Systems & Control Letters 79 (2015) 68–75

Contents lists available at ScienceDirect

Systems & Control Letters

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

Iterative identification of block-oriented nonlinear systems based on biconvex optimization

Guoqi Li [a,*], Changyun Wen [b], Wei Xing Zheng [c], Guangshe Zhao [d]

[a] Center for Brain-Inspired Computing Research, Department of Precision Instrument, Tsinghua University, Beijing, 100084, PR China
[b] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 639798, Singapore
[c] School of Computing, Engineering and Mathematics, University of Western Sydney, Penrith NSW 2751, Australia
[d] School of Aerospace, Xi'an Jiaotong University, Shaanxi, Xi'an, 710049, PR China

ARTICLE INFO

Article history:
Received 6 November 2013
Received in revised form
2 November 2014
Accepted 18 January 2015
Available online 10 April 2015

Keywords:
Nonlinear system identification
Biconvex optimization
Parameter estimation
Hammerstein–Wiener model
Iterative method
Global minimum point

ABSTRACT

We investigate the identification of a class of block-oriented nonlinear systems which is represented by a common model in this paper. Then identifying the common model is formulated as a biconvex optimization problem. Based on this, a normalized alterative convex search (NACS) algorithm is proposed under a given arbitrary nonzero initial condition. It is shown that we only need to find the unique partial optimum point of a biconvex cost function in order to obtain its global minimum point. Thus, the convergence property of the proposed algorithm is established under arbitrary nonzero initial conditions. By applying the results to Hammerstein–Wiener systems with an invertible nonlinear function, the long-standing problem on the convergence of iteratively identifying such systems under arbitrary nonzero initial conditions is also now solved.

1. Introduction

In science and engineering, convex optimization [1] has been investigated and applied extensively. If a particular optimization problem is convex, a local minimum point and a global minimum point become equivalent. So in parameter estimation [2] which is related to system identification, convexity naturally implies the convergence of the estimates. This is why convex optimization has been widely used. However, there are certain identification problems which cannot always be formulated as convex optimization problems. For example, the identification of block-oriented nonlinear systems may not be formulated as a convex optimization problem but a biconvex one. Different from convex optimization, biconvex optimization may have a large number of local minimum points. However, it has convex substructures, in the sense that a biconvex optimization problem can be divided into several convex optimization subproblems. Such structures can indeed be exploited in solving the whole biconvex optimization problem.

In this paper, biconvex optimization will be investigated in the identification of the well known and widely implemented block-oriented nonlinear systems [3,4]. A block-oriented system is composed of different blocks such as linear dynamic subsystems and nonlinear static functions, which are interconnected in different ways [4,5]. For example, Hammerstein–Wiener system [6], which includes Hammerstein system [7,8] and Wiener system [9,10] as its special cases, is one of the most well known member of block-oriented systems with the linear dynamic block between two nonlinear blocks of static functions. To achieve our identification objectives, we first use a common model to represent the block-oriented systems and formulate the identification of the common model as a biconvex optimization problem. Then a normalized alterative convex search (NACS) algorithm is proposed under a given arbitrary nonzero initial condition. It is shown that one only needs to find the unique partial optimum point of a biconvex cost function in order to obtain its global minimum point. In this way, the convergence property of the proposed algorithm can be established under arbitrary nonzero initial conditions. A unified framework for the iterative identification of the class of block-oriented systems considered in this paper is also provided.

Note that the proposed NACS algorithm belongs to the class of iterative algorithm [11–15] which is one popular algorithm in the identification of block-oriented systems. Basically, the iterative identification approach divides the unknown parameters into two sets, the linear part and the nonlinear part. At each iteration, one set of estimates is computed while the other set is fixed.

