input–output data if the number of the sampling points exceeds some fixed positive integer in a given time interval regardless of the sampling schemes (periodic, irregular, or random). The strongly convergent iterative and recursive algorithms have been developed to identify system parameters under noisy observations. The convergence properties of these algorithms depend critically on input richness, sampling schemes, and sampling sizes. The persistent excitation conditions under irregular and random sampling have been established.

There are still several significant open questions. Applications of the algorithms and input design developed in this paper to practical systems are important steps to verify true values and also limitations of the algorithms. Extension to nonlinear systems is possible and worth investigation.

Appendix

In the following proofs, we encounter frequently the partial derivatives (gradients) of $g(t, \theta)$ and $f(t, \theta)$ with respect to θ. It is straightforward to verify that in the region of convergence of the

Table 4
Iterative estimation under i.i.d. random sampling for different SNRs.

N	600	1200	1800	2400	3000
$SNR = 145.51$					
$a_1 = 1.5$	1.5003 (0.0246)	1.5005 (0.0165)	1.5034 (0.0127)	1.5030 (0.0151)	1.5043 (0.0139)
$a_2 = 1$	1.0049 (0.0190)	1.0022 (0.0112)	0.9963 (0.0090)	0.9962 (0.0102)	0.9961 (0.0102)
$b_1 = 2.5$	2.5003 (0.0199)	2.4977 (0.0159)	2.5013 (0.0111)	2.5022 (0.0135)	2.5041 (0.0112)
$b_2 = -1$	−0.9988 (0.0159)	−0.9981 (0.0107)	−1.0025 (0.0073)	−1.0028 (0.0093)	−1.0036 (0.0079)
SSEE	2.60E−05	1.41E−05	3.37E−05	3.62E−05	6.32E−05
$SNR = 9.09$					
$a_1 = 1.5$	1.5436 (0.1558)	1.5308 (0.1140)	1.4996 (0.0821)	1.5026 (0.0791)	1.4986 (0.0626)
$a_2 = 1$	0.9483 (0.0804)	0.9639 (0.0625)	0.9771 (0.0482)	0.9761 (0.0395)	0.9840 (0.0347)
$b_1 = 2.5$	2.5338 (0.1414)	2.5249 (0.1062)	2.4996 (0.0814)	2.4979 (0.0789)	2.4932 (0.0727)
$b_2 = -1$	−1.0340 (0.0971)	−1.0286 (0.0642)	−1.0089 (0.0485)	−1.0073 (0.0454)	−1.0029 (0.0393)
SSEE	6.87E−03	3.69E−03	6.05E−04	6.37E−04	3.11E−04
$SNR = 2.27$					
$a_1 = 1.5$	1.2593 (0.1894)	1.3085 (0.1568)	1.3589 (0.1581)	1.3809 (0.1680)	1.4043 (0.1555)
$a_2 = 1$	1.0779 (0.1533)	1.0490 (0.1346)	1.0308 (0.1344)	1.0345 (0.1115)	1.0394 (0.0906)
$b_1 = 2.5$	2.2796 (0.2163)	2.3283 (0.1589)	2.3816 (0.1403)	2.4004 (0.1475)	2.4264 (0.1494)
$b_2 = -1$	−0.8645 (0.1417)	−0.8988 (0.1159)	−0.9282 (0.1143)	−0.9347 (0.1130)	−0.9453 (0.1044)
SSEE	1.31E−01	7.88E−02	4.00E−02	2.96E−02	1.91E−02

Laplace transform

$$\frac{\partial G(s,\theta)}{\partial \theta} = \left[-\frac{b(s)[s^{n-1},\ldots,s,1]}{a^2(s)}, \frac{[s^{n-1},\ldots,s,1]}{a(s)}\right]'$$

and its inverse Laplace transform is

$$\frac{\partial g(t,\theta)}{\partial \theta} = \mathcal{L}^{-1}\left(\frac{\partial G(s,\theta)}{\partial \theta}\right),$$

which remains a stable exponential polynomial. It is clear that $g(t,\theta)$ and $\frac{\partial g(t,\theta)}{\partial \theta}$ are continuous in the region $(0, t_f) \times \mathcal{A}$ for any $0 < t < t_f$ and $\theta \in \mathcal{A}$.

Proof of Lemma 1. By (4), we have

$$h(t,\theta) = \frac{\partial f(t,\theta)}{\partial \theta} = \frac{\partial g(t,\theta)}{\partial \theta} \star u(t). \tag{22}$$

Its equivalent representation in the frequency domain is given by

$$\begin{aligned}\mathcal{L}(h(t,\theta)) &= \frac{\partial G(s,\theta)}{\partial \theta} U(s)\\ &= \left[-\frac{b(s)[s^{n-1},\ldots,s,1]}{a^2(s)}, \frac{[s^{n-1},\ldots,s,1]}{a(s)}\right]' U(s)\\ &= \frac{U(s)}{a^2(s)}\left[-b(s)[s^{n-1},\ldots,s,1], a(s)[s^{n-1},\ldots,s,1]\right]'. \end{aligned}\tag{23}$$

