Identification of Cascade Dynamic Nonlinear Systems: A Bargaining-Game-Theory-Based Approach

Zhenlong Xiao, *Member, IEEE*, Shengbo Shan, and Li Cheng

***Abstract*—Cascade dynamic nonlinear systems can describe a wide class of engineering problems, but little efforts have been devoted to the identification of such systems so far. One of the difficulties comes from its non-convex characteristic. In this paper, the identification of a general cascade dynamic nonlinear system is rearranged and transformed into a convex problem involving a double-input single-output nonlinear system. In order to limit the estimate error at the frequencies of interest and to overcome the singularity problem incurred in the least-square-based methods, the identification problem is, thereafter, decomposed into a multi-objective optimization problem, in which the objective functions are defined in terms of the spectra of the unbiased error function at the frequencies of interest and are expressed as a first-order polynomial of the model parameters to be identified. The coefficients of the first-order polynomial are derived in an explicit expression involving the system input and the measured noised output. To tackle the convergence performance of the multi-objective optimization problem, the bargaining game theory is used to model the interactions and the competitions among multiple objectives defined at the frequencies of interest. Using the game-theory-based approach, both the global information and the local information are taken into account in the optimization, which leads to an obvious improvement of the convergence performance. Numerical studies demonstrate that the proposed bargaining-game-theory-based algorithm is effective and efficient for the multi-objective optimization problem, and so is the identification of the cascade dynamic nonlinear systems.**

***Index Terms*—Cascade dynamic nonlinear systems, game theory, system identification, frequency domain.**

Manuscript received January 17, 2018; revised May 21, 2018 and June 14, 2018; accepted July 2, 2018. Date of publication July 20, 2018; date of current version August 2, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dennis Wei. This work was supported in part by the Xiamen Engineering Technology Center for Intelligent Maintenance of Infrastructures (No. TCIMI201814) and in part by the Research Grants Council of HKSAR through the project PolyU 152070/16E. *(Corresponding author: Zhenlong Xiao.)*

Z. Xiao is with the Fujian Key Laboratory of Sensing and Computing for Smart City, School of Information Science and Engineering and the Xiamen Engineering Technology Center for Intelligent Maintenance of Infrastructures, Xiamen University, Fujian 361005, China, and also with the Department of Mechanical Engineering, Hong Kong Polytechnic University, Hong Kong (e-mail: zlxiao@xmu.edu.cn).

S. Shan and L. Cheng are with the Department of Mechanical Engineering, Hong Kong Polytechnic University, Hong Kong (e-mail: shengbo.shan@connect.polyu.hk; li.cheng@polyu.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2018.2858212

I. Introduction

THE signal processing for nonlinear feature extractions in nonlinear systems attracts increasing attentions due to its significant application value in many engineering problems. For example, nonlinear characteristics (e.g., the second-order or the third-order harmonics) of Lamb waves can be used for structural health monitoring applications [1], [2]. In such systems, in addition to the damage-induced nonlinear components, nonlinearities can also be generated from other constituents of the system (which can be regarded as subsystems) such as transducers, bonding layers as well as measurement devices. Although some of the nonlinear components, e.g., power-amplifier and transducer nonlinearities, can be apprehended using pre-distortion techniques [3], [4], the handling of other components from the physical system itself is difficult to be apprehended. For example, the bonding-layer nonlinearity and the damage-induced nonlinearity. Moreover, due to the existence of the system damping, the corresponding subsystems behave like dynamic nonlinear systems, in which the past inputs and/or the past outputs intervene in the nonlinear systems. Such system can therefore be modeled as a general cascade dynamic nonlinear system [2], which is the model under the consideration in this paper.

Pioneering works in nonlinear system identification have been largely devoted to the Wiener model (linear-nonlinear system) [5]–[10] and the Hammerstein model (nonlinear-linear model) [11], [12]. The Hammerstein-Wiener model can deal with a cascade nonlinear system involving nonlinear-linear-nonlinear combinations, in which both nonlinear subsystems are static nonlinear. In this paper, we will investigate the cascade of two dynamic nonlinear systems. Existing techniques based on the Hammerstein-Wiener model, such as the iterative method [13], [14], the maximum-likelihood method [15], the over-parametrization method [16], [17], the biconvex based method [18], and the frequency domain method [19] cannot be directly applied to the identification of the cascade dynamic nonlinear systems.

Despite its wide application value in representing a large class of physical problems, cascade dynamic nonlinear systems have received little attention as far as the identification is concerned. One of the fundamental problems is that the identification problem is inherently non-convex. As an example, the input backlash nonlinearity and the output static nonlinearity were first identified together and then separated via the Kozen-Landau

decomposition algorithm [20], [21]. In [22], [23], an odd nonlinearity was assumed for the input system, and in [24]–[26] an iterative gradient based approach was proposed. The nonconvex characteristic has been shown to lead to problems like the local minimum.

Note that some of the nonlinear features needed to be described in the frequency domain, exemplified by the second-order harmonics and the third-order harmonics for power amplifiers [11] and structural health monitoring problems [2]. Therefore, it is desirable that the identification of the cascade dynamic nonlinear systems be conducted in the frequency domain such that the nonlinear features can be directly captured with the estimate error being guaranteed at the frequencies of interest. In this way, the nonlinear cost function (error function) can be defined and optimized directly in the frequency domain, which is more straightforward and more accurate to study the harmonics based features when compared with the indirect time-domain approach. The latter requires the identification in the time domain first and thereafter uses the fast Fourier transformation. The whole process contributes to a global estimate error which can be obtained. The error at the frequencies of interest, however, cannot be guaranteed. To tackle this problem, some previous attempts were made to conduct the frequency-domain based identification of cascade nonlinear systems. For example, the best linear approximation (BLA) method was proposed for the identification of the Wiener model [27], the Hammerstein model [28], [29], and the Wiener-Hammerstein model [30] in the frequency domain. BLA, however, fails in the identification of the cascade dynamic nonlinear systems because the Bussgang theorem does not hold for dynamic nonlinearities [31]. The Volterra series based frequency-domain method is a powerful tool for dynamic nonlinear systems [32]–[35] (e.g., nonlinear ordinary differential equations (NODE) and nonlinear autoregressive with exogenous input (NARX) model) under the convergence conditions [33], [36]–[38]. The generalized frequency response function [39], [40] and the output spectrum function were studied [41]. Thereafter, the characteristic relationship between the output spectrum and the parameters of interest was investigated [42]–[45]. However, very few efforts have been devoted to the identification of cascade nonlinear systems using Volterra series based frequency-domain method.

The objective of the present work is to investigate the identification of a general cascade dynamic nonlinear system in frequency domain. Firstly, to tackle the non-convex characteristic, a cascade dynamic nonlinear system is rearranged, and the estimate error function of the identification problem is cast into a double-input single-output (DISO) nonlinear function. In this way, the identification problem is transformed into a convex problem. Secondly, in order to guarantee the estimate error at each frequency of interest, the identification problem is decomposed into a set of sub-problems that are defined at the same frequency. The identification problem then becomes a multi-objective optimization one. Although the convex characteristic holds in the decomposition of the multi-objective problem, an effective and efficient algorithm is not straightforward. Despite the success of the gradient descent method, Newton method, Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm and the Levenberg-Marquardt (LM) algorithm in dealing with single-objective optimization problems, for the multi-objective case, the 'jump' phenomenon may exist between two consecutive iterations, which would be more obvious especially when the number of sub-problems is large, and therefore would deteriorate the convergence performance of the identification. Although evolution algorithms, for example, the genetic algorithm and the simulated annealing algorithm, are in principle applicable to this multi-objective optimization problem, potential limitations exist to hamper their practical implementation. For example, the adjustment of the parameters (e.g., the commute rate for GA) may depend on the experience, and the local search performance needs to be improved because the feedback information of the current candidate solution is not full employed for generating the new candidate solutions. To tackle these problems, the present work makes use of the bargaining game theory to model the interactions and competitions among various sub-problems in the multi-objective optimization. The sub-problems defined at the frequencies of interest can determine their own strategies according to the preceding competition result. The bargaining game theory based algorithm (BGTA) employs both the global and the local information in the multi-objective optimization process, therefore would greatly improve the convergence performance.