* Corresponding author.
E-mail addresses: liguoqi@mail.tsinghua.edu.cn (G. Li), ecywen@ntu.edu.sg (C. Wen), w.zheng@uws.edu.au (W.X. Zheng), zhaogs@mail.xjtu.edu.cn (G. Zhao).

http://dx.doi.org/10.1016/j.sysconle.2015.01.003
0167-6911/

Then the two sets alternate and their final parameter estimates are obtained iteratively. The iterative algorithm may be one of the simplest and the most efficient identification schemes. However, its convergence has been a long-standing open problem. The original iterative algorithm proposed in [11] may be divergent as pointed out in [12]. But with suitable normalization, global asymptotical convergence property for general Hammerstein systems under proper given initial condition is obtained in [14]. In [15], the convergence property for general Hammerstein systems under arbitrary nonzero initial conditions is achieved. But the convergence of the iterative algorithm for a more general block-oriented system, for example, the Hammerstein–Wiener system or Wiener–Hammerstein system, is still unknown. In this paper, through solving the formulated biconvex optimization problem in identifying a proposed common model by extending the normalization idea in [13] to the alterative convex search (ACS) method, the convergence property of the iterative algorithm for the Hammerstein–Wiener system is now established under arbitrary nonzero initial conditions.

When identifying nonlinear static functions in a block-oriented system, the functions can be regarded as parametric or nonparametric. For a parametric function, it is composed of a finite number of known basis functions with unknown coefficients. Then, estimating the functions can be reduced to determining the coefficients. Clearly, our proposed iterative method is applicable to identifying parametric functions. Recent advances on parametric identification of block-oriented systems include that a new maximum-likelihood based method is proposed in [16] for parametric identification of Hammerstein–Wiener model structures with non-invertible nonlinearities and colored stochastic disturbances. On the other hand, nonparametric functions involve an infinite number of basis functions and they can be identified by nonparametric methods [9,17]. Recently a novel kernel-based nonparametric approach is developed in [18] for identification of nonlinear systems within the framework of Gaussian regression.

The remaining part of this paper is organized as follows. The identification of a proposed common model is formulated as a biconvex optimization problem in Section 2. The convergence property is analyzed in Section 3. In Section 4, the NACS algorithm is presented for the identification of Hammerstein–Wiener systems, while its effectiveness is verified via simulation results in Section 5. Finally, this paper is concluded in Section 6.

2. Biconvex optimization in parameter estimation

2.1. A common model

In this paper, we consider formulating a biconvex optimization problem for the identification of block-oriented systems which are represented in the form of model (1) and (2) given below. Hammerstein systems, Wiener systems, Hammerstein–Wiener systems and Wiener–Hammerstein systems are subsets of such systems and it will be illustrated in Section 4 how a Hammerstein–Wiener system is transformed to the model

$$Y = \mathcal{G}d + \mathcal{K}\gamma + \upsilon, \tag{1}$$

$$Y - L(d) = b_0K_0a + b_1K_1a + \cdots + b_mK_ma + \upsilon \tag{2}$$

where $\mathcal{G} \in R^{N\times\kappa}$, $\mathcal{K} \in R^{N\times\iota}$ and $K_i \in R^{N\times l}$, for $i = 0, \ldots, m$ are known matrices constructed based on learning data points, $d = [d_1 \quad \cdots \quad d_\kappa]' \in R^{\kappa\times 1}$and $\gamma = [\gamma_1 \quad \cdots \quad \gamma_\iota]' \in R^{\iota\times 1}$, $b = [b_0 \quad b_1 \quad \cdots \quad b_m]' \in R^{(m+1)\times 1}$ and $a = [a_1 \quad \cdots \quad a_l]' \in R^{l\times 1}$ are unknown parameters, $L(d)$ is a known function with respect to parameter d in (1), $Y = [y_1 \quad \cdots \quad y_N]'$ is the observation vector, and $\upsilon = [\upsilon_1 \quad \cdots \quad \upsilon_N]'$ denotes the noise vector. Our objective is to obtain the estimates $\hat{a}$, $\hat{b}$, $\hat{d}$ and $\hat{\gamma}$ of a, b, d and γ, respectively, in (1) and (2). Note that γ can be easily estimated from an obtained $\hat{d}$. Then the objective can be reduced to estimate the parameters a, b and d. In the model (1) and (2), $L(d)$ and $\mathcal{G}d$ can be either the same or different. This is also the case for $\mathcal{K}\gamma$ and $b_0K_0a + b_1K_1a + \cdots + b_mK_ma$. But when $L(d) = \mathcal{G}d$, $\mathcal{K}\gamma$ and $b_0K_0a + b_1K_1a + \cdots + b_mK_ma$ should be equal, which corresponds to the case that $K = [K_1 \quad \cdots \quad K_i \quad \cdots \quad K_m]$ and $\gamma = [b_0a' \quad \cdots \quad b_ma']'$.