Suppose that $\gamma = [\alpha_1,\ldots,\alpha_n,\beta_1,\ldots,\beta_n]' \in \mathbb{R}^{2n}$ is a nonzero vector such that

$$h'(t,\theta)\gamma = 0, \quad \forall t \geq 0, \tag{24}$$

where $h'(t,\theta)$ is the transpose of $h(t,\theta)$. By (23), (24), and the linearity property of the Laplace transform, we have

$$\mathcal{L}\left(h'(t,\theta)\gamma\right) = \frac{U(s)}{a^2(s)}(a(s)\beta(s) - b(s)\alpha(s)) = 0 \quad \forall s \in \mathbb{C},$$

where $\beta(s) \triangleq \beta_1 s^{n-1} + \cdots + \beta_{n-1}s + \beta_n$ and $\alpha(s) \triangleq \alpha_1 s^{n-1} + \cdots + \alpha_{n-1}s + \alpha_n$. It follows from $u(t) \not\equiv 0$ that

$$a(s)\beta(s) - b(s)\alpha(s) = 0, \quad \forall s \in \mathbb{C}. \tag{25}$$

Apparently, if $\alpha(s) = 0$ (or $\beta(s) = 0$), then (25) implies $\beta(s) = 0$ (or $\alpha(s) = 0$). Therefore, suppose that neither $\beta(s)$ nor $\alpha(s)$ is the zero function. Then $b(s)/a(s) = \beta(s)/\alpha(s)$. Since $\deg \beta(s) \leq n-1$ and $\deg \alpha(s) \leq n-1$, the right-hand side is a rational function of order at most $n-1$ and the left-hand side is a rational function of order n. But this contradicts the fact that $a(s)$ and $b(s)$ are coprime (so that $gcd(a(s), b(s)) = 1$ and the order of $b(s)/a(s)$ cannot be reduced). As a result, the only solution is $\beta(s) = 0$ and $\alpha(s) = 0$, which implies that the vector $\gamma = [\alpha_1,\ldots,\alpha_n,\beta_1,\ldots,\beta_n]' = 0$. Therefore, the lemma is proved. □

Proof of Lemma 3. By (23), we have

$$\begin{aligned}&\mathcal{L}(h(t,\theta))\\ &= \left[\frac{-b(s)d(s)[s^{n-1},\ldots,s,1]}{a^2(s)c(s)}, \frac{d(s)[s^{n-1},\ldots,s,1]}{a(s)c(s)}\right]'.\end{aligned}\tag{26}$$

As a result, taking inverse Laplace transform on both sides of (26) gives

$$\begin{aligned}h(t,\theta) = \mathcal{L}^{-1}\Big(\Big[&-\frac{b(s)d(s)[s^{n-1},\ldots,s,1]}{a^2(s)c(s)},\\ &\frac{d(s)[s^{n-1},\ldots,s,1]}{a(s)c(s)}\Big]'\Big),\end{aligned}$$

Table 5
Iterative estimation under Poisson sampling process for different SNRs.

N	600	1200	1800	2400	3000
$SNR = 630.65$					
$a_1 = 1.5$	1.4958 (0.0066)	1.4973 (0.0052)	1.4975 (0.0057)	1.4984 (0.0039)	1.4978 (0.0042)
$a_2 = 1$	0.9953 (0.0093)	0.9987 (0.0049)	1.0003 (0.0051)	1.0001 (0.0047)	0.9998 (0.0043)
$b_1 = 2.5$	2.4952 (0.0100)	2.4966 (0.0070)	2.4970 (0.0086)	2.4980 (0.0063)	2.4967 (0.0070)
$b_2 = -1$	−1.0081 (0.0174)	−1.0031 (0.0087)	−0.9998 (0.0081)	−1.0008 (0.0078)	−1.0011 (0.0079)
SSEE	1.28E−04	3.04E−05	1.56E−05	6.87E−06	1.73E−05
$SNR = 39.42$					
$a_1 = 1.5$	1.5138 (0.0360)	1.5057 (0.0259)	1.5033 (0.0148)	1.5028 (0.0151)	1.5047 (0.0143)
$a_2 = 1$	0.9983 (0.0380)	0.9993 (0.0400)	1.0094 (0.0242)	1.0003 (0.0229)	1.0026 (0.0175)
$b_1 = 2.5$	2.5236 (0.0503)	2.5085 (0.0401)	2.5052 (0.0234)	2.5071 (0.0224)	2.5084 (0.0195)
$b_2 = -1$	−0.9925 (0.0729)	−0.9987 (0.0639)	−0.9848 (0.0402)	−0.9967 (0.0387)	−0.9912 (0.0279)
SSEE	8.05E−04	1.07E−04	3.57E−04	6.93E−05	1.78E−04
$SNR = 9.85$					
$a_1 = 1.5$	1.5451 (0.1358)	1.5435 (0.1217)	1.5121 (0.0435)	1.5034 (0.0375)	1.4907 (0.0310)
$a_2 = 1$	0.9339 (0.3217)	0.9072 (0.3073)	1.0083 (0.0271)	1.0043 (0.0301)	0.9960 (0.0200)
$b_1 = 2.5$	2.5575 (0.2055)	2.5742 (0.2033)	2.5176 (0.0655)	2.5010 (0.0517)	2.4834 (0.0396)
$b_2 = -1$	−1.0841 (0.4737)	−1.1281 (0.4783)	−0.9721 (0.0543)	−0.9873 (0.0593)	−1.0008 (0.0400)
SSEE	1.68E−02	3.24E−02	1.30E−03	1.92E−04	3.78E−04

which implies that each element of $h(t,\theta)$ is either $\mathcal{L}^{-1}\left(\frac{d(s)s^i}{a(s)c(s)}\right)$ or $\mathcal{L}^{-1}\left(\frac{b(s)d(s)s^i}{a^2(s)c(s)}\right)$, $0 \le i \le n-1$. Assume that $\mathcal{M}$ is the set that contains all the modes of the exponential polynomials generated by $a^2(s)c(s)$, i.e., $\mathcal{M} = \{p_{i,j}(t), i = 1,\ldots,l_w, j = 1,\ldots,n_i\}$ with $\sum_{i=1}^{l_w} n_i = 2n+m$ and $p_{i,j}(t) = \frac{t^{j-1}}{(j-1)!}\exp(\lambda_i t)$, and $\{\lambda_i, i = 1,\ldots,l_w\}$ are the l_w distinct roots of $a^2(s)c(s)$ with multiplicity n_i. Then each element of $h(t,\theta)$ is a weighted sum of the modes $\{p_{i,j}(t)\}$. As a result, $h'(t,\theta)$ can be expressed as the following linear combination