The contributions of this paper are summarized as follows: 1) To transform the non-convex problem into a convex problem, the cascade dynamic nonlinear systems are rearranged, allowing the identification problem to be modeled as a DISO Volterra system. Note that both of the two sub-systems are dynamic nonlinear; 2) To guarantee the estimate error at the frequencies of interest, the identification problem is decomposed into a multi-objective optimization problem; the unbiased error function is given, and the spectrum of the unbiased error function is demonstrated to be a first-order polynomial of the model parameters of the cascade dynamic nonlinear systems to be identified, such characteristic relationship will greatly facilitate the multi-objective optimization in the frequency domain; 3) To develop an effective and efficient algorithm for the multi-objective optimization problem, bargaining game theory is employed to model the interactions and competitions among various sub-problems, which would greatly improve the convergence performance of the multi-objective optimization problem.

The remaining part of the paper is arranged as follows: In Section II, the model under the consideration and the identification problem are formulated. In Section III, the bargaining game theory based identification of the cascade dynamic nonlinear systems is proposed. Numerical results and discussions are presented in Section IV. Finally, conclusions are presented in Section V.

II. Model Under the Consideration and Problem Formulation

A. Model Under the Consideration

Volterra model was widely reported in the literature to characterize a dynamic nonlinear system. Here we study a cascade dynamic nonlinear system as the cascade of two Volterra

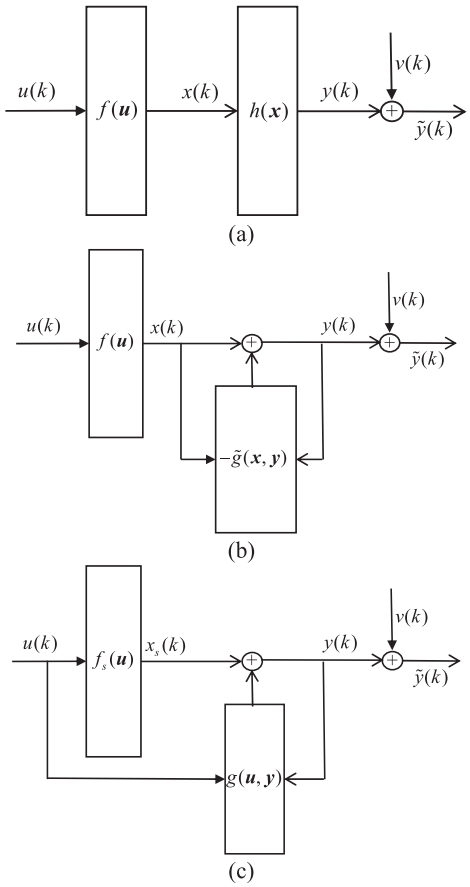

Fig. 1. Cascade dynamic nonlinear systems. (a) is a straightforward model of a cascade dynamic nonlinear system, (b) is an intermediate transformation of model (a), and (c) is the model under the consideration, which is a rearranged model of (a).

models as shown in Fig. 1(a), in which $f(\boldsymbol{u})$ and $h(\boldsymbol{x})$ are given as polynomial nonlinearities

$$x(k) = f(\boldsymbol{u}) = \sum_{i=1}^{I_f} \sum_{0 \le n_1 \le \cdots \le n_i \le N_f} \widehat{c}_i(n_1, \ldots, n_i) \times u(k-n_1) \cdots u(k-n_i) \tag{1}$$

and

$$y(k) = h(\boldsymbol{x}) = \sum_{i=1}^{I_h} \sum_{0 \le n_1 \le \cdots \le n_i \le N_h} \tilde{c}_i(n_1, \ldots, n_i) \times x(k-n_1) \cdots x(k-n_i), \tag{2}$$

where $x(k)$ is an unmeasurable intermediate variable, $u(k)$ and $y(k)$ are the system input and output, respectively. $n_1, \ldots, n_i$ are the difference orders with the maximum order N_f and N_h, and i is the nonlinear order with the maximum order I_f and I_h, respectively. $\widehat{c}_i(\cdot)$ and $\tilde{c}_i(\cdot)$ are the nonlinear parameters of model (1) and (2). Denote $\boldsymbol{x} = [x(k), \ldots, x(k-N)]$ and $\boldsymbol{y} = [y(k), \ldots, y(k-N)]$.

By substituting $x(k)$ in (1) into sub-system (2), it is obvious that the identification of the cascade dynamic nonlinear system shown in Fig. 1(a) turns into the optimization of a polynomial up to $(I_{\mathrm{h}}+1)$-th order (i.e., $\tilde{c}_i(\cdot)\widehat{c}_i^{\,I_h}(\cdot)$), which is a non-convex problem. Therefore, the parameters $\widehat{c}_i(\cdot)$ and $\tilde{c}_i(\cdot)$ cannot be uniquely determined. To overcome the non-convex characteristic, we split sub-system (2) into two parts as

$$h(\boldsymbol{x}) = x(k) + \tilde{h}(\boldsymbol{x}). \tag{3}$$

Given a $\tilde{h}(\boldsymbol{x})$, we can construct a nonlinear autoregressive exogenous (NARX) model

$$\tilde{h}(\boldsymbol{x}) + \tilde{g}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\tilde{p}=1}^{I_g} \sum_{0 \le n_1 \le \cdots \le n_{\tilde{p}} \le N} \tilde{d}_{\tilde{p},0}(n_1, \ldots, n_{\tilde{p}}) \prod_{s=1}^{\tilde{p}} x(k-n_s) + \sum_{\substack{i=1, q \ge 1 \\ \tilde{p}+q=i}}^{I_g} \sum_{\substack{0 \le n_1 \le \cdots \le n_{\tilde{p}} \le N \\ 0 \le n_{\tilde{p}+1} \le \cdots \le n_i \le N}} \tilde{d}_{\tilde{p},q}(n_1, \ldots, n_{\tilde{p}+q}) \prod_{s=1}^{\tilde{p}} x(k-n_s) \prod_{s=\tilde{p}+1}^{\tilde{p}+q} y(k-n_s) = \sum_{\substack{i=1 \\ \tilde{p}+q=i}}^{I_g} \sum_{\substack{0 \le n_1 \le \cdots \le n_{\tilde{p}} \le N \\ 0 \le n_{\tilde{p}+1} \le \cdots \le n_i \le N}} \tilde{d}_{\tilde{p},q}(n_1, \ldots, n_{\tilde{p}+q}) \prod_{s=1}^{\tilde{p}} x(k-n_s) \prod_{s=\tilde{p}+1}^{\tilde{p}+q} y(k-n_s) \tag{4}$$

such that

$$\left\| \tilde{h}(\boldsymbol{x}) + \tilde{g}(\boldsymbol{x}, \boldsymbol{y}) \right\| = \|e(\boldsymbol{x}, \boldsymbol{y})\| \to 0 \tag{5}$$

holds under certain norm $\|\cdot\|$ with appropriate nonlinear order I_g and difference order N. $\tilde{d}_{\tilde{p},q}(n_1, \ldots, n_{\tilde{p}+q})$ is the nonlinear parameter of the corresponding nonlinear term $\prod_{s=1}^{\tilde{p}} x(k-n_s) \prod_{s=\tilde{p}+1}^{\tilde{p}+q} y(k-n_s)$, $\tilde{p}$ and q are the nonlinear orders of the input $\boldsymbol{x}$ and output $\boldsymbol{y}$, respectively. Denote $\prod_{s=i+1}^{i}(\cdot) = 1$, and

$$\tilde{d}_{\tilde{p},0}(n_1, \ldots, n_{\tilde{p}}) = \begin{cases} \tilde{c}_1(0) - 1, & \tilde{p} = 1, n_1 = 0, \\ \tilde{c}_i(n_1, \ldots, n_i), & 1 \le \tilde{p} = i \le I_h, 0 \le n_1, \ldots, n_i \le N_h, \\ & \text{except } \tilde{p} = 1, n_1 = 0, \\ 0, & \text{others.} \end{cases} \tag{6}$$

