

Performance analysis of the generalized projection identification for time-varying systems

Feng Ding^{1*}, Ling Xu^{1,2}, Quanmin Zhu³

1. Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education),
School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, PR China

2. School of Internet of Things Technology, Wuxi Vocational Institute of Commerce, Wuxi 214153, PR China

3. Department of Engineering Design and Mathematics, Bristol BS16 1QY, University of the West of England

* Corresponding author: fding@jiangnan.edu.cn

February 23, 2016

Abstract: The least mean square methods include two typical parameter estimation algorithms, which are the projection algorithm and the stochastic gradient algorithm, the former is sensitive to noise and the latter is not capable of tracking the time-varying parameters. On the basis of these two typical algorithms, this paper presents a generalized projection identification algorithm for time-varying systems and studies its convergence by using the stochastic process theory. The analysis indicates that the generalized projection algorithm can track the time-varying parameters and requires less computational effort compared with the forgetting factor recursive least squares algorithm. The way of choosing the data window length is stated so that the minimum parameter estimation error upper bound can be obtained. The numerical examples are provided.

Keywords: Parameter estimation; recursive identification; time-varying system

1 Introduction

Establishing the mathematical models of things or systems is the main task of natural sciences. Mathematical models are very important in many areas such as controller design [1, 2], information filtering [3, 4], fault detection and diagnosis [5, 6], and state filtering and estimation [7–9]. System identification is the theory and methods of establishing the mathematical models of systems [10–12]. Parameter estimation methods are basic for system identification. Recently, Ding and Gu analyzed the performance analysis of the auxiliary model-based stochastic gradient parameter estimation algorithm for state space systems with one-step state delay [13]. This paper considers the identification algorithm and its performance analysis for time-varying systems [14, 15],

$$A(t, z)y(t) = B(t, z)u(t) + v(t), \quad (1)$$

where $\{u(t)\}$ and $\{y(t)\}$ are the input and output sequences of the system, respectively, $\{v(t)\}$ is a stochastic noise sequence with zero mean, and z^{-1} is a unit backward shift operator: $z^{-1}y(t) = y(t-1)$, $A(t, z)$ and $B(t, z)$ are time-varying coefficient polynomials in z^{-1} , and

$$A(t, z) := 1 + a_1(t)z^{-1} + a_2(t)z^{-2} + \dots + a_{n_a}(t)z^{-n_a},$$

* Corresponding author
E-mail: fding@jiangnan.edu.cn (F. Ding)

$$B(t, z) := b_1(t)z^{-1} + b_2(t)z^{-2} + \cdots + b_{n_b}(t)z^{-n_b}.$$

Define the time-varying parameter vector $\boldsymbol{\vartheta}(t-1) \in \mathbb{R}^n$ to be identified and the regressive information vector $\boldsymbol{\psi}(t) \in \mathbb{R}^n$ consisting of the observations up to and including time $(t-1)$,

$$\begin{aligned} \boldsymbol{\vartheta}(t-1) &:= [a_1(t), \cdots, a_{n_a}(t), b_1(t), \cdots, b_{n_b}(t)]^T \in \mathbb{R}^n, \\ \boldsymbol{\psi}(t) &:= [-y(t-1), -y(t-2), \cdots, -y(t-n_a), u(t-1), u(t-2), \cdots, u(t-n_b)]^T \in \mathbb{R}^n, \end{aligned}$$

where the superscript T denotes a vector transpose. Equation (1) can be written in vector form

$$y(t) = \boldsymbol{\psi}^T(t)\boldsymbol{\vartheta}(t-1) + v(t). \quad (2)$$

The forgetting factor recursive least squares (FF-RLS) algorithm is effective for estimating the time-varying parameter vector [16]. In this literature, Lozano [17], and Canetti and Espana [18] analyzed the performance of the FF-RLS algorithms for time-invariant and time-varying systems, respectively. Unfortunately, as the forgetting factor approaches unity, their results (i.e., the parameter estimation errors) goes to infinity even for time-invariant systems whose parameters are constant [15]. Bittanti et al studied the convergence properties of the directional FF-RLS algorithms for time-invariant deterministic systems [19]; for ergodic input-output data, Ljung and Prioret [20, 21] and Guo et al [22] obtained a parameter estimation error (PEE) upper bound like

$$\begin{aligned} \mathbb{E}[\|\hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}(t)\|^2] &\leq k_1(1-\lambda) \sup \mathbb{E}[v^2(t)] + \frac{k_2}{1-\lambda} \sup \mathbb{E}[\|\boldsymbol{w}(t)\|^2] \\ &\quad + O\left((1-\lambda)^{3/2} + c(1-\lambda)^{-1/2}\right) \end{aligned} \quad (3)$$

for large enough t , where $\hat{\boldsymbol{\vartheta}}(t)$ is the estimate of $\boldsymbol{\vartheta}(t)$, \mathbb{E} represents the expectation operator, $0 < \lambda < 1$ is the forgetting factor, k_1, k_2 and c are positive constants, $\boldsymbol{w}(t)$ is the parameter changing rate. However, for deterministic time-varying systems, i.e., the observation noise $v(t) \equiv 0$, as $\lambda \rightarrow 0$, the PEE upper bound [i.e., the expression on the right-hand side of (3)] is bounded. Unfortunately, this result is incompatible with the existing ones because as $\lambda \rightarrow 0$, the covariance matrix goes to infinity; it is impossible to obtain the bounded PEE. This motivates us to present a novel generalized stochastic gradient algorithm.

Although the forgetting factor recursive least squares algorithm can estimate the time-varying parameter vector $\boldsymbol{\vartheta}(t)$, its computational load is heavy due to computing the covariance matrix [15, 23]. From the perspective of decreasing computational complexity, the projection algorithm is sensitive to noise and the stochastic gradient (SG) algorithm is not capable of tracking the time-varying parameters [23]. On the basis of the work in [24], this paper combines the advantages of the projection algorithm and the SG algorithm to present a generalized projection identification algorithm for time-varying systems, and studies the convergence performance of the proposed algorithm by using the stochastic process theory. The generalized projection algorithm can track the time-varying parameters and requires less computational effort compared with the forgetting factor recursive least squares algorithm.

This paper is organized as follows. Section 2 gives several time-varying parameter estimation algorithms and derives the generalized projection algorithm. Section 3 provides several lemmas to prove the main convergence results in Section 4. Section 5 provides two numerical examples and summarizes some conclusions.

2 The generalized projection algorithm

Let us introduce some notations. Let \mathbf{I}_n be an identity matrix of order n , $\text{tr}[\mathbf{X}]$ denote the trace of a square matrix \mathbf{X} , and the norm of \mathbf{X} be $\|\mathbf{X}\| := \sqrt{\text{tr}[\mathbf{X}^T \mathbf{X}]} = \sqrt{\text{tr}[\mathbf{X} \mathbf{X}^T]}$ and $\lambda_{\max}[\mathbf{X}]$ represent the maximum eigenvalue of the positive definite matrix \mathbf{X} .

The following discusses several time-varying parameter estimation algorithms to real-time identify the parameter vector $\boldsymbol{\vartheta}(t)$ by using the input-output data, i.e., the observations $\{u(j), y(j), j \leq t\}$.

Using the Newton method and minimizing the cost function

$$J(\boldsymbol{\vartheta}(t)) := \sum_{j=1}^t [y(j) - \boldsymbol{\psi}^T(j) \boldsymbol{\vartheta}(t)]^2$$

give the following recursive algorithm of estimating the parameters $\boldsymbol{\vartheta}(t)$ [14]:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \mathbf{R}^{-1}(t) \boldsymbol{\psi}(t) [y(t) - \boldsymbol{\psi}^T(t) \hat{\boldsymbol{\vartheta}}(t-1)], \quad (4)$$

where $\hat{\boldsymbol{\vartheta}}(t)$ is the estimate of $\boldsymbol{\vartheta}(t)$ at time t . The difference of the matrix $\mathbf{R}(t) \in \mathbb{R}^{n \times n}$ will lead to different identification algorithms, e.g., the forgetting factor recursive least squares algorithm, the projection algorithm, the finite data window least squares algorithm, the stochastic gradient algorithm, and so on.

1. If we take $\mathbf{R}(t) := \mathbf{P}^{-1}(t)$, and

$$\mathbf{P}^{-1}(t) := \lambda \mathbf{P}^{-1}(t-1) + \boldsymbol{\psi}(t) \boldsymbol{\psi}^T(t), \quad 0 < \lambda < 1, \quad (5)$$

then Equations (4) and (5) form the FF-RLS algorithm [15]:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \mathbf{P}(t) \boldsymbol{\psi}(t) [y(t) - \boldsymbol{\psi}^T(t) \hat{\boldsymbol{\vartheta}}(t-1)], \quad (6)$$

$$\mathbf{P}^{-1}(t) = \lambda \mathbf{P}^{-1}(t-1) + \boldsymbol{\psi}(t) \boldsymbol{\psi}^T(t), \quad 0 < \lambda < 1. \quad (7)$$

As the forgetting factor λ goes to unity, the FF-RLS algorithm in (6) and (7) reduces to the recursive least squares (RLS) algorithm:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \mathbf{P}(t) \boldsymbol{\psi}(t) [y(t) - \boldsymbol{\psi}^T(t) \hat{\boldsymbol{\vartheta}}(t-1)], \quad (8)$$

$$\mathbf{P}^{-1}(t) = \mathbf{P}^{-1}(t-1) + \boldsymbol{\psi}(t) \boldsymbol{\psi}^T(t), \quad \mathbf{P}(0) = p_0 \mathbf{I}_n. \quad (9)$$

The number p_0 should be large enough if the regression variables take very small values.