$$h'(t,\theta) = [p_{1,1}(t),\ldots,p_{1,n_1}(t),\ldots,p_{l_w,n_{l_w}}(t)]\Lambda, \tag{27}$$

where Λ is the corresponding coefficient matrix. Λ is full column rank since $h(t,\theta)$ is linearly independent by Lemma 1.

Under the sampling points $(t_1,\ldots,t_N)$, denote

$$P_N = \begin{bmatrix} p_{1,1}(t_1) & \cdots & p_{1,n_1}(t_1) & \cdots & p_{l_w,n_{l_w}}(t_1) \\ p_{1,1}(t_2) & \cdots & p_{1,n_1}(t_2) & \cdots & p_{l_w,n_{l_w}}(t_2) \\ \vdots & & \vdots & & \vdots \\ p_{1,1}(t_N) & \cdots & p_{1,n_1}(t_N) & \cdots & p_{l_w,n_{l_w}}(t_N) \end{bmatrix}.$$

Apparently, $J_N(\theta) = P_N\Lambda$. Suppose that there exists a vector $\gamma = [\alpha_1,\ldots,\alpha_n,\beta_1,\ldots,\beta_n]' \neq 0$ such that $J_N(\theta)\gamma = P_N\Lambda\gamma = 0$. Define a function $w(t) = [p_{1,1}(t),\ldots,p_{1,n_1}(t),\ldots,p_{l_w,n_{l_w}}(t)]\Lambda\gamma$. Then, $J_N(\theta)\gamma = [w(t_1),\ldots,w(t_N)]' = 0$ implies that $w(t)$ has at least N zeros in $[0 \quad T]$. However, by Lemma 2, the number of zeros of $w(t)$ is bounded by μ_T^*.

This contradiction implies that $\Lambda\gamma = 0$. As a result, $\gamma = 0$ since Λ is full column rank. It follows that the Jacobian matrix $J_N(\theta)$ has full column rank. □

Proof of Theorem 3. The theorem is proved inductively. We first show that (12)–(13) hold for $k = 0$. Since $J_N(\theta)$ is Lipschitz continuous on some convex and compact subset Θ of $\mathbb{R}^{2n}$ containing the true value θ^*, there exists a constant $\eta > 0$ such that

$$\|J_N(x) - J_N(y)\| \le \eta\|x-y\|, \quad \forall x,y \in \Theta. \tag{28}$$

Furthermore, we have $\|J_N(x)\| \le \nu$, $\forall x \in \Theta$, where ν is a positive real number. Since $J_N(\theta^*)$ is full column rank by Lemma 3, the minimum eigenvalue λ of $J_N'(\theta^*)J_N(\theta^*)$ is positive. Selecting any initial value $\theta_0 \in O(\theta^*,\epsilon)$ with $\epsilon = \frac{\lambda}{4\nu\eta}$, then

$$\begin{aligned} &\|J_N'(\theta^*)J_N(\theta^*) - J_N'(\theta_0)J_N(\theta_0)\| \\ &\quad \le \|J_N'(\theta^*)(J_N(\theta^*) - J_N(\theta_0))\| + \|(J_N'(\theta^*) - J_N'(\theta^*))J_N(\theta_0)\| \\ &\quad \le 2\nu\eta\|\theta^* - \theta_0\| < \lambda/2. \end{aligned}$$

Since $\|\left(J_N'(\theta^*)J_N(\theta^*)\right)^{-1}\| \le 1/\lambda$, the conditions of Lemma 4 hold, and hence $J_N'(\theta_0)J_N(\theta_0)$ is nonsingular and $\left\|\left(J_N'(\theta_0)J_N(\theta_0)\right)^{-1}\right\| \le 2/\lambda$ whenever $\theta_0 \in O(\theta^*,\epsilon)$ with $\epsilon = \frac{\lambda}{4\nu\eta}$.

As $J_N(\theta)$ is Lipschitz continuous on Θ and $r(\theta^*) = 0$, we have

$$\begin{aligned} &\| - r(\theta_0) + J_N(\theta_0)(\theta^* - \theta_0)\| \\ &\quad = \left\| \int_0^1 -J_N(\theta_0 + s(\theta^* - \theta_0))(\theta^* - \theta_0)\mathrm{d}s + J_N(\theta_0)(\theta^* - \theta_0) \right\| \\ &\quad \le \int_0^1 \|J_N(\theta_0 + s(\theta^* - \theta_0)) - J_N(\theta_0)\| \, \|(\theta^* - \theta_0)\|\mathrm{d}s \\ &\quad \le \int_0^1 \eta\|s(\theta^* - \theta_0)\| \, \|(\theta^* - \theta_0)\|\mathrm{d}s \le \frac{\eta}{2}\|\theta^* - \theta_0\|^2. \end{aligned}$$

By the iterative algorithm (11), we have

$$\|\theta_1-\theta^*\| \le \frac{2}{\lambda}\frac{\nu\eta}{2}\|\theta_0-\theta^*\|^2 = \frac{\nu\eta}{\lambda}\|\theta_0-\theta^*\|^2 = \frac{4\nu\eta\|\theta_0-\theta^*\|}{\lambda}\frac{\|\theta_0-\theta^*\|}{4} < \frac{1}{4}\|\theta_0-\theta^*\|.$$