From (4) and (5), $\tilde{h}(\boldsymbol{x})$ can be approximated by $-\tilde{g}(\boldsymbol{x}, \boldsymbol{y})$, which can be denoted as $\tilde{h}(\boldsymbol{x}) = -\tilde{g}(\boldsymbol{x}, \boldsymbol{y})$. Therefore, $h(\boldsymbol{x}) = x(k) - \tilde{g}(\boldsymbol{x}, \boldsymbol{y})$ holds. Fig. 1(a) can then be converted to Fig. 1(b). But the unmeasurable intermediate variable $\boldsymbol{x}$ still exists in the model. Substitute $\boldsymbol{x}$ into $\tilde{g}(\boldsymbol{x}, \boldsymbol{y})$,

$$\tilde{g}(\boldsymbol{u}, \boldsymbol{y}) = \tilde{g}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\substack{i=1, q \ge 1 \\ p+q=i}}^{I} \sum_{\substack{0 \le n_1 \le \cdots \le n_p \le N+N_f \\ 0 \le n_{p+1} \le \cdots \le n_i \le N}} \times \widehat{d}_{p,q}(n_1, \ldots, n_{p+q}) \times \prod_{s=1}^{p} u(k-n_s) \prod_{s=p+1}^{p+q} y(k-n_s), \tag{7}$$

where

$$\widehat{d}_{p,q}(n_1,\ldots,n_{p+q}) = \sum_{\tilde{p}=1}^{p} \tilde{d}_{\tilde{p},q}(n_1,\ldots,n_{\tilde{p}+q}) \times \sum_{i_1+\cdots+i_{\tilde{p}}=p} \prod_{s=1}^{\tilde{p}} \widehat{c}_{i_s}(n_{s1},\ldots,n_{si_s}). \tag{8}$$

Denote the difference order of parameter $\tilde{d}_{\tilde{p},q}(\cdot)$ in (8) as $\tilde{n}_1 = n_1,\ldots,\tilde{n}_{\tilde{p}} = n_{\tilde{p}}$, then $n_1,\ldots,n_p$ in $\widehat{d}_{p,q}(\cdot)$ is the ascending sorting of $\{n_{11}+\tilde{n}_1, n_{12}+\tilde{n}_1,\ldots,n_{1i_1}+\tilde{n}_1,\ldots,n_{\tilde{p}1}+\tilde{n}_{\tilde{p}}, n_{\tilde{p}2}+\tilde{n}_{\tilde{p}},\ldots,n_{\tilde{p}i_{\tilde{p}}}+\tilde{n}_{\tilde{p}}\}$.

Substitute (1) and (7) into model Fig. 1(b), the following equation can be obtained:

$$\begin{aligned} y(k) &= x(k) - \tilde{g}(\boldsymbol{x},\boldsymbol{y}) = x(k) - \tilde{g}(\boldsymbol{u},\boldsymbol{y}) \\ &= \sum_{i=1}^{I_f} \sum_{0\le n_1\le\cdots\le n_i\le N_f} \widehat{c}_i(n_1,\ldots,n_i)u(k-n_1)\cdots u(k-n_i) \\ &\quad - \sum_{\substack{i=1,q\ge 1\\ p+q=i}}^{I} \sum_{\substack{0\le n_1\le\cdots\le n_p\le N+N_f\\ 0\le n_{p+1}\le\cdots\le n_i\le N}} \widehat{d}_{p,q}(n_1,\ldots,n_{p+q}) \\ &\quad \times \prod_{s=1}^{p} u(k-n_s) \prod_{s=p+1}^{p+q} y(k-n_s). \end{aligned} \tag{9}$$

Moving $y(k)$ in (9) from the left-hand side to the right-hand side, the corresponding parameter is $-(\widehat{d}_{0,1}(0)+1)$. Note that multiplying a non-zero constant onto both sides of the equation will not make any difference for identifying (9), which is the so-called scale deflection problem [18], [46], [47]. To avoid the scale deflection, we assume that the parameter of $y(k)$ in (9) is non-zero, i.e., $-(\widehat{d}_{0,1}(0)+1) \neq 0$. Multiplying both sides of (9) by a constant $1/(\widehat{d}_{0,1}(0)+1)$, the model Fig. 1(c) can be obtained as

$$\begin{aligned} &f_s(\boldsymbol{x}) + g(\boldsymbol{u},\boldsymbol{y}) - y(k) \\ &= \sum_{i=1}^{I_f} \sum_{0\le n_1\le\cdots\le n_i\le N_f} c_i(n_1,\ldots,n_i)u(k-n_1)\cdots u(k-n_i) \\ &\quad + \sum_{\substack{i=1,q\ge 1\\ p+q=i}}^{I} \sum_{\substack{0\le n_1\le\cdots\le n_p\le N+N_f\\ 0\le n_{p+1}\le\cdots\le n_i\le N}} d_{p,q}(n_1,\ldots,n_{p+q}) \\ &\quad \times \prod_{s=1}^{p} u(k-n_s) \prod_{s=p+1}^{p+q} y(k-n_s) = 0, \end{aligned} \tag{10}$$

$$\tilde{y}(k) = y(k) + v(k), \tag{11}$$

where

$$c_i(n_1,\ldots,n_i) = \frac{\widehat{c}_i(n_1,\ldots,n_i)}{\left(\widehat{d}_{0,1}(0)+1\right)}, \tag{12}$$

$$d_{p,q}(n_1,\ldots,n_i) = \begin{cases} -\dfrac{\widehat{d}_{p,q}(n_1,\ldots,n_i)}{\left(\widehat{d}_{0,1}(0)+1\right)}, & q \ge 1, \\ 0, & q = 0, \end{cases} \tag{13}$$

and obviously,

$$d_{0,1}(0) = -1. \tag{14}$$

$v(k)$ is the Gaussian white noise with zero mean and variance σ^2. $y(k)$ and $\tilde{y}(k)$ are the outputs without noise and with noise, respectively. Obviously,

$$x_s(k) = f_s(\boldsymbol{x}) = \frac{x(k)}{\left(\widehat{d}_{0,1}(0)+1\right)} \tag{15}$$

and

$$g(\boldsymbol{u},\boldsymbol{y}) = -\frac{\tilde{g}(\boldsymbol{x},\boldsymbol{y}) - y(k)}{\left(\widehat{d}_{0,1}(0)+1\right)} + y(k). \tag{16}$$

Remark 1: Given the cascade dynamic nonlinear model shown in Fig. 1(a), there exists an equivalent model shown in Fig. 1(c) that can be described by (10)–(14), which is the model under the consideration in this paper.