Due to its generality and also for convenience, we call the above model as "common model". To characterize the common model and estimate its parameters, we make the following assumptions.

Assumption 1. The noise υ_t is an i.i.d. random variable with zero mean and finite variance denoted as σ_υ^2, and is independent of the elements in matrices K_i for $i = 0, \ldots, m$.

Assumption 2. $\|b\|_2 = \sqrt{b'b}$ is known and the first nonzero entry of b is positive. Without loss of generality, it is assumed that $b_1 > 0$.

Assumption 3. In (1), $[\mathcal{G} \quad \mathcal{K}]$ is a full column rank matrix and $tr((\mathcal{G}'\mathcal{G})^{-1})$ approaches zero in probability as $N \to \infty$ where $tr(.)$ is the trace of a matrix.

Assumption 4. The matrix $K = [K_0 \quad \cdots \quad K_i \quad \cdots \quad K_m]$ in (2) satisfies that $\lim_{N\to\infty} \frac{K'K}{N} = I$ almost surely where I is an identity matrix.

Remark 1. Assumption 2 is to guarantee a unique expression of the common model, as any pair λa and b/λ for nonzero λ provides the same observations in Eq. (2). If (a, b) are the true parameters of the common model, both (a, b) and $(-a, -b)$ will be the global minimum points of objective function defined in function (3). So one needs to fix the norm $\|b\|_2$ and the sign of the first entry of b. Assumption 3 basically implies that the trace $tr(\mathcal{G}'\mathcal{G})$ approaches positive infinity as $N \to \infty$. For more details, please refer to Lemma 3.4 in Ref. [3].

In the next subsection, it will be shown how to estimate d in (1), a and b in (2), respectively. In Section 3, we will prove that the obtained estimates $\hat{d}$, $\hat{a}$ and $\hat{b}$ converge to their true values in probability under Assumptions 1–4.

When $\hat{d}$ in (7) is obtained without depending on other parameters, which is seen in (8) later, one needs to minimize the following cost function $J_N^{\hat{d}}$ for the estimation of a and b:

$$\begin{aligned} J_N^{\hat{d}}(\bar{a}, \bar{b}) &= \frac{1}{N}(\bar{Y} - Y)'(\bar{Y} - Y) \\ &= \frac{1}{N}(\bar{Y} - Y^* - \upsilon)'(\bar{Y} - Y^* - \upsilon) \\ \text{s.t. } &\sqrt{\bar{b}'\bar{b}} = 1, \\ &\bar{b}_1 > 0 \end{aligned} \tag{3}$$

where the observed system output $Y = Y^* + \upsilon$, $\bar{a}$ and $\bar{b}$ are variables of the cost function and

$$\begin{aligned} &\bar{a} = [\bar{a}_0 \quad \cdots \quad \bar{a}_l]', \qquad \bar{b} = [\bar{b}_1 \quad \cdots \quad \bar{b}_m]', \\ &\bar{Y} = L(\hat{d}) + \bar{b}_0K_0\bar{a} + \cdots + \bar{b}_mK_m\bar{a}, \\ &Y^* = L(d) + b_0K_0a + \cdots + b_mK_ma. \end{aligned} \tag{4}$$

Lemma 2.1. *Minimizing $J_N^{\hat{d}}(\bar{a}, \bar{b})$ in a bounded convex set is a biconvex optimization defined in Definition 4 in the Appendix.*

Proof. Firstly, it is easy to see that a bounded convex set is of course a biconvex set from Definition 1 in the Appendix. Secondly, $J_N^{\hat{d}}(\bar{a}, \bar{b})$ is a biconvex function with respect to separable $\bar{a}$ and $\bar{b}$ based on Definitions 2–4 in the Appendix. Thus, this lemma holds. □