Automatica 48 (2012) 1190–1196

Contents lists available at SciVerse ScienceDirect

Automatica

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

Brief paper

Weighted least squares based recursive parametric identification for the submodels of a PWARX system[☆]

Wen-Xiao Zhao[a,b], Tong Zhou[c]

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

[b] National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences, Beijing 100190, PR China

[c] Department of Automation, TNList, Tsinghua University, Beijing 100084, PR China

ARTICLE INFO

Article history:
Received 24 June 2010
Received in revised form
16 August 2011
Accepted 1 November 2011
Available online 5 April 2012

Keywords:
Hybrid system
Recursive identification
Kernel function
Weighted least squares
Strong consistency

ABSTRACT

A piecewise affine autoregressive system with exogenous inputs (PWARX) is composed of a finite number of ARX subsystems, each of which corresponds to a polyhedral partition of the regression space. In this work a weighted least squares (WLS) estimator is suggested to recursively estimate the parameters of the ARX submodels, in which a sequence of kernel functions are introduced. Conditions on the input signal and the PWARX system are imposed to guarantee the almost sure convergence of the WLS estimates. Some numerical examples are included to illustrate performances of the algorithm.

1. Introduction

Identification of nonlinear systems has become an active topic in recent years, both for the importance in practical applications and for the challenge in theoretical investigation. Available modeling approaches for nonlinear systems can roughly be divided into three categories, the first principles based approach, the grey-box approach and the black-box approach. The first principles based approach usually adopts physical/chemical laws to build accurate models for the nonlinear dynamics. This procedure is nontrivial and the resulting models may be intricate. In grey-box modeling, some *a prior* plant knowledge is available and Kalman filter-like algorithms are often adopted. This approach has found successful applications in area like process control (Bohlin, 1994; Raghavan et al., 2005). Another type of method is named as the black-box approach, in which the modeling procedure does not depend on the physical background of the system. The Wiener and Hammerstein systems, the nonlinear ARX system and the piecewise affine ARX (PWARX) system are widely adopted in this category for system dynamics description. There are also significant contributions on the black-box identification (Bai, Tempo, & Liu, 2007; Pajunen, 1992; Roll, Bemporad, & Ljung, 2004; Sjoberg et al., 1995; Wigren, 2006). The nonlinear dynamics in Pajunen (1992) and Wigren (2006) are transformed into linear regression form and the classical approaches, for example, the recursive prediction error approach, could be used. In Bai et al. (2007) the identification of nonlinear ARX system is considered and the value of the nonlinear function at a fixed point is estimated. In this paper, we will investigate the identification of another type of nonlinear systems, i.e. the PWARX system, which receives special attention from both theorists and engineers.

The PWARX system is composed of a finite number of ARX subsystems, each of which corresponds to a polyhedral partition of the regression space. Since neither the parameters of affine submodels nor the partition of the regression space are available, the identification of PWARX system becomes very difficult and many efforts have been devoted on this problem, for example, the data-clustering technique considered in Ferrari-Trecate, Muselli, Liberati, and Morari (2003) and Nakada, Takaba, and Katayama (2005), the bounded error approach presented in Bemporad, Garulli, Paoletti, and Vicino (2005), the Bayesian approach given in Juloski, Weiland, and Heemels (2005) and the mixed-integer programming method applied in Roll et al. (2004). Although a number of contributions have been made on the identification

[☆] The research of Tong Zhou is supported by NSFC No. 60625305, 60721003, 61174122, 61021063, and 973 Program of China No. 2009CB320602 and 2012CB316504. The research of Wen-Xiao Zhao is supported by NSFC No. 61104052, 61134013, and 91130008. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Martin Enqvist under the direction of Editor Torsten Söderström.

E-mail addresses: wxzhao@amss.ac.cn (W.-X. Zhao), tzhou@mail.tsinghua.edu.cn (T. Zhou).

0005-1098/$ – see front matter
doi:10.1016/j.automatica.2012.03.015

of PWARX system, this problem is not finished. Some important issues such as sufficient excitation of all submodels of the system, the recursive estimator and its asymptotical properties, are still open. See Paoletti, Juloski, Ferrari-Trecate, and Vidal (2007) for a detailed discussion.

In this paper, we will mainly consider identification of the parameters of affine submodels of the PWARX system. In existing literatures (see Paoletti et al., 2007, and references therein), the identification of parameters of affine submodels usually depends on data classification where each datum point must be associated to the most suitable submodel. Noticing also that the PWARX system is essentially nonlinear, the classical identification algorithms, for example, the least squares algorithm and the frequency domain approach, can not be applied. Motivated by the fact that the PWA function is *locally* linear, in this work we propose a kernel function-based weighted least squares (WLS) estimator to recursively estimate the parameters of affine submodels.