Remark 2: If the model in Fig. 1(a) is simply identified via a high-order Volterra model, the individual sub-system in the cascade cannot be uniquely and accurately determined because the identification is a non-convex problem. With the transformation from Fig. 1(a) to Fig. 1(c), once the model (10)–(14) is identified, the first sub-system can be obtained as $f_s(\boldsymbol{u}) = \sum_{i=1}^{I_f}\sum_{0\le n_1\le\cdots\le n_i\le N_f} c_i(n_1,\ldots,n_i)\prod_{j=1}^{i}u(k-n_j)$, and the second sub-system can be modeled by (10), i.e., $y(k) = x_s(k) + g(\boldsymbol{u},\boldsymbol{y}) = x(k) - \tilde{g}(\boldsymbol{x},\boldsymbol{y}) = h(\boldsymbol{x})$.

Remark 3: If we only consider the identification of model (10)–(14) without the transformation from Fig. 1(a) to Fig. 1(c), the two nonlinear sub-systems, $f(\boldsymbol{u})$ and $h(\boldsymbol{x})$, cannot be uniquely estimated either, and only the general Volterra model (10) can be obtained.

Assumption 1: A multi-frequency excitation, i.e.,

$$u(k) = \sum_{q=1}^{Q} A_q \sin(\omega_q k + \chi_q) \tag{17}$$

is assumed, where A_q, ω_q, χ_q are the amplitude, the frequency, and the phase for the q-th sinusoidal input, respectively. As assumed after (14), the noise satisfies $v(k) \sim N(0,\sigma^2)$.

B. Identification of Cascade Dynamic Nonlinear Systems

Given the model (10)–(14), the identification of the cascade dynamic nonlinear systems boils down to identifying the coefficients $c_i(\cdot)$ and $d_{p,q}(\cdot)$ from the given input $u(k)$ and the measurable output $\tilde{y}(k)$. The intermediate variable $x(k)$ is unmeasurable. The identification problem can be expressed in time

domain as

$$\varepsilon(k) = \sum_{i=1}^{I_f} \sum_{0\le n_1\le\cdots\le n_i\le N_f} c_i(n_1,\ldots,n_i)u(k-n_1)\cdots u \times (k-n_i) + \sum_{\substack{i=1,q\ge 1\\ p+q=i}}^{\mathrm{I}} \sum_{\substack{0\le n_1\le\cdots\le n_p\le N+N_f\\ 0\le n_{p+1}\le\cdots\le n_{p+q}\le N}} \times d_{p,q}(n_1,\ldots,n_{p+q}) \prod_{s=1}^{p} u(k-n_s) \prod_{s=p+1}^{p+q} \tilde{y}(k-n_s), \tag{18}$$

$$\arg\min_{\boldsymbol{\theta}} \|\boldsymbol{\varepsilon}\|, \tag{19}$$

where $\boldsymbol{\theta} = [c_1(0),\ldots,c_{I_f}(N_f,\ldots,N_f), d_{0,1}(0),\ldots,d_{0,I}(N,\ldots,N),\ldots,d_{I,0}(N+N_f,\ldots,N+N_f)]$ are the coefficients to be identified, $\boldsymbol{\varepsilon} = [\varepsilon(k),\ldots,\varepsilon(k-\mathrm{M})]$ with M the number of observed errors, and $\|\cdot\|$ denotes a metric with a certain norm. $\tilde{y}$ is the output with noise v.

Assumption 2: In this paper, we assume that the maximum nonlinear order I and I_{f}, and the maximum difference order N and N_{f} are known. The following work then focus on the identification of the model parameters $\boldsymbol{\theta}$.

Remark 4: The identification of the cascade model presented in Fig. 1(a) is non-convex, which may be solved using evolution algorithms. However, the evolution algorithms may be time-consuming, and the identification cannot be guaranteed to achieve the global optimum. After rearranging the model to the one shown in Fig. 1(c) and considering the identification problem (18), (19) as a DISO system with u and $\tilde{y}$ being the two new inputs and the error function ε as the new output, the error function (18) is linear in parameters, and therefore is a convex function. The rearrangement of the cascade dynamic nonlinear model from Fig. 1(a) to Fig. 1(c), therefore, transform the identification problem from a non-convex problem into a convex problem.

Remark 5: The identification problem (18), (19) may be solved with the least square method $\hat{\boldsymbol{\theta}} = (\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\varepsilon}$, which could potentially become singular especially when the number of parameters to be identified is large. Even when a global optimum is obtained, if the nonlinear features are to be described in the frequency domain and the estimate error at the frequencies of interest needs to be quantified, the evolution algorithms and the least square method can only provide the total estimate error but cannot guarantee the errors at the frequencies of interest. These motivate the present work to deal with the identification problem (18), (19) in the frequency domain directly.

III. Game Theory Based Frequency Domain Identification Approach

In this section, the identification problem (18), (19) is dealt with in the frequency domain. Firstly, the error function $\varepsilon(k)$ will be expressed as the output of a DISO nonlinear system, and its output spectrum will be given as an explicit polynomial function of the input spectra. Once the output spectrum of $\varepsilon(k)$ is obtained, the identification problem (18), (19) can be conducted in the frequency domain by optimizing the output spectra in the whole frequency range Ω. In this case, we need to consider the output spectrum of error $\varepsilon(k)$ at each frequency of interest, then, the identification problem is converted into a multi-objective optimization one. Note that the effective and efficient handling of a multi-objective optimization problem is not straightforward. To tackle this problem, the bargaining game theory is employed to model the interactions and competitions among the sub-problems defined at the frequencies of interest, leading to the proposition of BGTA, conducive to the multi-objective optimization problem.

A. Output Spectrum of the Error Function $\varepsilon(\cdot)$

It is well-known that in a forward problem, i.e., from input $u(k)$ to intermediate variable $x(k)$ and then the output $y(k)$ and noised output $\tilde{y}(k)$, the identification is unbiased because the noise only exists in the output $\tilde{y}(k)$, and $E(\tilde{y}(k)) = y(k)$ holds. Unfortunately, in model (18), the noise is involved in the new input $\tilde{y}(k)$, and the identification becomes a biased estimation. To tackle the problem, the bias of model (18) is given firstly.

Theorem 1: The bias of model (18) is given as

$$\delta(k) = \sum_{\substack{i=2,q\ge 1\\ p+q=i}}^{I} \sum_{\substack{0\le n_1\le\cdots\le n_p\le N+N_f\\ 0\le n_{p+1}\le\cdots\le n_i\le N}} \times \left[d_{p,q}(n_1,\ldots,n_{p+q}) \prod_{s=1}^{p} u(k-n_s) \sum_{j_0+j_1+\cdots+j_l=q} \times \left(\prod_{j=1}^{j_0} y(k-n_{i,j}) \times \prod_{\substack{s=1,\\ j_s \text{ is even}, j_s\ge 2}}^{l} \mu(j_s)\,\sigma^{j_s} \right) \right] \tag{20}$$

and

$$\prod_{j=1}^{j_0} y(k-n_{i,j}) = E\left(\prod_{j=1}^{j_0} \tilde{y}(k-n_{i,j})\right) \sum_{r_0+r_1+\cdots+r_l=j_0} - \left(\prod_{r=1}^{r_0} y(k-n_{j,r}) \times \prod_{\substack{s=1,\\ r_s \text{ is even}, r_s\ge 2}}^{l} \mu(j_s)\,\sigma^{r_s} \right), \tag{21}$$