2. If we take $\mathbf{R}(t) := \mathbf{P}^{-1}(t)$ and

$$\mathbf{P}^{-1}(t) = \sum_{i=0}^{q-1} \boldsymbol{\psi}(t-i) \boldsymbol{\psi}^T(t-i) \quad (10)$$

$$= \mathbf{P}^{-1}(t-1) + \boldsymbol{\psi}(t) \boldsymbol{\psi}^T(t) - \boldsymbol{\psi}(t-q) \boldsymbol{\psi}^T(t-q), \quad (11)$$

then Equations (4) and (11) form the finite data window recursive least squares (FDW-RLS) algorithm:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \mathbf{P}(t) \boldsymbol{\psi}(t) [y(t) - \boldsymbol{\psi}^T(t) \hat{\boldsymbol{\vartheta}}(t-1)], \quad (12)$$

$$\mathbf{P}^{-1}(t) = \mathbf{P}^{-1}(t-1) + \boldsymbol{\psi}(t) \boldsymbol{\psi}^T(t) - \boldsymbol{\psi}(t-q) \boldsymbol{\psi}^T(t-q), \quad \mathbf{P}(0) = p_0 \mathbf{I}_n, \quad (13)$$

where q is the length of the data window.

3. If we take $\mathbf{R}(t) := r(t)\mathbf{I}_n$ and the trace of both sides of Equation (7) and define

$$r(t) := \text{tr}[\mathbf{P}^{-1}(t)] = r(t-1) + \|\boldsymbol{\psi}(t)\|^2, \quad (14)$$

then Equations (4) and (14) form the forgetting factor stochastic gradient (FFSG) algorithm (the forgetting gradient algorithm for short):

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\boldsymbol{\psi}(t)}{r(t)}[y(t) - \boldsymbol{\psi}^\top(t)\hat{\boldsymbol{\vartheta}}(t-1)], \quad (15)$$

$$r(t) = \lambda r(t-1) + \|\boldsymbol{\psi}(t)\|^2, \quad 0 < \lambda < 1, \quad r(0) = 1. \quad (16)$$

As the forgetting factor λ goes to unity, the algorithm in (15)–(16) becomes the following stochastic gradient (SG) algorithm [23]:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\boldsymbol{\psi}(t)}{r(t)}[y(t) - \boldsymbol{\psi}^\top(t)\hat{\boldsymbol{\vartheta}}(t-1)], \quad (17)$$

$$r(t) = r(t-1) + \|\boldsymbol{\psi}(t)\|^2, \quad r(0) = 1. \quad (18)$$

Recently, a multi-innovation stochastic gradient algorithm was proposed to track time-varying parameters for linear regression models [25].

4. If we take $\lambda = 0$ in (16), then the forgetting factor stochastic gradient algorithm in (15)–(16) becomes the projection (PJ) identification algorithm:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\boldsymbol{\psi}(t)}{\|\boldsymbol{\psi}(t)\|^2}[y(t) - \boldsymbol{\psi}^\top(t)\hat{\boldsymbol{\vartheta}}(t-1)]. \quad (19)$$

5. If we take $\mathbf{R}(t) := r(t)\mathbf{I}_n$ and the trace of both sides of Equation (10) and define

$$r(t) := \text{tr}[\mathbf{P}^{-1}(t)] = \sum_{i=0}^{q-1} \|\boldsymbol{\psi}(t-i)\|^2, \quad (20)$$

then Equations (4) and (20) form the generalized projection (GPJ) identification algorithm:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\boldsymbol{\psi}(t)}{r(t)}[y(t) - \boldsymbol{\psi}^\top(t)\hat{\boldsymbol{\vartheta}}(t-1)], \quad (21)$$

$$r(t) = r(t-1) + \|\boldsymbol{\psi}(t)\|^2 - \|\boldsymbol{\psi}(t-q)\|^2, \quad r(0) = 1. \quad (22)$$

As the data window length $q = 1$, the GPJ algorithm is the projection algorithm in (19); as we take $q = t$, the GPJ algorithm becomes the stochastic gradient algorithm in (17)–(18).

3 The basic lemmas

Define the transition matrix

$$\boldsymbol{\Phi}(t+1, j) = [\mathbf{I}_n - \boldsymbol{\psi}(t)\boldsymbol{\psi}^\top(t)/r(t)] \boldsymbol{\Phi}(t, j), \quad \boldsymbol{\Phi}(j, j) = \mathbf{I}_n$$

and the maximum eigenvalue

$$\rho(t) := \lambda_{\max}[\boldsymbol{\Phi}^\top(t+s, t)\boldsymbol{\Phi}(t+s, t)].$$

It follows that

$$\boldsymbol{\Phi}(t+1, t) = \mathbf{I}_n - \boldsymbol{\psi}(t)\boldsymbol{\psi}^\top(t)/r(t) \in \mathbb{R}^{n \times n}.$$

Lemma 1 For the system in (2) and the GPJ algorithm in (21)–(22), assume that there exist positive constants α and β and an integer $s \geq n$ such that the following strong persistent excitation (SPE) condition holds,

$$(SPE) \quad \alpha \mathbf{I}_n \leq \frac{1}{s} \sum_{j=0}^{s-1} \boldsymbol{\psi}(t+j) \boldsymbol{\psi}^T(t+j) \leq \beta \mathbf{I}_n, \text{ a.s., } t > 0.$$

Then the maximum eigenvalue $\rho(t)$ satisfies

$$\rho(t) \leq 1 - \alpha^2 \{(q+s)(n\beta)^2(s+1)^2\}^{-1} =: \rho < 1, \text{ a.s.}$$

Proof Here, we refer to the way in [25] for proving this lemma. Let $\boldsymbol{\zeta}_0 \in \mathbb{R}^n$ be the unit eigenvector corresponding to the maximum eigenvalue $\rho(t)$ of the matrix $\boldsymbol{\Phi}^T(t+s, t) \boldsymbol{\Phi}(t+s, t)$, i.e., $\boldsymbol{\Phi}^T(t+s, t) \boldsymbol{\Phi}(t+s, t) \boldsymbol{\zeta}_0 = \rho(t) \boldsymbol{\zeta}_0$. Use the transition matrix $\boldsymbol{\Phi}(t+1, j)$ to construct the difference equation,

$$\boldsymbol{\xi}(j+1) = \boldsymbol{\Phi}(j+1, j) \boldsymbol{\xi}(j), \quad \boldsymbol{\xi}_t = \boldsymbol{\zeta}_0. \quad (23)$$

Using the relation $\boldsymbol{\Phi}(t, i) \boldsymbol{\Phi}(i, j) = \boldsymbol{\Phi}(t, j)$, we have

$$\boldsymbol{\xi}(t+s) = \boldsymbol{\Phi}(t+s, t) \boldsymbol{\xi}(t) = \boldsymbol{\Phi}(t+s, t) \boldsymbol{\zeta}_0.$$

Taking the norm to both sides and using the definition of the eigenvalue, we have

$$\|\boldsymbol{\xi}(t+s)\|^2 = \boldsymbol{\zeta}_0^T \boldsymbol{\Phi}^T(t+s, t) \boldsymbol{\Phi}(t+s, t) \boldsymbol{\zeta}_0 = \boldsymbol{\zeta}_0^T \rho(t) \boldsymbol{\zeta}_0 = \rho(t).$$

Taking the norm to both sides of (23) gives

$$\begin{aligned} \|\boldsymbol{\xi}(j+1)\|^2 &= \boldsymbol{\xi}^T(j+1) \boldsymbol{\xi}(j+1) \\ &= \boldsymbol{\xi}^T(j) \boldsymbol{\Phi}^T(j+1, j) \boldsymbol{\Phi}(j+1, j) \boldsymbol{\xi}(j) \\ &= \boldsymbol{\xi}^T(j) [\mathbf{I}_n - \boldsymbol{\psi}(j) \boldsymbol{\psi}^T(j)/r(j)]^2 \boldsymbol{\xi}(j) \\ &\leq \boldsymbol{\xi}^T(j) [\mathbf{I}_n - \boldsymbol{\psi}(j) \boldsymbol{\psi}^T(j)/r(j)] \boldsymbol{\xi}(j) \\ &= \boldsymbol{\xi}^T(j) \boldsymbol{\xi}(j) - [\boldsymbol{\psi}^T(j) \boldsymbol{\xi}(j)]^2 / r(j) \\ &= \|\boldsymbol{\xi}(j)\|^2 - [\boldsymbol{\psi}^T(j) \boldsymbol{\xi}(j)]^2 / r(j). \end{aligned}$$

Thus we have

$$\|\boldsymbol{\psi}^T(j) \boldsymbol{\xi}(j)\|^2 / r(j) \leq \|\boldsymbol{\xi}(j)\|^2 - \|\boldsymbol{\xi}(j+1)\|^2.$$

Replacing $t+j$ with j gives

$$\|\boldsymbol{\psi}^T(t+j) \boldsymbol{\xi}(t+j)\|^2 / r(t+j) \leq \|\boldsymbol{\xi}(t+j)\|^2 - \|\boldsymbol{\xi}(t+j+1)\|^2.$$

Summing for j from $j=0$ to $j=s-1$ gives

$$\begin{aligned} \sum_{j=0}^{s-1} \|\boldsymbol{\psi}^T(t+j) \boldsymbol{\xi}(t+j)\|^2 / r(t+j) &\leq \sum_{j=0}^{s-1} \|\boldsymbol{\xi}(t+j)\|^2 - \|\boldsymbol{\xi}(t+j+1)\|^2 \\ &= \|\boldsymbol{\xi}(t)\|^2 - \|\boldsymbol{\xi}(t+s)\|^2 = 1 - \rho(t). \end{aligned} \quad (24)$$