This means that (12) and (13) hold for $k=0$, and the constant $\varrho=\nu\eta/\lambda$ in (13). By induction, we see that (12) and (13) hold for $k\ge 1$ by the similar proof as that used for $k=0$. Hence, it follows that the conclusion of the theorem holds true. □

Proof of Theorem 7. It is clear that $|G(j\omega i,\theta)|$ and $\angle G(j\omega i,\theta)$ for any $1\le i\le L$ are continuous with respect to θ in a small neighborhood Θ of θ^*. It follows that

$$f(t,\theta)=\phi(t,\theta)+O(\mu^t), \tag{29}$$

where $\phi(t,\theta)=\sum_{i=1}^{L}A_i|G(j\omega i,\theta)|\cos(i\omega t+\angle G(j\omega i,\theta))$ is a periodic function with period T_0, i.e., $\phi(t+T_0,\theta)=\phi(t,\theta)$. It follows from (29) that

$$h(t_k,\theta)=\frac{\partial f(t_k,\theta)}{\partial\theta}=\frac{\partial\phi(t_k,\theta)}{\partial\theta}+O(\mu^{t_k}),$$

where

$$\begin{aligned}\frac{\partial\phi(t_k,\theta)}{\partial\theta} &= \sum_{i=1}^{L}A_i\cos(i\omega t_k+\angle G(j\omega,\theta))\frac{\partial|G(j\omega i,\theta)|}{\partial\theta}\\ &\quad-\sum_{i=1}^{L}A_i|G(j\omega i,\theta)|\sin(i\omega t_k\\ &\quad+\angle G(j\omega i,\theta))\frac{\partial\angle G(j\omega i,\theta)}{\partial\theta}\\ &= \mathcal{G}(\theta)\zeta_k,\end{aligned}$$

is also a periodic function with period T_0, and

$$\begin{aligned}\zeta_k=[&\cos\big(wt_k+\angle G(j\omega,\theta)\big),\sin\big(wt_k+\angle G(j\omega,\theta)\big),\ldots,\\ &\cos\big(wLt_k+\angle G(j\omega L,\theta)\big),\sin\big(wLt_k+\angle G(j\omega L,\theta)\big)]'.\end{aligned}$$

Therefore, for any fixed $\overline{m}>0$ we have

$$\begin{aligned}&\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}h(t_k,\theta)h'(t_k,\theta)\\ &=\frac{1}{N}\left\lfloor\frac{N}{N_0}\right\rfloor\sum_{l=0}^{N_0-1}\frac{\partial\phi(t_{\overline{m}+l},\theta)}{\partial\theta}\left(\frac{\partial\phi(t_{\overline{m}+l},\theta)}{\partial\theta}\right)'+O\left(\frac{1}{N}\right)\\ &\xrightarrow{N\to\infty}\frac{1}{N_0}\sum_{l=0}^{N_0-1}\frac{\partial\phi(t_{\overline{m}+l},\theta)}{\partial\theta}\left(\frac{\partial\phi(t_{\overline{m}+l},\theta)}{\partial\theta}\right)'\\ &=\frac{1}{N_0}\sum_{l=1}^{N_0}\frac{\partial\phi(t_l,\theta)}{\partial\theta}\left(\frac{\partial\phi(t_l,\theta)}{\partial\theta}\right)'\triangleq M(\theta),\end{aligned}$$

where $\lfloor a\rfloor$ represents the integer part of a positive real number a. As a result, we have

$$\begin{aligned}&\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}h(t_k,\theta^*)h'(t_k,\theta^*)\\ &\xrightarrow[N\to\infty]{}\frac{1}{N_0}\sum_{l=1}^{N_0}\frac{\partial\phi(t_l,\theta^*)}{\partial\theta}\left(\frac{\partial\phi(t,\theta^*)}{\partial\theta}\right)'\\ &=\frac{1}{N_0}\Phi'(\theta^*)\Phi(\theta^*)=M(\theta^*),\end{aligned} \tag{30}$$

where

$$\begin{aligned}\Phi'(\theta^*) &= \left[\frac{\partial\phi(t_1,\theta^*)}{\partial\theta},\frac{\partial\phi(t_2,\theta^*)}{\partial\theta},\ldots,\frac{\partial\phi(t_{N_0},\theta^*)}{\partial\theta}\right]\\ &= \mathcal{G}(\theta^*)\left[\zeta_1,\zeta_2,\ldots,\zeta_{N_0}\right].\end{aligned}$$

It follows that $\Phi(\theta^*)$ has full column rank by the similar treatment as that used in Lemmas 1 and 3 if $\mathcal{G}(\theta^*)$ is full rank and $N_0>\mu_{T_0^*}$. As a result, $M(\theta^*)$ is positive definite since N_0 is finite. □

Proof of Theorem 9. According to the strong law of large numbers, under Assumption 1 the limit

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}f^2(t_k,\theta) &= Ef^2(t_1,\theta)\\ &= \int_{p_1}^{p_2}f^2(t,\theta)p(t)\mathrm{d}t\ \text{w.p.1}.\end{aligned} \tag{31}$$