where σ is the standard derivation of the noise v, and $\{\{n_{i,1},\ldots,n_{i,j_0}\},\{n_{i,j_0+1},\ldots,n_{i,j_0+j_1}\},\ldots,\{n_{i,j_0+\cdots+j_{l-1}+1},\ldots,n_{i,j_0+\cdots+j_l}\}\}$ is a partition of the set $\{n_{p+1},\ldots,n_{p+q}\}$ where $n_{i,j_0+1} = \cdots = n_{i,j_0+j_1},\ldots,$ and $n_{i,j_0+\cdots+j_{l-1}+1} = \cdots = n_{i,j_0}+\cdots+j_{l-1}+j_l$. $j_1,\ldots,j_l$ are even numbers larger than 2, and $j_0+\cdots+j_l = q$. Denote $\{j_0,\ldots,j_l\}$ a partition of set $\{n_1,\ldots,n_i\}$, the summation $\sum_{j_0+\cdots+j_l=q}(\cdot)$ stands for the sum of $(\cdot)$ on all partitions of the set $\{n_{p+1},\ldots,n_{p+q}\}$, $\prod_{j=1}^{0}(\cdot) = 1$, and

$$\mu(j_s) = 1\times 3\times 5\times\cdots\times(j_s-1). \tag{22}$$

It can be observed that (20) and (21) form a recursion calculation because a high-order product $\prod_{j=1}^{j_0}(\cdot)$ in (20) can be recursively calculated using the low-order product $\prod_{r=1}^{r_0}(\cdot)$ in (21), i.e., $j_0 > r_0$. The recursion terminates with the following condition:

$$\begin{aligned} &y(k-n_{2,1})y(k-n_{2,2}) = E\left(\tilde{y}(k-n_{2,1})\tilde{y}(k-n_{2,2})\right) \\ &\quad - E\left(v(k-n_{2,1})v(k-n_{2,2})\right) \\ &= \begin{cases} E\left(\tilde{y}(k-n_{2,1})\tilde{y}(k-n_{2,2})\right), & n_{2,1} \neq n_{2,2}, \\ E\left(\tilde{y}(k-n_{2,1})\tilde{y}(k-n_{2,2})\right) - \sigma^2, & n_{2,1} = n_{2,2}. \end{cases} \end{aligned} \tag{23}$$

Proof: For model (18), the following equation holds,

$$\begin{aligned} E\left(\varepsilon(k)\right) &= \sum_{i=1}^{I_f} \sum_{0\le n_1\le\cdots\le n_i\le N_f} c_i(n_1,\ldots,n_i) \\ &\quad \times u(k-n_1)\cdots u(k-n_i) + \sum_{\substack{i=1,q\ge 1\\ p+q=i}}^{I} \sum_{\substack{0\le n_1\le\cdots\le n_p\le N+N_f\\ 0\le n_{p+1}\le\cdots\le n_i\le N}} \\ &\quad \times d_{p,q}(n_1,\ldots,n_{p+q}) \prod_{s=1}^{p} u(k-n_s) E\left(\prod_{s=p+1}^{p+q}\tilde{y}(k-n_s)\right), \end{aligned} \tag{24}$$

where $E(\cdot)$ denotes the expectation of $(\cdot)$, and

$$\begin{aligned} &E\left(\tilde{y}(k-n_{p+1})\times\cdots\times\tilde{y}(k-n_{p+q})\right) \\ &= E\left\{\prod_{s=p+1}^{p+q}\left[y(k-n_s)+v(k-n_s)\right]\right\} \\ &= \prod_{s=p+1}^{p+q} y(k-n_s) + E\left(\prod_{s=p+1}^{p+q} v(k-n_s)\right) \\ &\quad + \sum_{i_1+i_2=q}\left[\prod_{s_1=1}^{i_1} y(k-n_{i,s_1}) \times E\left(\prod_{s_2=1}^{i_2} v(k-n_{i,i_1+s_2})\right)\right]. \end{aligned} \tag{25}$$

The problem then becomes the high-order moments of the noise $v(k)$, which can be given as [48]:

$$E(v^s(k)) = \begin{cases} 0 & s \text{ is odd}, \\ \mu(s)\sigma^s & s \text{ is even}, \end{cases} \tag{26}$$

where $\mu(s) = 1\times 3\times\cdots\times(s-1)$. Substituting (26) into (25), and note that (10) holds, the bias $\delta(k) = E(\varepsilon(k))$ can be obtained.

Let $q = j_0$ and $p = 0$, and substitute these two values into (25), then (21) is straightforward. This completes the proof. □

Example 1: Given $\tilde{g} = d_4(1,1,1,1)\tilde{y}(k-1)^4$. All the partitions of the set $\{n_1,n_2,n_3,n_4\}$ can be given as $\{\{n_1,n_2,n_3,n_4\}\}$, $\{\{n_1,n_2\},\{n_3,n_4\}\}$, $\{\{n_1,n_3\},\{n_2,n_4\}\}$, $\{\{n_1,n_4\},\{n_2,n_3\}\}$, $\{\{n_2,n_3\},\{n_1,n_4\}\}$, $\{\{n_2,n_4\},\{n_1,n_3\}\}$, and $\{\{n_3,n_4\},\{n_1,n_2\}\}$. Then the bias can be obtained as $\delta(k) = \mu(4)\sigma^4 + 6\times y(k-1)^2\mu(2)\sigma^2 = 3\sigma^4 + 6y(k-1)^2\sigma^2$. Note that $y(k-1)^2 = E(\tilde{y}(k-1)^2) - \sigma^2$ holds. Therefore, $\delta(k) = -3\sigma^4 + 6E(\tilde{y}(k-1)^2)\sigma^2$.

Theorem 2: The spectrum of the expectation of the unbiased error function $E(\varepsilon(k)-\delta(k))$ can be given as a first-order polynomial with respect to the parameters to be identified as

$$\begin{aligned} \Psi(\omega) &= F\left[E\left(\varepsilon(k)\right) - \delta(k)\right] \\ &= \sum_{i=1}^{I_f} \sum_{0\le n_1\le\cdots\le n_i\le N_f} c_i(n_1,\ldots,n_i)\varphi_{n_1,\ldots,n_i}(\omega) \\ &\quad + \sum_{\substack{i=1,p\ge 1\\ p+q=i}}^{I} \sum_{\substack{0\le n_1\le\cdots\le n_p\le N+N_f\\ 0\le n_{p+1}\le\cdots\le n_i\le N}} d_{p,q}(n_1,\ldots,n_{p+q}) \\ &\quad \times \phi_{n_1,\ldots,n_{p+q}}(\omega), \end{aligned} \tag{27}$$

where $c_i(\cdot)$ and $d_{p,q}(\cdot)$ are the model parameters to be identified, and $F(\cdot)$ denotes the Fourier transform, and the coefficients of the first-order polynomial can be obtained as

$$\varphi_{n_1,\ldots,n_i}(\omega) = \sum_{\omega_1+\cdots+\omega_i=\omega} \prod_{s=1}^{i}\left(e^{-jn_s\omega_s}U(\omega_s)\right), \tag{28}$$

$$\begin{aligned} \phi_{n_1,\ldots,n_i}(\omega) &= \sum_{\substack{\omega_1+\cdots+\omega_{p+q}=\omega\\ p+q=i}} \left[\prod_{s=1}^{p}\left(e^{-jn_s\omega_s}U(\omega_s)\right)\right. \\ &\quad \left.\times E\left(\prod_{s=p+1}^{p+q}\left(e^{-jn_s\omega_s}\tilde{Y}(\omega_s)\right)\right)\right] - \sum_{\substack{\omega_1+\cdots+\omega_p+\\ \omega_{i,1}+\cdots+\omega_{i,j_0}=\omega}} \\ &\quad \times \left[\prod_{s=1}^{p}\left(e^{-jn_s\omega_s}U(\omega_s)\right) \times \sum_{j_0+j_1+\cdots+j_l=q}\right. \\ &\quad \left.\times\left(G_{j_0}(\omega_{i,1},\ldots,\omega_{i,j_0})\left(\prod_{\substack{s=1,\\ j_s \text{ is even},j_s\ge 2}}^{l}\mu(j_s)\sigma^{j_s}\right)\right)\right], \end{aligned} \tag{29}$$