From (23), we have

$$\begin{aligned}
\xi(t+1) &= [\mathbf{I}_n - \boldsymbol{\psi}(t)\boldsymbol{\psi}^\top(t)/r(t)]\xi(t) \\
&= \xi(t) - \boldsymbol{\psi}(t)\boldsymbol{\psi}^\top(t)/r(t)\xi(t), \\
\xi(t+2) &= \xi(t+1) - \boldsymbol{\psi}(t+1)\boldsymbol{\psi}^\top(t+1)/r(t+1)\xi(t+1) \\
&= \xi(t) - \boldsymbol{\psi}(t)\boldsymbol{\psi}^\top(t)/r(t)\xi(t) - \boldsymbol{\psi}(t+1)\boldsymbol{\psi}^\top(t+1)/r(t+1)\xi(t+1), \\
\xi(t+3) &= \xi(t+1) - \boldsymbol{\psi}(t+1)\boldsymbol{\psi}^\top(t+1)/r(t+1)\xi(t+1) - \boldsymbol{\psi}(t+2)\boldsymbol{\psi}^\top(t+2)/r(t+2)\xi(t+2) \\
&= \xi(t) - \boldsymbol{\psi}(t)\boldsymbol{\psi}^\top(t)/r(t)\xi(t) - \boldsymbol{\psi}(t+1)\boldsymbol{\psi}^\top(t+1)/r(t+1)\xi(t+1) - \boldsymbol{\psi}(t+2)\boldsymbol{\psi}^\top(t+2)/r(t+2)\xi(t+2) \\
&= \zeta_0 - \sum_{j=0}^2 \boldsymbol{\psi}(t+j)\boldsymbol{\psi}^\top(t+j)/r(t+j)\xi(t+j), \\
&\vdots \\
\xi(t+i) &= \zeta_0 - \sum_{j=0}^{i-1} \boldsymbol{\psi}(t+j)\boldsymbol{\psi}^\top(t+j)/r(t+j)\xi(t+j).
\end{aligned}$$

That is

$$\xi(t+i) = \zeta_0 - \sum_{j=0}^{i-1} \boldsymbol{\psi}(t+j)\boldsymbol{\psi}^\top(t+j)/r(t+j)\xi(t+j).$$

Taking the norm to both sides and using the inequality $(\sum a_i b_i)^2 \leq (\sum a_i^2)(\sum b_i^2)$, $\|\boldsymbol{\psi}(t+j)\|^2/r(t+j) \leq 1$ and (24), we have

$$\begin{aligned}
\|\xi_{t+i} - \zeta_0\|^2 &= \left\| \sum_{j=0}^{i-1} \boldsymbol{\psi}(t+j)\boldsymbol{\psi}^\top(t+j)/r(t+j)\xi(t+j) \right\|^2 \\
&\leq \left[\sum_{j=0}^{i-1} \|\boldsymbol{\psi}(t+j)\|^2/r(t+j) \right] \left[\sum_{j=0}^{i-1} \|\boldsymbol{\psi}^\top(t+j)\xi_{t+j}\|^2/r(t+j) \right] \\
&\leq i[1 - \rho(t)] \leq s[1 - \rho(t)], \quad i \leq s-1.
\end{aligned} \tag{25}$$

Taking the trace of the SPE condition gives $\|\boldsymbol{\psi}(t)\|^2 \leq \delta_1 := ns\beta$, a.s. $r(t)$ in (22) satisfies [26]

$$(q-s+1)n\alpha \leq r(t) \leq (q+s)n\beta + 1, \quad \text{a.s., } q \geq s, \tag{26}$$

Pre-multiplying and post-multiplying the SPE condition by ζ_0^\top and ζ_0 , respectively, using (24), (25) and (26), we have

$$\begin{aligned}
\alpha s &\leq \zeta_0^\top \sum_{i=0}^{s-1} \boldsymbol{\psi}(t+i)\boldsymbol{\psi}^\top(t+i)\zeta_0 \\
&\leq \sqrt{(q+s)n\beta + 1} \zeta_0^\top \sum_{i=0}^{s-1} \boldsymbol{\psi}(t+i)\boldsymbol{\psi}^\top(t+i)/\sqrt{r(t+i)} \zeta_0 \\
&\leq \sqrt{(q+s)n\beta + 1} \left\| \sum_{i=0}^{s-1} \boldsymbol{\psi}(t+i)\boldsymbol{\psi}^\top(t+i)/\sqrt{r(t+i)} [\zeta_0 - \xi(t+i) + \xi(t+i)] \right\| \\
&\leq \sqrt{(q+s)n\beta + 1} \left\| \sum_{i=0}^{s-1} \boldsymbol{\psi}(t+i)\boldsymbol{\psi}^\top(t+i)/\sqrt{r(t+i)} [\zeta_0 - \xi(t+i)] \right\| \\
&\quad + \sqrt{(q+s)n\beta + 1} \left\| \sum_{i=0}^{s-1} \boldsymbol{\psi}(t+i)\boldsymbol{\psi}^\top(t+i)\xi(t+i)/\sqrt{r(t+i)} \right\|
\end{aligned}$$

$$\begin{aligned}
&\leq \sqrt{(q+s)n\beta+1} \sum_{i=0}^{s-1} \sqrt{\delta_1} \|\boldsymbol{\xi}(t+i) - \boldsymbol{\zeta}_0\| \\
&\quad + \sqrt{(q+s)n\beta+1} \left[\sum_{i=0}^{s-1} \|\boldsymbol{\psi}(t+i)\|^2 \right]^{1/2} \left[\sum_{i=0}^{s-1} \|\boldsymbol{\psi}^\top(t+i)\boldsymbol{\xi}(t+i)\|^2 / r(t+i) \right]^{1/2} \\
&\leq \sqrt{(q+s)n\beta+1} [s\sqrt{\delta_1}\sqrt{s(1-\rho(t))} + \sqrt{s\delta_1}\sqrt{1-\rho(t)}] \\
&= \sqrt{(q+s)n\beta+1} (s+1)\sqrt{s\delta_1}\sqrt{1-\rho(t)}, \text{ a.s.}
\end{aligned}$$

Solving $\rho(t)$ from this inequality gives the result of Lemma 1. \square

Lemma 2 For the system in (2) and the SG algorithm in (17)–(18), if the conditions in Lemma 1 hold, then the following inequality holds,

$$\rho(t) \leq 1 - \alpha^2 \{n(s+1)^2 \beta [n(t+2s-1)\beta + 1]\}^{-2}, \text{ a.s., } t > 0.$$

Proof Using the SPE condition and referring to the method in [25, 26], we can conclude that $r(t)$ in (16) satisfies

$$(t-s+1)n\alpha + 1 \leq r(t) \leq (t+s)n\beta + 1, \text{ a.s.} \quad (27)$$

Since $r(t)$ in (18) is nondecreasing, using (27), a similar derivation of Lemma 1 and referring to the way in [25], we have

$$\begin{aligned}
\alpha s &\leq \boldsymbol{\zeta}_0^\top \sum_{i=0}^{s-1} \boldsymbol{\psi}(t+i)\boldsymbol{\psi}^\top(t+i)\boldsymbol{\zeta}_0 \\
&\leq \sqrt{r(t+s-1)} \boldsymbol{\zeta}_0^\top \sum_{i=0}^{s-1} \boldsymbol{\psi}(t+i)\boldsymbol{\psi}^\top(t+i) / \sqrt{r(t+i)} \boldsymbol{\zeta}_0 \\
&\leq \sqrt{n(t+2s-1)\beta+1} \{s\sqrt{\delta_1}\sqrt{s[1-\rho(t)]} + \sqrt{s\delta_1}\sqrt{1-\rho(t)}\} \\
&= \sqrt{n(t+2s-1)\beta+1} (s+1)\sqrt{s\delta_1[1-\rho(t)]}, \text{ a.s.}
\end{aligned}$$

Solving $\rho(t)$ from this inequality gives the results of Lemma 2.

Lemma 3 [27] Let the nonnegative sequences $\{x(t)\}$, $\{a_t\}$ and $\{b_t\}$ satisfy $x(t+1) \leq (1-a_t)x(t) + b_t$, $t \geq 0$ and $a_t \in [0, 1)$ or $0 < 1-a_t \leq 1$, $\sum_{t=1}^{\infty} a_t = \infty$, $x(0) < \infty$, then $\lim_{t \rightarrow \infty} x(t) \leq \lim_{t \rightarrow \infty} b_t/a_t$, where it is assumed that the above limits exist.

The proof is easy and omitted here.

4 The main results

Theorem 1 For the system in (2) and the algorithm in (21)–(22), assume that the observation noise $\{v(t)\}$ and the parameter changing rate $\{\boldsymbol{w}(t) := \boldsymbol{\vartheta}(t) - \boldsymbol{\vartheta}(t-1)\}$ are uncorrelated random variable sequences with zero mean and satisfy

$$(A1) \quad \mathbb{E}[v(t)] = 0, \quad \mathbb{E}[\boldsymbol{w}(t)] = \mathbf{0}, \quad \mathbb{E}[v(t)\boldsymbol{w}(i)] = \mathbf{0},$$

$$(A2) \quad \mathbb{E}[v(t)v(i)] = 0, \quad \mathbb{E}[\boldsymbol{w}(t)\boldsymbol{w}^\top(i)] = \mathbf{0}, \quad i \neq t,$$

$$(A3) \quad \begin{aligned} \mathbb{E}[v^2(t)] &= \sigma_v^2(t) \leq \sigma_v^2 < \infty, \\ \mathbb{E}[\|\mathbf{w}(t)\|^2] &= \sigma_w^2(t) \leq \sigma_w^2 < \infty. \end{aligned}$$

If the conditions in Lemma 1 hold, then the parameter estimation error vector $\hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}(t)$ given by the generalized projection algorithm is mean square bounded, i.e.,

$$\mathbb{E}[\|\hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}(t)\|^2] \leq \rho^{\lfloor \frac{t}{s} \rfloor} \mathbb{E}[\|\hat{\boldsymbol{\vartheta}}(0) - \boldsymbol{\vartheta}(0)\|^2] + k_1 \frac{(q+s)\sigma_v^2}{(q-s+1)^2} + k_2(q+s)\sigma_w^2,$$

where $k_1 = 2ns^3(s+1)^2\beta^3/\alpha^4$, $k_2 = 2n^2s^2(s+1)^2\beta^2/\alpha^2$.