Since $f(t,\theta)$ is a bounded exponential polynomial, the integral (31) exists if p_2 is finite. When $p_2=\infty$, noting that $p(t)$ is a density function, for any $A_2>A_1>0$,

$$\begin{aligned}\left|\int_{A_1}^{A_2}f^2(t,\theta)p(t)\mathrm{d}t\right| &\le \widetilde{M}\left|\int_{A_1}^{A_2}\big(\mu^t+\cos(\omega t)+1\big)p(t)\mathrm{d}t\right|\\ &\le \widetilde{M}\left|\int_{A_1}^{A_2}p(t)\mathrm{d}t\right|\xrightarrow{A_1\to\infty}0,\end{aligned}$$

where $\widetilde{M}$ is a positive constant and $0<\mu<1$. This means that the integral (31) also exists under this setting. Similarly, we have

$$\begin{aligned}Q(\theta) &= \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}(f(t_k,\theta)-f(t_k,\theta^*))^2\\ &= E(f(t_k,\theta)-f(t_k,\theta^*))^2\\ &= \int_{p_1}^{p_2}(f(t,\theta)-f(t,\theta^*))^2p(t)\mathrm{d}t\ \text{w.p.1}\end{aligned} \tag{32}$$

which is well defined. It follows that the Hessian matrix of $Q(\theta)$ at the point θ^*

$$Q_{\theta\theta}(\theta^*)=2\int_{p_1}^{p_2}h(t,\theta^*)h'(t,\theta^*)p(t)\mathrm{d}t.$$

In what follows, we will show that the Hessian matrix $Q_{\theta\theta}(\theta^*)$ is positive definite. Let $\beta=[\beta_1,\beta_2,\ldots,\beta_{2n}]'\ne 0$ be a column vector such that

$$\beta'Q_{\theta\theta}(\theta^*)\beta=2\int_{p_1}^{p_2}\beta'h(t,\theta^*)h'(t,\theta^*)\beta p(t)\mathrm{d}t=0.$$

Then we have $\beta'h(t,\theta^*)=0$ since $p(t)$ is a density function. By Lemma 1, $h(t,\theta^*)$ is linearly independent, which derives that $\beta=0$. As a result, we have proved that $Q_{\theta\theta}(\theta^*)$ is positive definite. Hence, $Q(\theta)$ has a unique minimum at $\theta=\theta^*$. It follows that Assumption 3(i) is true. Assumption 3(ii) can be proved similarly. As the set $\mathcal{A}$ defined in Section 2.1 is open, the true value θ^* is an interior point of some Θ containing θ^*. Since

$$\begin{aligned}M(\theta) &= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}h(t_k,\theta)h'(t_k,\theta)=Eh(t_k,\theta)h'(t_k,\theta)\\ &= \int_{p_1}^{p_2}h(t,\theta)h'(t,\theta)p(t)\mathrm{d}t\ \text{w.p.1},\end{aligned}$$

also exists for any $\theta\in\Theta$, we have $M(\theta^*)$ is positive definitive by the proof above. It follows that Assumption 3 (iii) also holds. □

Lemma 5. *Under Assumption 6, for any initial phase ξ, the following assertions hold at the frequency ω satisfying $0 < |\varphi(\omega)| < 1$ and $|\varphi(2\omega)/\varphi(\omega)| < 1$:*

$$\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\cos(\omega t_k+\xi)\xrightarrow[N\to\infty]{}0 \text{ w.p.1}, \tag{33}$$

$$\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\sin(\omega t_k+\xi)\xrightarrow[N\to\infty]{}0 \text{ w.p.1}, \tag{34}$$

while at the frequency ω satisfying $0 < |\varphi(2\omega)| < 1$ and $|\varphi(4\omega)/\varphi(2\omega)| < 1$ there hold

$$\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\cos^2(\omega t_k+\xi)\xrightarrow[N\to\infty]{}1/2 \text{ w.p.1}, \tag{35}$$

$$\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\sin^2(\omega t_k+\xi)\xrightarrow[N\to\infty]{}1/2 \text{ w.p.1}. \tag{36}$$

Proof. Because

$$\cos(\omega t_k+\xi)=\cos(\omega t_k)\cos(\xi)-\sin(\omega t_k)\sin(\xi),$$

to prove (33) it suffices to verify that

$$\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\cos(\omega t_k)\xrightarrow[N\to\infty]{}0 \text{ w.p.1} \quad \textit{and}$$

$$\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\sin(\omega t_k)\xrightarrow[N\to\infty]{}0 \text{ w.p.1}$$

which is equivalent to

$$\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\exp(j\omega t_k)\xrightarrow[N\to\infty]{}0 \text{ w.p.1}. \tag{37}$$

Using $t_k=\sum_{i=1}^{k}\tau_i$, the characteristic function of t_k equals

$$\begin{aligned}E\exp(j\omega t_k)&=E\exp\Big(j\omega\sum_{i=1}^{k}\tau_i\Big)\\&=\prod_{i=1}^{k}E\exp(j\omega\tau_i)=\varphi^k(\omega).\end{aligned}$$

Since $\{\tau_k\}$ is an i.i.d. sequence, we have

$$\begin{aligned}&E\Big[\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\exp(j\omega t_k)\Big]^2\\&=\frac{1}{N^2}\sum_{k=\overline{m}}^{\overline{m}+N-1}\sum_{l=\overline{m}}^{\overline{m}+N-1}E\exp(j\omega t_k)\exp(j\omega t_l)\\&=\frac{2}{N^2}\sum_{k=\overline{m}}^{\overline{m}+N-1}\Big[\frac{\varphi(2\omega)}{\varphi(\omega)}\Big]^k\frac{\varphi^{k+1}(\omega)-\varphi^{\overline{m}+N}(\omega)}{1-\varphi(\omega)}\\&\quad+\frac{1}{N^2}\frac{\varphi^{\overline{m}}(2\omega)-\varphi^{\overline{m}+N}(2\omega)}{1-\varphi(2\omega)}=O\Big(\frac{1}{N^2}\Big).\end{aligned} \tag{38}$$