where

$$\begin{aligned} G_{j_0}(\omega_{i,1},\ldots,\omega_{i,j_0}) &= E\left(\prod_{s=1}^{j_0}\left(e^{-jn_{i,s}\omega_{i,s}}\tilde{Y}(\omega_{i,s})\right)\right) \\ &\quad - \sum_{r_0+r_1+\cdots+r_l=j_0}\left(G_{r_0}(\omega_{j_0,1},\ldots,\omega_{j_0,r_0})\right. \\ &\quad \left.\times \prod_{\substack{s=1,\\ r_s \text{ is even},r_s\ge 2}}^{l}\mu(r_s)\sigma^{r_s}\right), \end{aligned} \tag{30}$$

$$G_2(\omega_{2,1},\omega_{2,2}) == \begin{cases} E\left(\prod_{s=1}^{2}\left(e^{-jn_{2,s}\omega_{2,s}}\tilde{Y}(\omega_{2,s})\right)\right), & n_{2,1}\neq n_{2,2}, \\ E\left(\prod_{s=1}^{2}\left(e^{-jn_{2,s}\omega_{2,s}}\tilde{Y}(\omega_{2,s})\right)\right)-\sigma^2, & n_{2,1}=n_{2,2}, \end{cases} \tag{31}$$

and $U(\omega)$ and $\tilde{Y}(\omega)$ are the spectra of $u(k)$ and the noised output $\tilde{y}(k)$, respectively. $E(\cdot)$ denotes the expectation of $(\cdot)$.

Proof: Note that the unbiased error function can be given as

$$\begin{aligned} E(\varepsilon(k))-\delta(k) &= \sum_{i=1}^{I_f}\sum_{0\le n_1\le\cdots\le n_i\le N_f} c_i(n_1,\ldots,n_i) \\ &\times u(k-n_1)\cdots u(k-n_i)+\sum_{\substack{i=1,q\ge 1\\ p+q=i}}^{I}\sum_{\substack{0\le n_1\le\cdots\le n_p\le N+N_f\\ 0\le n_{p+1}\le\cdots\le n_i\le N}} \\ &\times d_{p,q}(n_1,\ldots,n_{p+q})\prod_{s=1}^{p}u(k-n_s)E\left(\prod_{s=p+1}^{p+q}\tilde{y}(k-n_s)\right) \\ &-\delta(k), \end{aligned} \tag{32}$$

which can be considered as a double-input single-output nonlinear system because both of $u(k)$ and $\tilde{y}(k)$ are measurable. The following equations hold [49]:

$$\begin{aligned} F(x(k))|_{\omega} = X(\omega) &= \sum_{i=1}^{I_f}\sum_{0\le n_1\le\cdots\le n_i\le N_f} c_i(n_1,\ldots,n_i) \\ &\times \sum_{\omega_1+\cdots+\omega_i=\omega} e^{-jm_1\omega_1}U(\omega_1)\times\cdots\times e^{-jm_i\omega_i}U(\omega_i) \end{aligned}$$

and

$$\begin{aligned} F\left(\prod_{s=1}^{i}y(k-n_s)\right)\Big|_{\omega} &= \sum_{\omega_1+\cdots+\omega_i=\omega} e^{-jn_1\omega_1}Y(\omega_1)\times\cdots \\ &\times e^{-jn_i\omega_i}Y(\omega_i). \end{aligned}$$

Taking the Fourier transform on both sides of (32) and substituting the above two equations into the Fourier transform lead to the results in (27)–(29). For (30) and (31), by taking the Fourier transform on both sides of (21) and (23), respectively, and then following the similar procedures for proving (27)–(29), the results can thereafter be obtained. This completes the proof. □

Remark 6: It is worth noting that the spectrum of the expectation of the unbiased error function is a first-order polynomial of the model parameters to be identified (i.e., $\boldsymbol{\theta}$), and the coefficients of the first-order polynomial, $\varphi_{n_1,\ldots,n_i}(\omega)$ and $\phi_{n_1,\ldots,n_i}(\omega)$, are both independent of $\boldsymbol{\theta}$. According to [50], (27) is therefore a convex function with respect to the model parameters to be identified, $\boldsymbol{\theta}$, which means that the optimization of (27) at each given frequency ω is a convex optimization problem.

B. Frequency Domain Based Modeling of Multi-Objective Convex Optimization

Given the output spectrum of the expectation of the unbiased error function at frequency ω, i.e., $\Psi(\omega)$ in (27), the identification problem (18), (19) can then be investigated in the frequency domain based on the L_∞-norm by minimizing the maximum error over the entire frequency range Ω,

$$\arg\min_{\boldsymbol{\theta}} \max\{|\Psi(\omega)||\forall\omega\in\Omega\}, \tag{33}$$

where $|\Psi(\omega)|$ denotes the module of $\Psi(\omega)$. Problem (33) is a convex optimization problem with $\Psi(\omega)$ given in (27).

Proof: From Remark 5, it is shown that $\Psi(\omega)$ is a convex function with respect to the model parameters to be identified, $\boldsymbol{\theta}$, and so is $|\Psi(\omega)|$. The pointwise maximum $\max\{|\Psi(\omega)||\forall\omega\in\Omega\}$ is a convex function [50]. Problem (33) then comes to minimize a convex function over a convex region, and therefore is a convex optimization problem. This completes the proof. □

Remark 7: The convex optimization problem (33) needs further tactic treatment in order to get an effective and efficient algorithm. Noted that the objective function in (33) is nondifferentiable, the module $|\Psi(\omega)|$ at each frequency ω is differentiable exclude the singularities. A naive gradient based method, e.g., in each step choose the gradient of $|\Psi(\omega_s)|$ that $|\Psi(\omega_s)|=\max\{|\Psi(\omega)||\forall\omega\in\Omega\}$, may lead to the 'jump' phenomenon, i.e., the sudden increase of the cost function between two consecutive iterations, which is a key issue affecting the convergence performance of the multi-objective optimization problem.

In the following, the game theory is introduced to model the behaviors of the optimization process and to establish an effective and efficient algorithm for the multi-objective optimization problem (33).

C. Bargaining Game Theory Based Multi-Objective Optimization

Given the multi-objective optimization problem (33), classical optimization algorithms can well employ the local information to update the coefficients, by using, for example, the gradient based strategy. Problems, however, may occur in such optimization process such as the jump phenomenon, resulting from the lack of global information in the coefficient updating process. Evolution algorithms may offer an alternative to the problem, but most of them such as the genetic algorithm cannot make full use of the local information, detrimental to the local searching. In this section, the bargaining game theory is used to overcome such problems through a simultaneous consideration of both the global and local information. Firstly, the definition of the game model for a multi-objective optimization problem (33) is given. Then the game theory based behavior modeling of the interactions and competitions among all multiple objectives is presented, which provides a feasible way to achieve the optimal strategy (i.e., the estimated optimal coefficients). The Nash equilibrium of the behavior modeling is demonstrated.

Finally, a game theory based algorithm is summarized for the multi-objective optimization problem (33).

Definition 1: The multi-player bargaining game model involves:

Players: The frequency ω in the frequency range Ω, i.e., $\forall\omega \in \Omega$. Denote the number of players as λ.