Proof Define the parameter estimation error vector $\tilde{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}(t)$. Assume that $\tilde{\boldsymbol{\vartheta}}(0)$ and $\{v(t)\}$ are uncorrelated, and $\mathbb{E}[\|\tilde{\boldsymbol{\vartheta}}(0)\|^2] < \infty$. Subtracting $\boldsymbol{\vartheta}(t)$ from both sides of (21), we have

$$\begin{aligned} \tilde{\boldsymbol{\vartheta}}(t) &= \hat{\boldsymbol{\vartheta}}(t) - [\boldsymbol{\vartheta}(t-1) + \mathbf{w}(t)] \\ &= \tilde{\boldsymbol{\vartheta}}(t-1) + \frac{\boldsymbol{\psi}(t)}{r(t)} [-\boldsymbol{\psi}^\top(t)\tilde{\boldsymbol{\vartheta}}(t-1) + v(t)] - \mathbf{w}(t) \\ &= \boldsymbol{\Phi}(t+1, t)\tilde{\boldsymbol{\vartheta}}(t-1) + \frac{\boldsymbol{\psi}(t)}{r(t)}v(t) - \mathbf{w}(t) \\ &= \boldsymbol{\Phi}(t+1, t-s+1)\tilde{\boldsymbol{\vartheta}}(t-s) \\ &\quad + \sum_{i=0}^{s-1} \boldsymbol{\Phi}(t+1, t-i+1) \left[\frac{\boldsymbol{\psi}(t-i)}{r(t-i)}v(t-i) - \mathbf{w}(t-i) \right]. \end{aligned}$$

Furthermore, in order to prove that the parameter estimation error is mean squared bounded, taking the norm to both sides of the above equation and using Lemma 1, we have

$$\begin{aligned} \|\tilde{\boldsymbol{\vartheta}}(t)\|^2 &= \tilde{\boldsymbol{\vartheta}}^\top(t-s)\boldsymbol{\Phi}^\top(t+1, t-s+1)\boldsymbol{\Phi}(t+1, t-s+1)\tilde{\boldsymbol{\vartheta}}(t-s) \\ &\quad + 2\tilde{\boldsymbol{\vartheta}}^\top(t-s)\boldsymbol{\Phi}^\top(t+1, t-s+1) \sum_{i=0}^{s-1} \boldsymbol{\Phi}(t+1, t-i+1) \left[\frac{\boldsymbol{\psi}(t-i)}{r(t-i)}v(t-i) - \mathbf{w}(t-i) \right] \\ &\quad + \left\| \sum_{i=0}^{s-1} \boldsymbol{\Phi}(t+1, t-i+1) \left[\frac{\boldsymbol{\psi}(t-i)}{r(t-i)}v(t-i) - \mathbf{w}(t-i) \right] \right\|^2 \\ &\leq \rho(t)\|\tilde{\boldsymbol{\vartheta}}(t-s)\|^2 + 2\tilde{\boldsymbol{\vartheta}}^\top(t-s)\boldsymbol{\Phi}^\top(t+1, t-s+1) \\ &\quad \times \sum_{i=0}^{s-1} \boldsymbol{\Phi}(t+1, t-i+1) \left[\frac{\boldsymbol{\psi}(t-i)}{r(t-i)}v(t-i) - \mathbf{w}(t-i) \right] \\ &\quad + s \sum_{i=0}^{s-1} \left\| \boldsymbol{\Phi}(t+1, t-i+1) \left[\frac{\boldsymbol{\psi}(t-i)}{r(t-i)}v(t-i) - \mathbf{w}(t-i) \right] \right\|^2. \end{aligned}$$

Here, we have used the relation $(a_1 + a_2 + \dots + a_n)^2 \leq n(a_1^2 + a_2^2 + \dots + a_n^2)$. Note that the maximum eigenvalue of $\boldsymbol{\Phi}^\top(t+1, t-i+1)\boldsymbol{\Phi}(t+1, t-i+1)$ is not more than unity for any $i \geq 1$. Using (A1)–(A3) and taking the expectation of both sides of the above equation and using Lemma 1 give

$$\begin{aligned} \mathbb{E}[\|\tilde{\boldsymbol{\vartheta}}(t)\|^2] &\leq \rho \mathbb{E}[\|\tilde{\boldsymbol{\vartheta}}(t-s)\|^2] + 0 + s \sum_{i=0}^{s-1} \mathbb{E} \left\{ \left\| \boldsymbol{\Phi}(t+1, t-i+1) \left[\frac{\boldsymbol{\psi}(t-i)}{r(t-i)}v(t-i) - \mathbf{w}(t-i) \right] \right\|^2 \right\} \\ &\leq \rho \mathbb{E}[\|\tilde{\boldsymbol{\vartheta}}(t-s)\|^2] + 2s \sum_{i=0}^{s-1} \left\{ \delta_1 \mathbb{E} \left[\frac{v^2(t-i)}{r^2(t-i)} \right] + \sigma_w^2 \right\} \\ &\leq \rho \mathbb{E}[\|\tilde{\boldsymbol{\vartheta}}(t-s)\|^2] + 2s \sum_{i=0}^{s-1} \left[\frac{\delta_1 \sigma_v^2}{(q-s+1)^2(n\alpha)^2} + \sigma_w^2 \right] \end{aligned}$$

$$\leq \rho \mathbb{E}[\|\tilde{\boldsymbol{\vartheta}}(t-s)\|^2] + 2s^2 \left[\frac{\delta_1 \sigma_v^2}{(q-s+1)^2 (n\alpha)^2} + \sigma_w^2 \right].$$

Let $t = si + k$, $0 \leq k \leq s-1$, we have

$$\begin{aligned} \mathbb{E}[\|\tilde{\boldsymbol{\vartheta}}(t)\|^2] &= \mathbb{E}[\|\tilde{\boldsymbol{\vartheta}}(si+k)\|^2] \leq \rho T(s(i-1)+k) + 2s^2 \left[\frac{\delta_1 \sigma_v^2}{(q-s+1)^2 (n\alpha)^2} + \sigma_w^2 \right] \\ &\leq \rho^i \mathbb{E}[\|\tilde{\boldsymbol{\vartheta}}(k)\|^2] + \frac{2s^2}{1-\rho} \left[\frac{\delta_1 \sigma_v^2}{(q-s+1)^2 (n\alpha)^2} + \sigma_w^2 \right] \\ &\leq \rho^i \mathbb{E}[\|\tilde{\boldsymbol{\vartheta}}(k)\|^2] + \frac{2s^2(q+s)(n\beta)^2(s+1)^2}{\alpha^2} \left[\frac{\delta_1 \sigma_v^2}{(q-s+1)^2 (n\alpha)^2} + \sigma_w^2 \right] \\ &\leq \rho^{\lfloor \frac{t}{s} \rfloor} \mathbb{E}[\|\tilde{\boldsymbol{\vartheta}}(k)\|^2] + \frac{2s^2(s+1)^2(n\beta)^2}{\alpha^2} \left[\frac{s\beta(q+s)\sigma_v^2}{(q-s+1)^2 n\alpha^2} + (q+s)\sigma_w^2 \right]. \end{aligned}$$

Taking the limit and using Lemma 3 lead to

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbb{E}[\|\tilde{\boldsymbol{\vartheta}}(t)\|^2] &= \lim_{i \rightarrow \infty} \mathbb{E}[\|\tilde{\boldsymbol{\vartheta}}(si+k)\|^2] \\ &\leq \frac{2s^2(q+s)(n\beta)^2(s+1)^2}{\alpha^2} \left[\frac{\delta_1 \sigma_v^2}{(q-s+1)^2 (n\alpha)^2} + \sigma_w^2 \right] \\ &\leq \frac{2s^2(s+1)^2(n\beta)^2}{\alpha^2} \left[\frac{s\beta(q+s)\sigma_v^2}{(q-s+1)^2 n\alpha^2} + (q+s)\sigma_w^2 \right]. \end{aligned}$$

This completes the proof of Theorem 1.

Theorem 1 shows that the larger α is and/or the smaller β is, the smaller the parameter estimation error is that is the stationarity of the input-output data can improve the identification accuracy.

From Theorem 1, we can see that the identification algorithms encounter difficulties for fast-changing-parameter systems because the fast-changing parameters have large variance σ_w^2 and the parameter estimation error upper bound become large. Otherwise, the slowly time-varying parameters and small observation noise variance lead to a small parameter estimation error.

Theorem 1 gives the PEE upper bound. The following studies how to obtain the minimum estimation error upper bounds and the data window length which leads to the minimum PEE upper bound. From Theorem 1, we have the following corollaries.