It follows from (38) that for any $\epsilon>0$,

$$\begin{aligned}&\sum_{N=1}^{\infty}P\left(\left|\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\exp(j\omega t_k)\right|>\epsilon\right)\\&\le\sum_{N=1}^{\infty}\frac{E\Big[\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\exp(j\omega t_k)\Big]^2}{\epsilon^2}\\&=\sum_{N=1}^{\infty}O\Big(\frac{1}{N^2}\Big)<\infty.\end{aligned} \tag{39}$$

Hence, by the Borel–Cantelli lemma (Chow & Teicher, 2003), one derives that

$$\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\exp(j\omega t_k)\to 0 \text{ w.p.1 as } N\to\infty.$$

Similarly, we can prove (34). Noting $\cos^2(\omega t_k+\xi)=1/2+1/2\cos(2\omega t_k+2\xi)$, by (33) we have

$$\begin{aligned}&\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\cos^2(\omega t_k+\xi)\\&=\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}(1/2+1/2\cos(2\omega t_k+2\xi))\\&=\frac{1}{2}+\frac{1}{2N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\cos(2\omega t_k+2\xi)\xrightarrow[N\to\infty]{}\frac{1}{2} \text{ w.p.1.}\end{aligned}$$

By the triangle identity $\sin^2(\omega t_k+\xi)=1/2-1/2\cos(2\omega t_k+2\xi)$, the assertion (36) can be verified similarly. The lemma is thus proved. □

Proof of Theorem 10. Under Assumption 5, the noise-free output is

$$f(t,\theta)=\phi(t,\theta)+O(\mu^t), \tag{40}$$

where $\phi(t,\theta)=\sum_{i=1}^{L}A_i|G(j\omega_i,\theta)|\cos(\omega_i t+\angle G(j\omega_i,\theta))$ and $0<\mu<1$.

Similar to the derivations in Theorem 7, under Assumptions 5 and 6, the gradient of (40) is given by

$$\begin{aligned}h(t_k,\theta)=\frac{\partial f(t_k,\theta)}{\partial\theta}&=\frac{\partial\phi(t_k,\theta)}{\partial\theta}+O(\mu^{t_k})\\&=\widetilde{\mathcal{G}}(\theta)\widetilde{\zeta}_k+O(\mu^{t_k}),\end{aligned}$$

where

$$\begin{aligned}\widetilde{\zeta}_k=[&\cos\big(w_1t_k+\angle G(j\omega_1,\theta)\big),\sin\big(w_1t_k+\angle G(j\omega_1,\theta)\big),\ldots,\\&\cos\big(w_Lt_k+\angle G(j\omega_L,\theta)\big),\sin\big(w_Lt_k+\angle G(j\omega_L,\theta)\big)]'.\end{aligned}$$

Denote $\boldsymbol{\zeta}_N=\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\widetilde{\zeta}_k\widetilde{\zeta}_k'$. Then the diagonal elements of $\boldsymbol{\zeta}_N$ are either

$$\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\cos^2\big(w_it_k+\angle G(j\omega_i,\theta)\big),\quad 1\le i\le L$$

or

$$\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\sin^2\big(w_it_k+\angle G(j\omega_i,\theta)\big)\quad 1\le i\le L,$$

which tend to 1/2 via (35) and (36). On the other hand, the non-diagonal elements of ζ_N have the forms:

$$\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\cos(w_pt_k+\angle G(j\omega_p,\theta))\sin(w_qt_k+\angle G(j\omega_q,\theta)),\quad 1\le p,q\le L;$$

$$\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\sin(w_pt_k+\angle G(j\omega_p,\theta))\sin(w_qt_k+\angle G(j\omega_q,\theta)),\quad 1\le p,q\le L\quad\text{and}\quad p\ne q;$$

or

$$\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\cos(w_pt_k+\angle G(j\omega_p,\theta))\cos(w_qt_k+\angle G(j\omega_q,\theta)),\quad 1\le p,q\le L\quad\text{and}\quad p\ne q.$$

It follows that the non-diagonal elements of ζ_N approach to zeros via (33), (34), and the triangular formulas: $\cos(\alpha)\sin(\beta)=1/2(\sin(\alpha+\beta)-\sin(\alpha-\beta))$, $\sin(\alpha)\sin(\beta)=-1/2(\cos(\alpha+\beta)-\cos(\alpha-\beta))$, and $\cos(\alpha)\cos(\beta)=1/2(\cos(\alpha+\beta)+\cos(\alpha-\beta))$. As a result, we have $\zeta_N\xrightarrow[N\to\infty]{}\frac{1}{2}I$ w.p.1.

Noting that $t_k=\sum_{i=1}^{k}\tau_i$ and the sequence $\{\tau_k\}$ is i.i.d, we have $E\mu^{t_k}=E\mu^{\sum_{i=1}^{k}\tau_i}=E\prod_{i=1}^{k}\mu^{\tau_i}=(E\mu^{\tau_1})^k$. Since $E\mu^{\tau_1}=\int\mu^t p(t)\mathrm{d}t<1$ and $E\mu^{2\tau_1}/E\mu^{\tau_1}<1$ for any $0<\mu<1$, where $p(t)$ is the density function of τ_k, we have $E\left[\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\mu^{t_k}\right]^2=O\left(\frac{1}{N^2}\right)$ and

$$\sum_{N=1}^{\infty}P\left(\left|\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\mu^{t_k}\right|>\epsilon\right)=\sum_{N=1}^{\infty}O\left(\frac{1}{N^2}\right)<\infty$$

similar to the derivations in (38) and (39). By the Borel–Cantelli lemma (Chow & Teicher, 2003), we get $\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}\mu^{t_k}\xrightarrow[N\to\infty]{}0$. It follows that

$$\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}h(t_k,\theta)h'(t_k,\theta)=\widetilde{\mathcal{G}}(\theta)\zeta_N\widetilde{\mathcal{G}}'(\theta)+\frac{1}{N}\sum_{k=\overline{m}}^{\overline{m}+N-1}O(\mu^{t_k})\xrightarrow[N\to\infty]{}\frac{1}{2}\widetilde{\mathcal{G}}(\theta)\widetilde{\mathcal{G}}'(\theta)=M(\theta)\ \text{w.p.1.}\tag{41}$$