Under some mild conditions, we will prove the strong consistency of the WLS estimates. As far as the convergence analysis of recursive estimates is concerned, the ODE method is probably the most widely used approach in various application areas. See Ljung (1975, 1977) and the recent paper Brus (2007). A restriction of the ODE method is that it *a priori* assumes that the estimates are bounded, which is difficult to verify in general case. And sometimes the stationarity of the signals is required (Brus, 2007). To avoid imposing such kinds of conditions, in this paper we adopt the Markov chain method in Zhao (2008) and Zhao, Chen, and Zheng (2010) for the convergence analysis of the WLS algorithm.

The rest of this work is organized as follows. The identification algorithm is introduced in Section 2 while its asymptotical properties are given in Section 3. Two numerical examples are given in Section 4. Some concluding remarks can be found in Section 5. Appendix contains proofs of some technical results.

Notations: Let $(\Omega, \mathscr{F}, \mathbb{P})$ be the basic probability space and $\mathbf{E}(\cdot)$ be the expectation operator. Denote the Borel σ-algebra on $\mathbb{R}^{p+q}$ by $\mathscr{B}^{p+q}$ and the Lebesgue measure on $(\mathbb{R}^{p+q}, \mathscr{B}^{p+q})$ by $\mu_{p+q}(\cdot)$. The total variation norm of a signed measure $\nu(\cdot)$ is denoted by $\|\nu\|_{\mathrm{var}}$. For a Markov chain $\{\varphi_k\}_{k\geq 1}$, denote by $P_k(A) = \mathbb{P}\{\varphi_k \in A\}$ the marginal distribution and by $P_{\mathrm{IV}}(\cdot)$ the invariant probability. For real numbers a and b and a set A, denote $a \wedge b = \min\{a, b\}$ and A^c the complement of A. For vectors $a_i \in \mathbb{R}^n$, $i = 1, \ldots, N$, define $\mathrm{row}\{a_i \mid i = 1, \ldots, N\} \triangleq [a_1 \cdots a_N] \in \mathbb{R}^{n\times N}$ and $\mathrm{col}\{a_i \mid i = 1, \ldots, N\} \triangleq [a_1^T \cdots a_N^T]^T \in \mathbb{R}^{nN\times 1}$.

2. Kernel-based recursive estimation

The single-input single-output (SISO) PWARX system is formulated as follows,

$$y_{k+1} = f(y_k, \ldots, y_{k+1-p}, u_k, \ldots, u_{k+1-q}) + \varepsilon_{k+1}, \tag{1}$$

$$f(x) = \theta(l)^T \begin{bmatrix} x^T & 1 \end{bmatrix}^T, \quad x \in \mathscr{X}_l,\ l = 1, \ldots, s \tag{2}$$

where u_k and y_k are the system input and output, respectively, ε_k is the system noise, $f(\cdot)$ is a piecewise affine (PWA) function, $\{\mathscr{X}_l\}_{l=1}^s$ is an unknown polyhedral partition of $\mathbb{R}^{p+q}$, i.e., $\bigcup_{l=1}^s \mathscr{X}_l = \mathbb{R}^{p+q}$, $\mathscr{X}_m \cap \mathscr{X}_n = \emptyset, \forall m \neq n, m, n = 1, \ldots, s$, and $\theta(l) = [\beta(l)^T \alpha(l)]^T \in \mathbb{R}^{p+q+1}$ with $\beta(l) \in \mathbb{R}^{p+q}$ and $\alpha(l) \in \mathbb{R}$ is the unknown parameter vector corresponding to $\mathscr{X}_l$.

The estimates for the parameters of the linear submodel are given by the following weighted least squares (WLS) algorithm,

$$(\beta_N, \gamma_N) = \underset{(\gamma,\beta)}{\mathrm{argmin}} J_N(\beta, \gamma) \tag{3}$$

$$J_N(\beta, \gamma) = \sum_{k=1}^N w_k(\varphi^*) \left(y_{k+1} - \beta^T(\varphi_k - \varphi^*) - \gamma\right)^2, \tag{4}$$

where $\varphi_k = [y_k \cdots y_{k+1-p} u_k \cdots u_{k+1-q}]^T$, $\varphi^* = [y^{(1)} \cdots y^{(p)} u^{(1)} \cdots u^{(q)}]^T \in \mathbb{R}^{p+q}$,