Corollary 1 For the time invariant stochastic systems $y(t) = \boldsymbol{\psi}^T(t)\boldsymbol{\vartheta} + v(t)$, we have

$$\limsup_{t \rightarrow \infty} \mathbb{E}[\|\hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}\|^2] \leq k_1 \frac{(q+s)\sigma_v^2}{(q-s+1)^2}.$$

Let $q = t$ in (22), i.e., $r(t) = r(t-1) + \|\boldsymbol{\psi}(t)\|^2$, we have

$$\lim_{t \rightarrow \infty} \mathbb{E}[\|\hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}\|^2] \leq \lim_{t \rightarrow \infty} k_1 \frac{(t+s)\sigma_v^2}{(t-s+1)^2} = 0.$$

Thus, for the time invariant stochastic systems, the parameter estimates given by the stochastic gradient algorithm converges to the true parameters – see Theorem 2.

Corollary 2 For the deterministic time-varying systems $y(t) = \boldsymbol{\psi}^T(t)\boldsymbol{\vartheta}(t-1)$, we have

$$\limsup_{t \rightarrow \infty} \mathbb{E}[\|\hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}(t)\|^2] \leq k_2(q+s)\sigma_w^2 =: f_2(q), \quad q \geq s.$$

As $q = s$, $f_2(q) = 2k_2s\sigma_w^2 = \min$, i.e., the PEE upper bound given by the projection algorithm is minimal.

Corollary 3 For the time-varying stochastic systems in (2), letting $f_1'(q) = 0$ in Theorem 1 gives

$$q^3 + 3(1-s)q^2 + \left[3(1-s)^3 - \frac{k_1\sigma_v^2}{k_2\sigma_w^2}\right]q + \left[(1-s)^3 + (1-3s)\frac{k_1\sigma_v^2}{k_2\sigma_w^2}\right] = 0.$$

This equation has three solutions, q_1 , q_2 and q_3 , and the solution $q = [q_i]$ or $q = [q_i] + 1$ which leads to $f(q_i) = \min$ is the best data window length, $[x]$ represents the maximum integer no more than x , and the corresponding minimal estimation error upper bound is $f([q_i])$ or $f([q_i] + 1)$.

Remark x For practical identification problem, the task first is collecting the input-output data $\{u(t), y(t)\}$ and uses them to construct the information vector $\boldsymbol{\psi}(t)$, and then

Remark x A longer window permits better performance of the GPJ algorithm for slowly time varying systems. However, if the parameters of system is changing fast, the window cannot be too large. These three corollaries can give the practical guidance for users on how to choose the window length. Also, these three corollaries may be used for evaluating the PEE upper bound given by the generalized projection algorithm for time-varying systems and may guide the choice of the data window length so as to obtain the minimal parameter estimation errors.

Theorem 2 For time invariant stochastic systems in (2), if the conditions in Theorem 1 hold, then the estimation error, $E[\|\hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}\|^2]$, given by the stochastic gradient algorithm in (17)–(18) converges to zero at the rate of $O(\frac{1}{t})$.

Proof A similar derivation of Theorem 2 leads to

$$E[\|\hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}\|^2] \leq \rho(t-s+1)E[\|\hat{\boldsymbol{\vartheta}}(t-s) - \boldsymbol{\vartheta}\|^2] + s \sum_{i=0}^{s-1} E\left[\left\|\Phi(t+1, t-i+1) \frac{\boldsymbol{\psi}(t-i)}{r(t-i)} v(t-i)\right\|^2\right].$$

Using (27), we have

$$\begin{aligned} E[\|\hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}\|^2] &\leq \rho(t-s+1)E[\|\hat{\boldsymbol{\vartheta}}(t-s) - \boldsymbol{\vartheta}\|^2] + s \sum_{i=0}^{s-1} \frac{\delta_1\sigma_v^2}{[(t-s+1-i)n\alpha+1]^2} \\ &\leq \left(1 - \frac{\alpha^2}{n(s+1)^2\beta[n(t+s)\beta+1]}\right)T(t-s) + \frac{ns^3\beta\sigma_v^2}{[(t-2s+2)n\alpha+1]^2}. \end{aligned}$$

Using Lemma 3, it is not difficult to obtain

$$\lim_{t \rightarrow \infty} E[\|\hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}\|^2] \leq \lim_{t \rightarrow \infty} \frac{ns^3\beta\sigma_v^2}{[(t-2s+2)n\alpha+1]^2} \frac{n(s+1)^2\beta[n(t+s)\beta+1]}{\alpha^2} = 0.$$

This proves Theorem 2.

If the information vector $\boldsymbol{\psi}(t)$ has the lower and upper bounds with $0 < \alpha \leq \|\boldsymbol{\psi}(t)\|^2 \leq \beta$, then $r(t)$ in the FFSG algorithm (15)–(16) satisfies

$$\frac{\alpha}{1-\lambda} \leq \lim_{t \rightarrow \infty} r(t) \leq \frac{\beta}{1-\lambda},$$

then $r(t)$ in (20) or in the GPJ algorithm (21)–(22) satisfies

$$q\alpha \leq r(t) \leq q\beta.$$

Thus, we may take $q = 1/(1-\lambda)$ as the data window length.

5 Examples

Example 1 Consider the following time-invariant stochastic system:

$$y(t) + a_1y(t-1) + a_2y(t-2) = b_1u(t-1) + b_2u(t-2) + v(t).$$

In simulation, the input $\{u(t)\}$ is taken as an uncorrelated stochastic sequence with zero mean and unit variance and $\{v(t)\}$ as a white noise sequence with zero mean and variance $\sigma_v^2 = 0.20^2$, the noise-to-signal ratio is $\delta_{\text{ns}} = 24.51\%$. Applying the SG algorithm (i.e., the GPJ algorithm with $q = t$), the projection algorithm (i.e., the GPJ algorithm with $q = 1$) and the GPJ algorithm to estimate the parameters of this example system, the parameter estimates and estimation errors $\delta := \|\hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}\|/\|\boldsymbol{\vartheta}\|$ versus t are shown in Tables 1–4 and Figure 2.

$$\sigma_v^2 = 0.10^2, \delta_{\text{ns}} = 12.25\%$$

$$\sigma_v^2 = 0.50^2, \delta_{\text{ns}} = 61.27\%$$

Table 1: The SG estimates and errors

t	a_1	a_2	b_1	b_2	δ (%)
100	-0.63990	-0.12408	0.04561	0.04426	76.27480
200	-0.67470	-0.09366	0.05148	0.07421	73.43301
500	-0.72528	-0.06592	0.06173	0.11083	69.98076
1000	-0.75978	-0.03849	0.07081	0.13611	67.31626
2000	-0.79828	-0.00705	0.08255	0.16353	64.31563
3000	-0.81718	0.01129	0.08838	0.17689	62.74391
0	-1.60000	0.80000	0.30900	0.52900	0.00000
100	-0.66122	-0.11141	0.04657	0.04386	75.09508
200	-0.69318	-0.08165	0.05240	0.07311	72.38488
500	-0.74034	-0.05029	0.06233	0.10938	68.93372
1000	-0.77558	-0.02162	0.07140	0.13455	66.19804
2000	-0.81409	0.01027	0.08310	0.16155	63.18492
3000	-0.83325	0.02907	0.08869	0.17448	61.59524
100	-0.74698	-0.05187	0.04739	0.03718	70.23421
200	-0.77582	-0.01730	0.05240	0.05986	67.58660
500	-0.81194	0.02151	0.05996	0.08823	64.41216
1000	-0.84693	0.05143	0.06761	0.10922	61.71967
2000	-0.88303	0.08312	0.07773	0.13183	58.87245
3000	-0.90194	0.10179	0.08218	0.14245	57.34344
0	-1.60000	0.80000	0.30900	0.52900	0.00000
True values	-1.60000	0.80000	0.30900	0.52900	

Example 2 Consider the following time-varying system:

$$\begin{aligned} y(t) + a(t)y(t-1) &= b(t)u(t-1) + v(t), \\ a(t) &= 0.3 + 0.0002t + 0.1 \sin(0.0021\pi t), \\ b(t) &= 1.25 + 0.1(t+100)^{0.4}, \\ \boldsymbol{\vartheta}(t) &= [a(t), b(t)]^T. \end{aligned}$$

Simulation conditions are the same as in Example 1, where the noise variance $\sigma_v^2 = 0.50^2$, the noise-to-signal ratio is between $\delta_{\text{ns}} = 13.36\%$ and $\delta_{\text{ns}} = 26.55\%$. Taking $q = 8$ and $q = 1$, the simulation results are shown in Tables 5–6 and Figures 5–6 with the estimation errors $\delta := \|\hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}(t)\|/\|\boldsymbol{\vartheta}(t)\|$.

From Tables 1 to 6 and Figures 2 to 6, we can draw the next conclusions.