It is obtained that $M(\theta^*)$ is positive definite if the matrix $\widetilde{\mathcal{G}}(\theta^*)$ has full rank. Similarly, the limit in (19) also exists as that used in the proof above. □

References

Åström, K. J., & Bernhardsson, B. (1999). Comparison of periodic and event based sampling for first-order stochastic systems. In *Proceedings of the 14th IFAC world congress. Vol. 11* (pp. 301–306).

Åström, K. J., & Wittenmark, B. (1997). *Computer-controlled systems: theory and design* (3rd ed.). New Jersey: Prentice Hall.

Berenstein, C. A., & Gay, R. (1995). *Complex analysis and special topics in harmonic analysis. Vol. 262*. New York: Springer.

Chen, T., & Francis, B. A. (1995). *Optimal sampled-data control systems*. Springer.

Chow, Y. S., & Teicher, H. (2003). *Probability theory: independence, interchangeability, martingales* (3rd ed.). New York: Springer-Verlag.

Ding, F., Qiu, L., & Chen, T. (2009). Reconstruction of continuous-time systems from their non-uniformly sampled discrete-time systems. *Automatica*, *45*, 324–332.

Garnier, H., & Wang, L. (2008). *Identification of continuous-time models from sampled data*. Springer.

Gillberg, J., & Ljung, L. (2010). Frequency domain identification of continuous-time output error models, part II: Non-uniformly sampled data and b-spline output approximation. *Automatica*, *46*, 11–18.

Granas, A., & Dugundji, J. (2003). *Fixed point theory*. New York: Springer-Verlag.

Istratescu, V. I. (1981). *Fixed point theory: an introduction. Vol. 7*. Netherlands: Springer.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, *40*, 633–643.

Johansson, R. (2010). Continuous-time model identification and state estimation using non-uniformly sampled data. In *Proceedings of the 19th international symposium on mathematical theory of networks and systems–MTNS. Vol. 5*.

Kushner, H. J., & Yin, G. (2003). *Stochastic approximation algorithms and applications* (2nd ed.). New York: Springer-Verlag.

Larsson, E. K., Mossberg, M., & Soderstrom, T. (2007). Identification of continuous-time ARX models from irregularly sampled data. *IEEE Transactions on Automatic Control*, *52*, 417–427.

Larsson, E. K., & Söderström, T. (2002). Identification of continuous-time AR processes from unevenly sampled data. *Automatica*, *38*, 709–718.

Ljung, L. (1999). *System identification: theory for the user*. Englewood Cliffs: Prentice-Hall.

Ljung, L., & Vicino, A. (Eds.) (2005). *IEEE Transactions on Automatic Control*, *50*, Special Issue on Identification.

Marelli, D., & Fu, M. (2010). A continuous-time linear system identification method for slowly sampled data. *IEEE Transactions on Signal Processing*, *58*, 2521–2533.

Mas-Colell, A. (1979). Homeomorphisms of compact, convex sets and the Jacobian matrix. *SIAM Journal on Mathematical Analysis*, *10*, 1105–1109.

Mossberg, M. (2008). High-accuracy instrumental variable identification of continuous-time autoregressive processes from irregularly sampled noisy data. *IEEE Transactions on Signal Processing*, *56*, 4087–4091.

Persson, N., & Gustafsson, F. (2001). Event based sampling with application to vibration analysis in pneumatic tires. In *2001 IEEE International conference on acoustics, speech, and signal processing, 2001. proceedings. Vol. 6*, (ICASSP'01). (pp. 3885–3888). IEEE.

Phillips, C. L., & Nagle, H. T. (2007). *Digital control system analysis and design*. Prentice Hall Press.

Proakis, J. G., & Manolakis, D. G. (2007). *Digital signal processing*. Pearson Prentice Hall.

Rudin, W. (1973). *Functional analysis*. New York: McGraw-Hill.

Sanderberg, I. W. (1980). Global inverse function theorems. *IEEE Transactions on Circuits and Systems*, *27*(11), 998–1004.

Vajda, S., Valko, P., & Godfrey, K. R. (1987). Direct and indirect least squares methods in continuous-time parameter estimation. *Automatica*, *23*, 707–718.

Wang, L. Y., Feng, W., & Yin, G. G. (2013). Joint state and event observers for linear switching systems under irregular sampling. *Automatica*, *49*, 894–905.

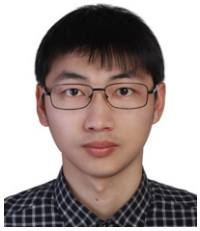
Wang, L. Y., Li, C., Yin, G. G., Guo, L., & Xu, C.-Z. (2011). State observability and observers of linear-time-invariant systems under irregular sampling and sensor limitations. *IEEE Transactions on Automatic Control*, *56*, 2639–2654.