Strategy set: $\boldsymbol{\theta} \in R^{\varpi}$, where ϖ denotes the dimension of the coefficients $\boldsymbol{\theta}$ with R being a real number set.

Utility set: $\{\Phi(\omega)|\forall\omega \in \Omega\}$, where $\Phi(\omega)$ is the utility of player ω given as

$$\Phi(\omega) = 1 - |\Psi(\omega)| \tag{34}$$

and the system utility is defined as

$$\Phi = \arg\max_{\boldsymbol{\theta}} \min_{\omega\in\Omega} \Phi(\omega) = \arg\min_{\boldsymbol{\theta}} \max_{\omega\in\Omega} |\Psi(\omega)|. \tag{35}$$

Assumption 2: In order to guarantee an agreement in the bargaining game, it is assumed that $\Phi(\omega) = 0, \forall\omega \in \Omega$ holds if player ω decides not to enroll in the bargaining game.

Behavior modeling: In the s th-round of the bargaining, the behavior can be modeled as:

Step 1: Given the strategy $\boldsymbol{\theta}_s$ and the indicator κ, each player ($\forall\omega_\alpha \in \Omega$) calculates its utility $\Phi_s(\omega_\alpha)$. If $\Phi_s(\omega_\alpha) < \kappa$ holds, player ω_α generates its new strategy

$$\boldsymbol{\theta}_{s+1}(\omega_\alpha) = \boldsymbol{\theta}_s(\omega_\alpha) + \boldsymbol{r}\triangle_\theta \tag{36}$$

such that $\Phi_{s+1}(\omega_\alpha) > \Phi_s(\omega_\alpha)$ holds, where $\boldsymbol{r} = [r_1, \ldots, r_{\varpi}]^{\mathrm{T}}$ is a random vector with r_s randomly generated in $[-1, 1], s = 1, \ldots, \varpi$, and $\triangle_\theta$ is a given constant step; otherwise, the new strategy for player ω_α should reduce its utility, i.e., $\Phi_{s+1}(\omega_\alpha) < \Phi_s(\omega_\alpha)$.

Step 2: Each player ω_α sends their new strategy $\boldsymbol{\theta}_{s+1}(\omega_\alpha)$ to all other players $\forall\omega_\beta \in \Omega$, $\omega_\beta \neq \omega_\alpha$. Each player ω_β should calculate its utility $\Phi_s(\omega_\alpha; \omega_\beta)$ based on the player ω_α's strategy and feedback to player ω_α.

Step 3: Each player ω_α calculates the minimum utility of its own strategy, i.e.,

$$\Phi_{s+1,\min}(\omega_\alpha) = \min_{\forall\omega_\beta\in\Omega} \Phi_s(\omega_\alpha; \omega_\beta). \tag{37}$$

Step 4: All the players negotiate and choose the maximum utility, i.e.,

$$\Phi_{s+1,\max} = \max_{\omega_\alpha\in\Omega} \Phi_{s+1,\min}(\omega_\alpha). \tag{38}$$

If $\Phi_{s+1,\max} > \Phi_{s,\max}$, choose the corresponding player's strategy as the strategy of this round's bargaining, i.e.,

$$\boldsymbol{\theta}_{s+1} = \boldsymbol{\theta}_{s+1}(\omega_\alpha), \tag{39}$$

and calculate the indicator as

$$\kappa_{s+1} = \sum_{\omega_\beta\in\Omega} \Phi_{s+1}(\omega_\alpha; \omega_\beta) \Big/ \lambda, \tag{40}$$

where λ is the number of the players; go to step 5. Otherwise, if $\Phi_{s+1,\max} < \Phi_{s,\max}$, the new strategy of this round should be directly rejected, then go to step 1 for generating new strategies.

Step 5: End of the round's negotiation.

Remark 8: According to (38) in step 4, the maximum system utility $\Phi_{s+1,\max}$ is chosen among all the players' utilities. Therefore, the new strategy $\boldsymbol{\theta}_{s+1}$ in (39) and the indicator κ_{s+1} in (40) would involve the global information. In step 1, each player generates its own new strategy independently via a local searching, and obviously, the local information is employed. Both the global and local information are considered in the multi-objective optimization problem, which would improve the convergence performance of the optimization.

Remark 9: The indicator κ in (40) represents an approximation of the targeted system utility. For $\Phi_s(\omega_\alpha) > \kappa$ in step 1, if player ω_α still chooses to maximize its own utility, the corresponding strategy will have a larger chance to reduce the other players' utilities, and consequently will reduce the system utility. Other players would have a greater chance to reject this strategy. In order to help achieve an agreement, player ω_α should choose its new strategy such that $\Phi_{s+1}(\omega_\alpha) < \Phi_s(\omega_\alpha)$ holds.

Theorem 3: After several rounds of negotiations, the multi-player bargaining game finally converges to the Nash equilibrium, which means that no unilateral deviation in the strategy by any single player ω_α is profitable for that player, i.e.,

$$\Phi(\omega_\alpha|\boldsymbol{\theta}^*(\omega_\alpha), \boldsymbol{\theta}^*(\omega_{-\alpha})) \geq \Phi(\omega_\alpha|\boldsymbol{\theta}(\omega_\alpha), \boldsymbol{\theta}^*(\omega_{-\alpha})), \quad \forall\omega_\alpha \in \Omega, \tag{41}$$

where $\boldsymbol{\theta}^*(\omega_\alpha)$ is the optimal strategy set of the player ω_α, $\boldsymbol{\theta}^*(\omega_{-\alpha})$ is the optimal strategy set of all other players except for player ω_α, and $\Phi(\omega_\alpha|\cdot)$ is the utility of player ω_α with strategy set $\cdot$. The strategy $\boldsymbol{\theta}^*(\omega_1) = \cdots = \boldsymbol{\theta}^*(\omega_\lambda)$ is the optimal solution of the multi-objective optimization problem (33).

Proof: Suppose that there exists a strategy $\boldsymbol{\theta}(\omega_\alpha)$ such that $\Phi(\omega_\alpha|\boldsymbol{\theta}(\omega_\alpha), \boldsymbol{\theta}^*(\omega_{-\alpha})) \geq \Phi(\omega_\alpha|\boldsymbol{\theta}^*(\omega_\alpha), \boldsymbol{\theta}^*(\omega_{-\alpha}))$ for all strategies $\boldsymbol{\theta}^*(\omega_{-\alpha})$ but it is not a Nash equilibrium. Resulting from the competition characteristic among all the players in the multi-objective optimization, at least one of the player's utilities will be decreased by unilaterally deviating from the strategy, namely:

$$\Phi(\omega_\beta|\boldsymbol{\theta}(\omega_\alpha), \boldsymbol{\theta}^*(\omega_{-\alpha})) \leq \Phi(\omega_\beta|\boldsymbol{\theta}^*(\omega_\alpha), \boldsymbol{\theta}^*(\omega_{-\alpha})), \quad \exists\omega_\beta \in \Omega, \omega_\beta \neq \omega_\alpha,$$

which means that the minimum utility has been decreased. Note that the goal of the optimization is to maximize the minimum utility. Therefore, in the negotiation (step 4), the new strategy will be chosen as $\boldsymbol{\theta}_{s+1} \in \boldsymbol{\theta}^*(\omega_{-\alpha})$, and $\boldsymbol{\theta}(\omega_\alpha)$ will be rejected. The assumption does not hold and none of the players can benefit by unilaterally deviating its strategy. Therefore, the strategy $\boldsymbol{\theta}^*(\omega_1) = \cdots = \boldsymbol{\theta}^*(\omega_\lambda)$ is the Nash equilibrium of the multi-objective bargaining game model. This completes the proof. □

The following algorithm is summarized to provide a step-by-step negotiation process to achieve the optimal strategy of the multi-player bargaining game model, and so is the optimal estimate of the multi-objective frequency-domain based identification problem (33).