Table 2: The PJ estimates and errors

t	a_1	a_2	b_1	b_2	δ (%)
100	-1.60164	0.83634	0.34112	0.57545	3.55259
200	-1.51181	0.69587	0.36761	0.53951	7.87395
500	-1.64902	0.82572	0.33176	0.49647	3.60251
1000	-1.61285	0.81293	0.31649	0.52869	1.04238
2000	-1.62737	0.82240	0.34041	0.52507	2.51030
3000	-1.61010	0.84570	0.29190	0.46246	4.39632
100	-1.59916	0.86724	0.36279	0.63290	7.13682
200	-1.48822	0.67767	0.43466	0.51300	11.03108
500	-1.67738	0.82095	0.35241	0.45838	6.09886
1000	-1.63227	0.83254	0.33222	0.52168	2.74453
2000	-1.65701	0.84612	0.37483	0.53225	5.21421
3000	-1.58256	0.84410	0.28384	0.40917	6.94442
100	-1.57101	0.92632	0.34899	0.82722	17.32636
200	-1.52203	0.70473	0.31872	0.34265	11.82260
500	-1.64611	0.69499	0.36130	0.32424	12.71648
1000	-1.73863	0.94351	0.42876	0.48346	12.54059
2000	-1.74698	0.91363	0.48531	0.60833	14.18015
3000	-1.92382	0.70698	0.52336	0.18495	27.87628
True values	-1.60000	0.80000	0.30900	0.52900	

Table 3: The GPJ estimates and errors with $q = 15$

t	a_1	a_2	b_1	b_2	δ (%)
100	-0.91770	0.04455	0.12037	0.22818	57.01671
200	-1.17278	0.27532	0.19062	0.47134	36.45479
500	-1.40248	0.61240	0.29415	0.57342	14.61838
1000	-1.56708	0.76683	0.31385	0.55392	2.81283
2000	-1.59489	0.79616	0.31693	0.53373	0.59387
3000	-1.60399	0.80514	0.30934	0.53366	0.42395
100	-0.93315	0.07057	0.11537	0.22428	55.64646
200	-1.19104	0.27069	0.19866	0.47432	35.96993
500	-1.39412	0.60825	0.29053	0.57230	15.08611
1000	-1.57221	0.77119	0.32041	0.55535	2.60531
2000	-1.59345	0.79589	0.32706	0.53921	1.17085
3000	-1.61020	0.80628	0.31094	0.54025	0.87509
100	-1.01603	0.16690	0.11053	0.21993	49.52007
200	-1.25622	0.36277	0.19939	0.38114	30.98398
500	-1.44978	0.66410	0.27826	0.52266	10.84120
1000	-1.60273	0.79714	0.33833	0.53741	1.62733
2000	-1.59772	0.80427	0.35593	0.55821	2.93452
3000	-1.62390	0.80174	0.32060	0.55976	2.15136
True values	-1.60000	0.80000	0.30900	0.52900	

- For time-invariant systems, the projection algorithm is sensitive to noise, the parameter estimation errors becomes smaller as the noise variance becomes smaller; the SG algorithm does not have the ability of tracking the (time-varying) parameters, the parameter estimation errors given by the SG algorithm are very large; the GPJ algorithm can give more accurate parameter estimates for larger q – see Tables 1 to 6 and Figure 2.

Table 4: The GPJ estimates and errors with $q = 30$

t	a_1	a_2	b_1	b_2	δ (%)
100	-0.75513	-0.09749	0.06946	0.10555	70.08128
200	-0.93937	0.06321	0.12505	0.28114	54.82236
500	-1.18183	0.38008	0.22243	0.49617	31.72171
1000	-1.41882	0.60908	0.28078	0.56771	14.14873
2000	-1.55850	0.75779	0.31029	0.55918	3.51474
3000	-1.59287	0.79425	0.30940	0.54396	0.92814
<hr/>					
100	-0.77488	-0.07713	0.06962	0.10337	68.72439
200	-0.95799	0.06103	0.12729	0.28232	54.24711
500	-1.18095	0.38103	0.21932	0.48937	31.76493
1000	-1.42303	0.61280	0.28049	0.56542	13.84179
2000	-1.56188	0.75929	0.31500	0.55844	3.35039
3000	-1.59787	0.79760	0.31050	0.54824	1.03463
<hr/>					
100	-0.84659	-0.00151	0.06635	0.08697	63.99679
200	-1.01188	0.11247	0.10425	0.20689	51.93255
500	-1.23240	0.42417	0.18200	0.40096	29.39363
1000	-1.45674	0.64633	0.26024	0.51537	11.42873
2000	-1.58228	0.77261	0.32052	0.54509	2.01803
3000	-1.61392	0.80084	0.31508	0.55832	1.74694
True values	-1.60000	0.80000	0.30900	0.52900	

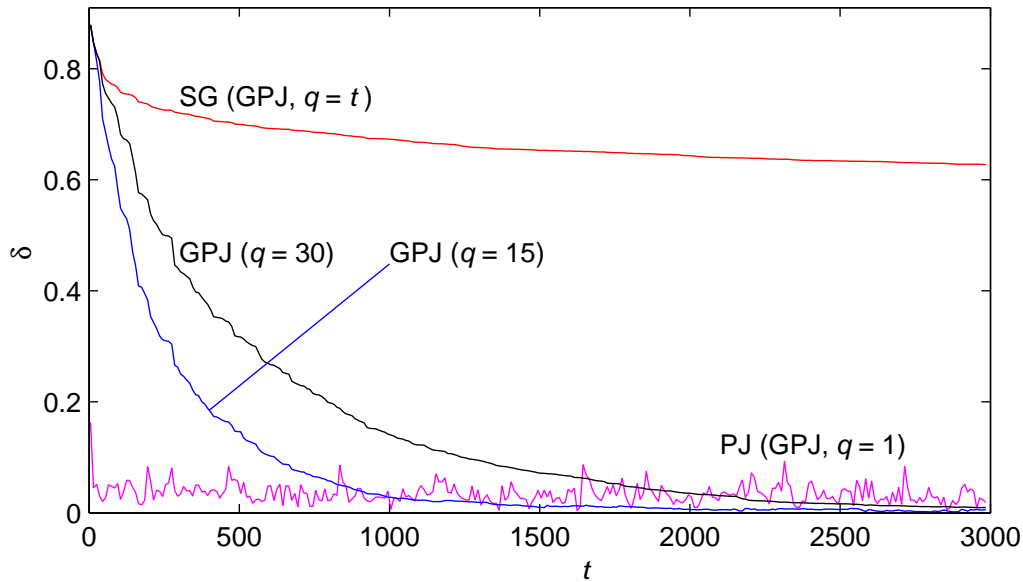


Figure 1: The estimation errors δ versus t for Example 1 ($\sigma^2 = 0.10^2$)

- For time-varying systems, the projection algorithm is still sensitive to noise, the parameter estimation errors becomes large as the noise variance becomes large, but the GPJ algorithm can give smaller parameter estimation errors for larger q – see Tables 5 to 6 and Figures 5 and 6.

6 Conclusions

This paper simply summarizes several time-varying parameter identification methods and derives a generalized projection identification algorithm. It can track the time-varying parameters and gives a bounded parameter

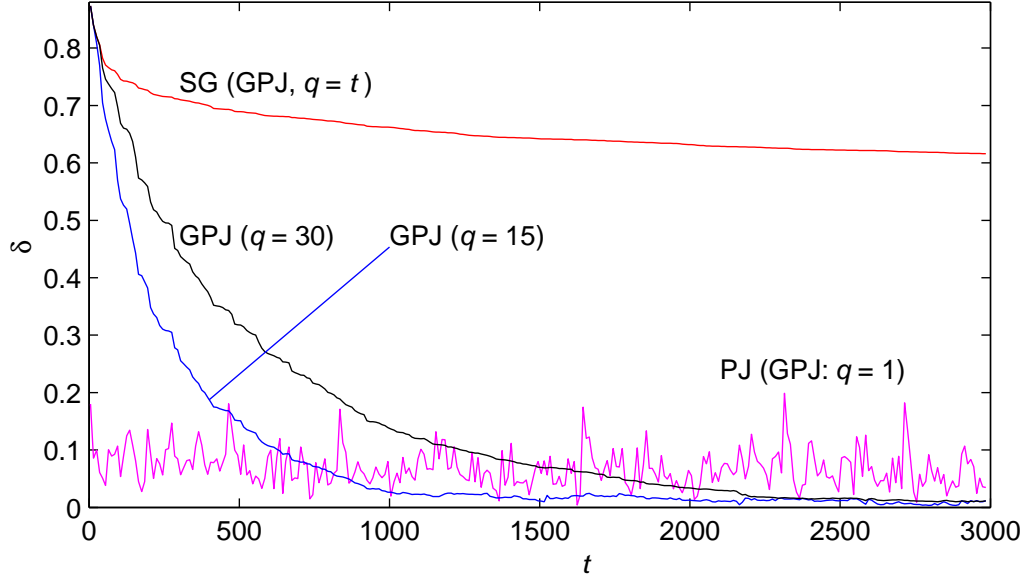


Figure 2: The estimation errors δ versus t for Example 1 ($\sigma^2 = 0.20^2$)

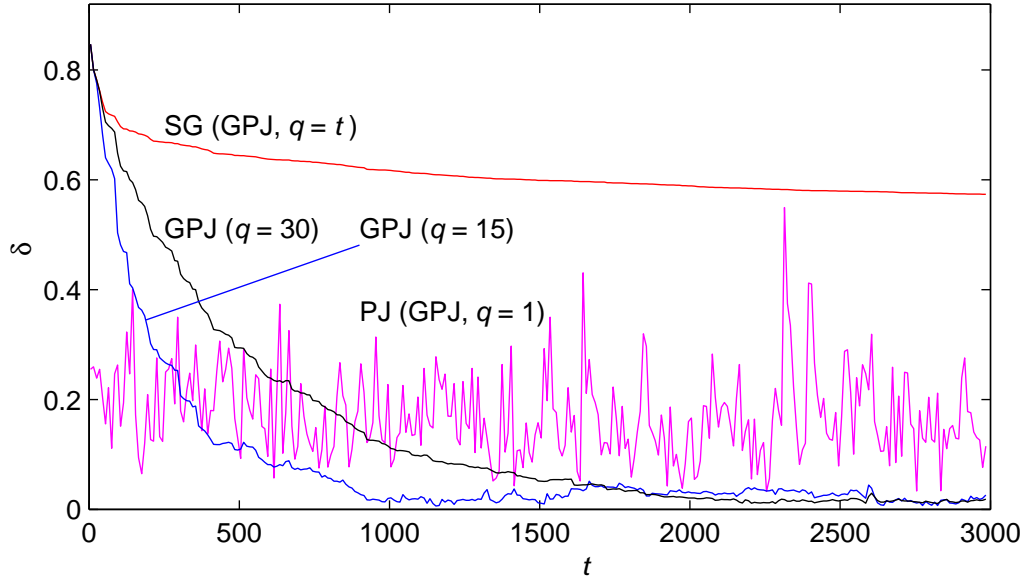


Figure 3: The estimation errors δ versus t for Example 1 ($\sigma^2 = 0.50^2$)

estimation error. For small noise-to-signal ratio, the projection algorithm can give more accurate parameter estimates, especially for time-invariant deterministic systems. The generalized projection algorithm has better stability performance for a larger data window length and its performance is superior to the projection algorithm and stochastic gradient algorithm. Therefore, when the generalized projection algorithm works at the beginning of operation, we choose a smaller data window length and then a large data window length as the time passes. The proposed method in the paper can be extended to study identification problems of other time-varying or time-invariant scalar or multivariable systems [?, 28–31].