Wang, L. Y., Yin, G. G., Zhang, J., & Zhao, Y. (2010). *System identification with quantized observations*. Boston: Birkhäuser.

Wu, F. F., & Desoer, C. A. (1972). Global inverse function theorem. *IEEE Transactions on Circuit Theory*, *19*, 199–201.

Yuz, J., Alfaro, J., Agüero, J. C., & Goodwin, G. C. (2011). Identification of continuous-time state-space models from non-uniform fast-sampled data. *IET Control Theory & Applications*, *5*, 842–855.

Zhu, Y., Telkamp, H., Wang, J., & Fu, Q. (2009). System identification using slow and irregular output samples. *Journal of Process Control*, *19*, 58–67.



Biqiang Mu was born in Sichuan, China, 1986. He received the Bachelor of Engineering degree in Material Formation and Control Engineering from Sichuan University in 2008, and the Ph.D. degree in Operations Research and Cybernetics from the Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences in 2013. He was a postdoctoral fellow at the Wayne State University from 2013 to 2014. He is currently a postdoctoral fellow at the University of Western Sydney. His research interests include recursive identification for Block-oriented nonlinear systems, identifiability of multivariable linear systems, identification of continuous-time linear systems, and variable selection for high-dimensional nonlinear systems.

Jin Guo received the B.S. degree in Mathematics from Shandong University, China, in 2008, and Ph.D. degree in System Modeling and Control Theory from Academy of Mathematics and Systems Science, Chinese Academy of Sciences in 2013. Now, he is a postdoctoral researcher in the School of Automation and Electrical Engineering, University of Science and Technology Beijing. His current research interests are identification and control of set-valued output systems and systems biology.

Le Yi Wang received the Ph.D. degree in Electrical Engineering from McGill University, Montreal, Canada, in 1990. Since 1990, he has been with Wayne State University, Detroit, Michigan, where he is currently a professor in the Department of Electrical and Computer Engineering. His research interests are in the areas of complexity and information, system identification, robust control, H-infinity optimization, time-varying systems, adaptive systems, hybrid and nonlinear systems, information processing and learning, as well as medical, automotive, communications, power systems, and computer applications of control methodologies. He was a keynote speaker in several international conferences. He serves on the IFAC Technical Committee on Modeling, Identification and Signal Processing. He was an Associate Editor of the IEEE Transactions on Automatic Control and several other journals, and currently is an Associate Editor of Journal of Control Theory and Applications. He was a Visiting Faculty at University of Michigan in 1996 and Visiting Faculty Fellow at University of Western Sydney in 2009 and 2013. He is a member of a Foreign Expert Team in Beijing Jiao Tong University and a member of the Core International Expert Group at Academy of Mathematics and Systems Science, Chinese Academy of Sciences. He is a Fellow of IEEE.

George Yin received the B.S. degree in Mathematics from the University of Delaware in 1983, M.S. degree in Electrical Engineering, and Ph.D. in Applied Mathematics from Brown University in 1987. He joined Wayne State University in 1987, and became a professor in 1996. His research interests include stochastic systems and applications. He severed on the IFAC Technical Committee on Modeling, Identification and Signal Processing, and many conference program committees; he was Co-Chair of SIAM Conference on Control & Its Application, 2011, and Co-Chair of a couple AMS-IMS-SIAM Summer Research Conferences; he also chaired a number SIAM prize selection committees. He is Chair SIAM Activity Group on Control and Systems Theory, and serves on the Board of Directors of American Automatic Control Council. He is an associate editor of SIAM Journal on Control and Optimization, and on the editorial board of a number of other journals. He was an Associate Editor of Automatica (2005–2011) and IEEE Transactions on Automatic Control (1994–1998). He is a Fellow of IEEE and a Fellow of SIAM.

Lijian Xu received the B.S. degree from the University of Science and Technology Beijing, China in 1998, M.S. degree from the University of Central Florida in 2001, Ph.D. degree from Wayne State University in 2014. He worked for AT&T, Florida, USA and Telus Communication, Calgary, Alberta, Canada as an engineer and engineering manager from 2001 to 2007. He is currently an Assistant Professor in the Department of Electrical and Computer Engineering Technology at Farmingdale State College, the State University of New York. His research interests are in the areas of networked control systems, digital wireless communications, consensus control, vehicle platoon and safety. He received the Best Paper Award from 2012 IEEE International Conference on Electro/Information Technology. He is a member of IEEE.

Wei Xing Zheng received the B.Sc. degree in Applied Mathematics in 1982, the M.Sc. degree in Electrical Engineering in 1984, and the Ph.D. degree in Electrical Engineering in 1989, all from Southeast University, Nanjing, China. He is currently a Professor at University of Western Sydney, Australia. Over the years he has also held various faculty/research/visiting positions at Southeast University, China, Imperial College of Science, Technology and Medicine, UK, University of Western Australia, Curtin University of Technology, Australia, Munich University of Technology, Germany, University of Virginia, USA, and University of California-Davis, USA. His research interests are in the areas of systems and controls, signal processing, and communications.

He is a Fellow of IEEE. Previously, he served as an Associate Editor for IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications, IEEE Transactions on Automatic Control, IEEE Signal Processing Letters, and IEEE Transactions on Circuits and Systems-II: Express Briefs, and as a Guest Editor for IEEE Transactions on Circuits and Systems-I: Regular Papers. Currently, he is an Associate Editor for Automatica, IEEE Transactions on Automatic Control (the second term), IEEE Transactions on Fuzzy Systems, IEEE Transactions on Cybernetics, IET Control Theory & Applications, and other scholarly journals. He is also an Associate Editor of IEEE Control Systems Society's Conference Editorial Board.