$$w_k(\varphi^*) = \frac{1}{(2\pi)^{\frac{p+q}{2}}} \frac{1}{b_k^{p+q}} \cdot \exp\left\{ -\frac{1}{2}\sum_{i=1}^p \left(\frac{y_{k+1-i} - y^{(i)}}{b_k}\right)^2 - \frac{1}{2}\sum_{j=1}^q \left(\frac{u_{k+1-j} - u^{(j)}}{b_k}\right)^2 \right\}, \tag{5}$$

with $b_k = 1/k^\delta$ for some fixed $\delta \in (0, 1/2(p+q+1))$. Assume $\varphi^* \in \mathscr{X}_l$ for some $l = 1, \ldots, s$. From the definition of $w_k(\varphi^*)$, it can be found that the kernel $w_k(\varphi^*)$ decays to zero exponentially as $\|\varphi_k - \varphi^*\|$ tends to infinity. While for those regressors close to φ^*, the corresponding kernel $w_k(\varphi^*)$ can be approximated by $w_k(\varphi^*) = O\left(1/b_k^{p+q}\right) = O\left(k^{(p+q)\delta}\right)$. So the kernel function provides a *smoothed* classification for the data set $\{\varphi_k, y_{k+1}\}_{k=1}^N$ by assigning different weights to each datum point and only for those regressors close to φ^* the corresponding kernels take effective values. So (γ_N, β_N) generated from (3) and (4) may serve as the estimates for the parameters of the submodel to which the given point φ^* belongs and the data classification difficulty can be overcome. This is the key idea of the paper. For the convergence analysis of the WLS algorithm, here we adopt the Markov chain method as follows. Denoting $\Phi(\varphi_k) \triangleq [f(\varphi_k) y_k \cdots y_{k+2-p} 0 u_k \cdots u_{k+2-q}]^T$, $\xi_k \triangleq [\varepsilon_k 0 \cdots 0 u_k 0 \cdots 0]^T$, then the PWARX system (1) can be reformulated as

$$\varphi_{k+1} = \Phi(\varphi_k) + \xi_{k+1}, \tag{6}$$

which indicates that the regressor sequence $\{\varphi_k\}_{k\geq 1}$ is a Markov chain. Based on the stochastic stability results of Markov chains, the strong consistency of WLS estimates can be established.

3. Properties of identification algorithms

Assume that N pairs of input–output data $\{\varphi_k, y_{k+1}\}_{k=1}^N$ are available and a point $\varphi^* \in \mathscr{X}_l$ is given for some $l = 1, \ldots, s$. We make the following assumptions.

(A0) $\{u_k\}_{k\geq 1}$ is selected to be an iid sequence with $0 < \mathbf{E}(u_k^2) < \infty$ and with a probability density function (pdf), denoted by $f_u(\cdot)$, which is positive and continuous on $\mathbb{R}$.

(A1) The number of submodels s and the system orders (p, q) are available. Further, we assume that $s \geq 2$.

(A2) $\{\varepsilon_k\}_{k\geq 1}$ is an iid sequence with $\mathbf{E}(\varepsilon_k) = 0, 0 < \mathbf{E}\left(\varepsilon_k^2\right) < \infty$ and with a pdf denoted by $f_\varepsilon(\cdot)$, which is positive and uniformly continuous on $\mathbb{R}$; $\{u_k\}$ and $\{\varepsilon_k\}$ are mutually independent.

(A3) $f(\cdot)$ defined by (2) is bounded on $\mathbb{R}^{p+q}$.

Define matrices

$$X_N \triangleq \mathrm{row}\left\{\begin{bmatrix} 1 & (\varphi_k - \varphi^*)^T \end{bmatrix}^T \mid k = 1, \ldots, N\right\}, \tag{7}$$

$$W_N \triangleq \mathrm{diag}\left\{w_k(\varphi^*) \mid k = 1, \ldots, N\right\}, \tag{8}$$

$$Y_{N+1} \triangleq \mathrm{col}\left\{y_{k+1} \mid k = 1, \ldots, N\right\}. \tag{9}$$

Then the WLS estimate defined by (4) can be computed by the following formula,

$$\theta_N = (X_N W_N X_N^T)^+ (X_N W_N Y_{N+1}), \tag{10}$$

where A^+ denotes the Moore–Penrose inversion of a matrix A. In this paper, we can show that matrices $X_N W_N X_N^T$ are nonsingular for all N large enough. Hence the Moore–Penrose inverse matrix equals the classical inverse matrix for all N large enough. Noticing that $X_N W_N X_N^T$ is nonnegative, to preclude the possible singularity of $X_N W_N X_N^T$ with N being small, we modify the WLS estimates as