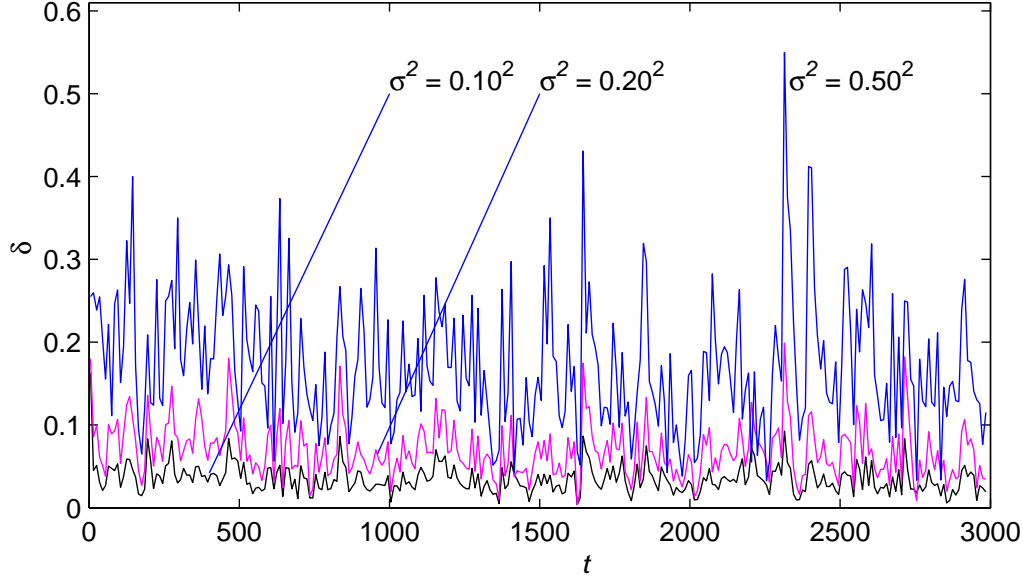


Figure 4: The estimation errors δ versus t for Example 1 ($q = 1$)

Table 5: The PJ estimates and errors

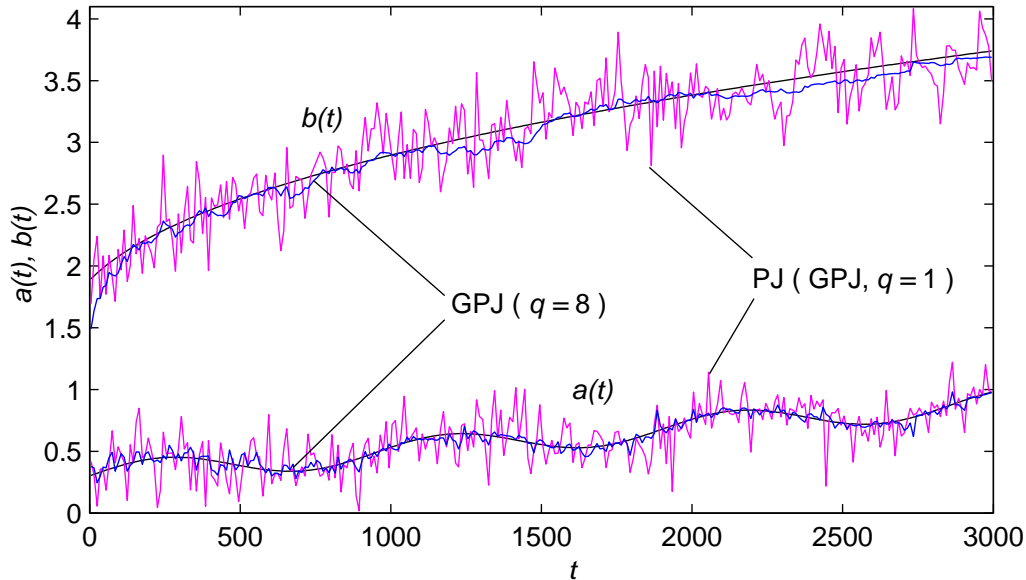
t	$a(t)$	$\hat{a}(t)$	$b(t)$	$\hat{b}(t)$	δ (%)
100	0.38201	0.65964	2.08422	2.14393	13.40192
200	0.43722	0.51625	2.23045	2.09556	6.87846
500	0.38391	0.45203	2.54285	2.61695	3.91406
1000	0.53173	0.48242	2.89708	2.97894	3.24453
2000	0.75951	0.89789	3.38290	3.29157	4.78192
3000	0.98149	1.01477	3.74230	3.77847	1.27034
100	0.38201	0.65000	2.08422	2.24654	14.78668
200	0.43722	0.47598	2.23045	2.17156	3.10191
500	0.38391	0.46177	2.54285	2.71951	7.50692
1000	0.53173	0.50263	2.89708	2.87053	1.33733
2000	0.75951	0.77689	3.38290	3.40412	0.79111
3000	0.98149	1.04752	3.74230	3.70194	2.00031
100	0.38201	0.54569	2.08422	2.10976	7.81807
200	0.43722	0.46353	2.23045	2.21902	1.26220
500	0.38391	0.45696	2.54285	2.69075	6.41416
1000	0.53173	0.50690	2.89708	2.86102	1.48632
2000	0.75951	0.76263	3.38290	3.43356	1.46393
3000	0.98149	1.01972	3.74230	3.68337	1.81576
100	0.38201	0.53387	2.08422	2.16492	8.11588
200	0.43722	0.45309	2.23045	2.21768	0.89647
500	0.38391	0.42519	2.54285	2.69018	5.94946
1000	0.53173	0.51162	2.89708	2.88971	0.72702
2000	0.75951	0.74387	3.38290	3.42877	1.39785
3000	0.98149	1.05190	3.74230	3.68910	2.28108

7 Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 61273194).

Table 6: The GPJ estimates and errors with $q = 8$

t	$a(t)$	$\hat{a}(t)$	$b(t)$	$\hat{b}(t)$	δ (%)
100	0.38201	0.50411	2.08422	2.03813	6.15913
200	0.43722	0.45194	2.23045	2.19503	1.68757
500	0.38391	0.41014	2.54285	2.57266	1.54407
1000	0.53173	0.51492	2.89708	2.90383	0.61478
2000	0.75951	0.74315	3.38290	3.39651	0.61375
3000	0.98149	1.01750	3.74230	3.69095	1.62114
100	0.38201	0.47382	2.08422	1.96660	7.04176
200	0.43722	0.43177	2.23045	2.17320	2.53011
500	0.38391	0.39470	2.54285	2.52280	0.88560
1000	0.53173	0.52518	2.89708	2.88000	0.62118
2000	0.75951	0.74470	3.38290	3.37162	0.53700
3000	0.98149	0.99648	3.74230	3.67379	1.81276
100	0.38201	0.46217	2.08422	1.85243	11.57455
200	0.43722	0.42596	2.23045	2.10329	5.61666
500	0.38391	0.38949	2.54285	2.46982	2.84838
1000	0.53173	0.51952	2.89708	2.84169	1.92569
2000	0.75951	0.74050	3.38290	3.32887	1.65196
3000	0.98149	0.97657	3.74230	3.61783	3.21989

Figure 5: The parameter estimates versus t

References

- [1] H. Li, Y. Shi and W. Yan, On neighbor information utilization in distributed receding horizon control for consensus-seeking, *IEEE Transactions on Cybernetics*, 2016. doi: 10.1109/TCYB.2015.2459719
- [2] H. Li, Y. Shi, W. Yan, Distributed receding horizon control of constrained nonlinear vehicle formations with guaranteed γ -gain stability, *Automatica* 68 (2016) 148-154, 2016.
- [3] S. Zhao, B. Huang, F. Liu, Linear optimal unbiased filter for time-variant systems without apriori information on initial condition, *IEEE Transactions on Automatic Control* (2016) doi: 10.1109/TAC.2016.2557999.
- [4] S. Zhao, Yuriy S. Shmaliy, F. Liu, Fast Kalman-like optimal unbiased FIR filtering with applications, *IEEE Transactions on Signal Processing* 64 (9) (2016) 2284-2297.

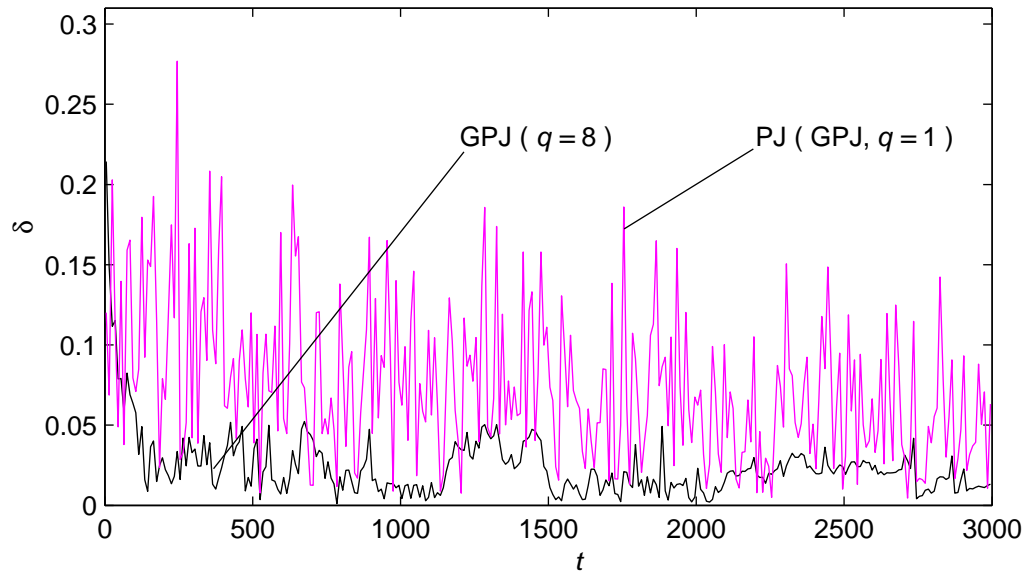


Figure 6: The estimation errors δ versus t for Example 2

- [5] T.Z. Wang, H. Wu, M.Q. Ni, et al, An adaptive confidence limit for periodic non-steady conditions fault detection, *Mechanical Systems and Signal Processing* 72-73 (2016) 328-345.
- [6] T.Z. Wang, J. Qi, H. Xu, et al, Fault diagnosis method based on FFT-RPCA-SVM for cascaded-multilevel inverter, *ISA Transactions* 60 (2016) 156-163.
- [7] Y. Shi, H. Fang, Kalman filter based identification for systems with randomly missing measurements in a network environment. *Int. J. Control* **83**(3), 538-551 (2010)
- [8] H. Li, Y. Shi, Robust H-infinity filtering for nonlinear stochastic systems with uncertainties and random delays modeled by Markov chains. *Automatica* **48**(1), 159-166 (2012)
- [9] Y. Shi, B. Yu, Robust mixed H-2/H-infinity control of networked control systems with random time delays in both forward and backward communication links. *Automatica* **47**(4), 754-760 (2011)
- [10] J. Vörös, Iterative algorithm for parameter identification of Hammerstein systems with two-segment nonlinearities. *IEEE Trans. Autom. Control* **44**(11), 2145-2149 (1999)
- [11] J. Vörös, Modeling and parameter identification of systems with multi-segment piecewise-linear characteristics. *IEEE Trans. Autom. Control* **47**(1), 184-188 (2002)
- [12] J. Vörös, Recursive identification of Hammerstein systems with discontinuous nonlinearities containing dead-zones. *IEEE Trans. Autom. Control* **48**(12), 2203-2206 (2003)
- [13] F. Ding, Y. Gu, Performance analysis of the auxiliary model-based stochastic gradient parameter estimation algorithm for state space systems with one-step state delay. *Circuits Syst. Signal Proces.* **32**(2), 585-599 (2013)
- [14] L. Ljung, *System Identification: Theory for the User*, 2nd edn. (Prentice-Hall, Englewood Cliffs, NJ, 1999)
- [15] F. Ding, T. Chen, Performance bounds of the forgetting factor least squares algorithm for time-varying systems with finite measurement data. *IEEE Trans. on Circuits Syst.–I: Regular Papers* **52**(3), 555-566 (2005)
- [16] F. Ding, J. Ding, Least squares parameter estimation with irregularly missing data. *Int. J. Adapt. Control Signal Process.* **24**(7), 540-553 (2010)
- [17] L.R. Lozano, Convergence analysis of recursive identification algorithms with forgetting factor. *Automatica* **19**(1), 95-97 (1983)
- [18] R.M. Canetti, M.D. Espana, Convergence analysis of the least-squares identification algorithm with a variable forgetting factor for time-varying linear systems. *Automatica* **25**(4), 609-612 (1989)
- [19] S. Bittanti, P. Bolzern, M. Campi, Convergence and exponential convergence of identification algorithms with directional forgetting factor. *Automatica* **26**(5), 929-932 (1990)
- [20] L. Ljung, P. Priouret, A result on the mean square error obtained using general tracking algorithms. *Int. J. Adapt. Control Signal Process.* **5**(4), 231-250 (1991)

- [21] L. Ljung, P. Priouret, Remarks on the mean square tracking error. *Int. J. Adapt. Control Signal Process.* **5**(6), 395-403 (1991)
- [22] L. Guo, L. Ljung, P. Priouret, Performance analysis of the forgetting factor RLS algorithm. *Int. J. Adapt. Control Signal Process.* **7**(6), 525-527 (1993)
- [23] G.C. Goodwin, K.S. Sin, *Adaptive Filtering Prediction and Control*. (Prentice-Hall, Englewood Cliffs, NJ, 1984)
- [24] F. Ding, The generalized projection algorithm for time-varying systems. The 32nd Chinese Control Conference (2013 CCC), July 26-28, 2013, Xi'an, China, pp. 1905-1910
- [25] F. Ding, T. Chen, Performance analysis of multi-innovation gradient type identification methods. *Automatica* **43**(1), 1-14 (2007)
- [26] F. Ding, X.P. Liu, H.Z. Yang, Parameter identification and intersample output estimation for dual-rate systems. *IEEE Trans. Syst. Man Cybern., Part A: Syst. Humans* **38**(4), 966-975 (2008)
- [27] L. Guo, *Time-Varying Stochastic Systems – Stability, Estimation and Control*. (Changchun, China: Jilin Science and Technology Press, 1993)
- [28] X. Su, L. Wu, P. Shi, Y.D. Song, H-infinity model reduction of T-S fuzzy stochastic systems. *IEEE Trans. Syst. Man Cybern., Part B: Cybern.* **42**(6), 1573-1585 (2012)
- [29] X. Su, P. Shi, L. Wu, Y.D. Song, A novel approach to filter design for T-S fuzzy discrete-time systems with time-varying delay. *IEEE Trans. Fuzzy Syst.* **20**(6), 1114-1129 (2012)
- [30] L. Wu, X. Su, P. Shi, J. Qiu, Model approximation for discrete-time state-delay systems in the T-S fuzzy framework. *IEEE Trans. Fuzzy Syst.* **19**(2), 366-378 (2011)
- [31] L. Wu, X. Su, P. Shi, J. Qiu, A new approach to stability analysis and stabilization of discrete-time T-S fuzzy time-varying delay systems. *IEEE Trans. Syst. Man Cybern., Part B: Cybern.* **41**(1), 273-286 (2011)
- =====
- [32] [1] H. Salhi, S. Kamouna, N. Essounboulib, A.Hamzaoui, Adaptive discrete-time sliding-mode control of nonlinear systems described by Wiener models, *International Journal of Control* **89** (3) (2016) 611-622.
- In this paper, we propose an adaptive control scheme that can be applied to nonlinear systems with unknown parameters. The considered class of nonlinear systems is described by the block-oriented models, specifically, the Wiener models. These models consist of dynamic linear blocks in series with static nonlinear blocks. The proposed adaptive control method is based on the inverse of the nonlinear function block and on the discrete-time sliding-mode controller. The parameters adaptation are performed using a new recursive parametric estimation algorithm. This algorithm is developed using the adjustable model method and the least squares technique. A recursive least squares (RLS) algorithm is used to estimate the inverse nonlinear function. A time-varying gain is proposed, in the discrete-time sliding mode controller, to reduce the chattering problem. The stability of the closedloop nonlinear system, with the proposed adaptive control scheme, has been proved. An application to a pH neutralisation process has been carried out and the simulation results clearly show the effectiveness of the proposed adaptive control scheme.
- [33] H. Salhi, S. Kamoun, A recursive parametric estimation algorithm of multivariable nonlinear systems described by Hammerstein mathematical models, *Applied Mathematical Modelling* **39** (16) (2015) 4951-4962.
- [34] K.V. Yuen, S.C. Kuok, Online Updating and Uncertainty Quantification Using Nonstationary Output-only Measurement, *Mechanical Systems and Signal Processing* **66** (2016) 62-77.
- Extended Kalman filter (EKF) is widely adopted for state estimation and parametric identification of dynamical systems. In this algorithm, it is required to specify the covariance matrices of the process noise and measurement noise based on prior knowledge. However, improper assignment of these noise covariance matrices leads to unreliable estimation and misleading uncertainty estimation on the system state and model parameters. Furthermore, it may induce diverging estimation. To resolve these problems, we propose a Bayesian probabilistic algorithm for online estimation of the noise parameters which are used to characterize the noise covariance matrices. There are three major appealing features of the proposed approach. First, it resolves the divergence problem in the conventional usage of EKF due to improper choice of the noise covariance matrices. Second, the proposed approach ensures the reliability of the uncertainty quantification. Finally, since the noise parameters are allowed to be time-varying, nonstationary process noise and/or measurement noise are explicitly taken into account. Examples using stationary/nonstationary response of linear/nonlinear time-varying dynamical systems are presented to demonstrate the efficacy of the proposed approach. Furthermore, comparison with the conventional usage of EKF will be provided to reveal the necessity of the proposed approach for reliable model updating and uncertainty quantification.
- [35] H.Q. Mu, K.V. Yuen, Novel outlier-resistant extended Kalman filter for robust online structural identification, *Journal of Engineering Mechanics* **141** (1) (2015), doi: 10.1061/(ASCE)EM.1943-7889.0000810.