In the initialization, the coefficients $\boldsymbol{\theta}_0$ and the indicator κ_0 can be randomly given. After several rounds of bargaining,

TABLE I
ALGORITHM TO ACHIEVE THE NASH EQUILIBRIUM

Algorithm1: Achieving the optimal strategy of the multi-player bargaining game model

1. **Initialization:** s=1, given $\boldsymbol{\theta}_s = \boldsymbol{\theta}_0$, $\kappa = \kappa_0$, the step $\triangle_\theta$ for updating the strategy $\boldsymbol{\theta}_s$, and the targeted system utility Φ_{targeted}.
2. **Bargaining:**
   while $\Phi_s < \Phi_{\text{targeted}}$
   2.1 each player generates its new strategy $\boldsymbol{\theta}_{s+1}(\omega_\alpha)$ according to **(36)** in step 1 in the behavior modeling. The player's utility is given in **(34)** where the corresponding unbiased error function is given in **(27)**.
   2.2 each player calculates the minimum utility $\Phi_{s+1,\min}(\omega_\alpha)$ of its own strategy $\boldsymbol{\theta}_{s+1}(\omega_\alpha)$ according to **(37)**.
   2.3 calculate the maximum system utility $\Phi_{s+1,\max}$ according to **(38)**. If $\Phi_{s+1,\max} < \Phi_{s,\max}$, the strategy is directly rejected, then go to 2.1; otherwise, update the strategy $\boldsymbol{\theta}_{s+1}$ and the indicator κ_{s+1} according to **(39)** and **(40)**, respectively, and s=s+1.
   end

Algorithm 1 will converge to the Nash equilibrium, and the optimal estimation can therefore be obtained. Owing to the use of the global information, the system utility in Algorithm 1 increases successively as the iterative bargaining goes on, that is, the maximum error in (33) monotonously decreases, which will greatly improve the convergence performance of the multi-objective identification problem (33).

Remark 10: Although finding the Nash equilibrium in a game is a NP problem, Algorithm 1 together with the behavior modelling can achieve the Nash equilibrium. The reasons are: 1) the multi-objective optimization problem in (33) (or (35)) is a convex problem, so there only exists a single Nash equilibrium in the game; 2) according to Algorithm 1 and the behavior modelling, a higher utility (i.e., a smaller estimate error at the frequencies of interest) can be guaranteed after each negotiation until reach the Nash equilibrium, which means that the current strategy gets closer to the Nash equilibrium after each negotiation, and after several rounds of negotiations (can be determined by the targeted utility Φ_{targeted} or the estimate error acceptable at the frequencies of interest), the strategy can reach the Nash equilibrium.

Remark 11: The bargaining game theory algorithm (BGTA) is proposed to solve the multi-objective optimization problem where the objective function given in (33) is based on L_∞-norm. There also exist other criteria to characterize the objective function, e.g., the L_1-norm or L_2-norm that integrate the module or squared module over the entire frequency range. With these norms, the estimate error at the frequencies of interest would be more conservative (much larger than the estimate error provided by the proposed BGTA method) because it is a total estimate error that integrates the error over the whole frequency range. If the L_1-norm or L_2-norm is selected as the criterion to formulate the objective function, the proposed BGTA method cannot be directly applied to the single-objective optimization problem. We will consider these criteria in our further study.

Remark 12: To guarantee a unique solution, the frequencies taken into account in the identification problem (33) (i.e., the number of players considered in the bargaining game) should be greater than or at least equal to the number of model parameters to be identified.

Remark 13: In the above, we demonstrate that the bargaining game theory based multi-objective optimization is a convex problem and a unique optimal solution exists. Generally, a small step $\triangle_\theta$ can ensure the convergence of the numerical simulation to the optimal solution with a large number of bargaining, and a large step $\triangle_\theta$ with a small number of bargaining. If the step $\triangle_\theta$ is too large, the numerical simulation may not converge to the optimal solution. It is difficult to provide an analytical and explicit criterion for the step $\triangle_\theta$ to guarantee the convergence of the numerical simulation. Instead, we can give two or more different steps $\triangle_\theta$, if the solutions of these different steps are the same, the numerical simulations can be considered as converging to the optimal solution. Otherwise, we can reduce the step $\triangle_\theta$, until the results using different steps are the same.

IV. NUMERICAL SIMULATIONS AND DISCUSSIONS

Examples are investigated using the cascade dynamic nonlinear system described by (42)–(44). Note model (42) is a Volterra model, and model (43) a NARX model, which can be merged to the general system (10)–(14) that describes the cascade dynamic nonlinear systems.

$$x(k) = 0.6u(k) + 0.2u(k-2) + 0.6u(k)u(k-1), \quad (42)$$

$$\begin{aligned} y(k) = {} & x(k) + 0.4y(k-1)y(k-2) + 0.6y(k-1)u(k) \\ & - 0.1y(k-1)u(k-2) - 0.1y(k-1)u(k-4) \\ & + 0.6y(k-1)u(k)u(k-1) - 0.3y(k-1) \\ & \times u(k-2)u(k-3), \end{aligned} \quad (43)$$

$$\tilde{y}(k) = y(k) + v(k), \quad (44)$$

where

$$c_1(0) = 0.6, c_1(2) = 0.2, c_2(0,1) = 0.6, d_{0,1}(0) = -1,$$
$$d_{0,2}(1,2) = 0.4, d_{1,1}(0,1) = 0.6, d_{1,1}(2,1) = -0.1,$$
$$d_{1,1}(4,1) = -0.1, d_{2,1}(0,1,1) = 0.6, d_{2,1}(2,3,1) = -0.3.$$

The input is given as

$$u(k) = 0.3\sin(2\pi T_s k) + 0.3\sin(4\pi T_s k + \pi/2), \quad (45)$$

where $T_s = 0.025$ s is the sampling time interval. In typical structural health monitoring applications, the received signal has a signal-to-noise ratio (SNR) around 60 dB [51]. Therefore, the SNR of the white Gaussian noise $v(k)$ is firstly set to be

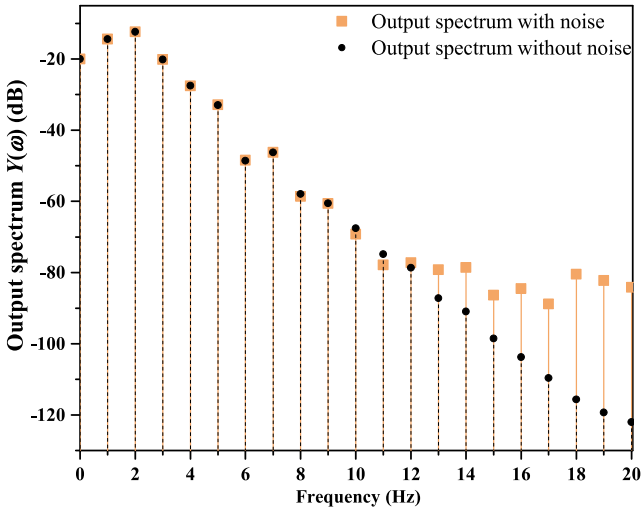

Fig. 2. Output spectrum $Y(\omega)$.

60 dB in the simulation. The effect of signals with different SNR will be studied later.

Adopting least square method mentioned in Remark 5, $\hat{\boldsymbol{\theta}} = (\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\varepsilon}$, and considering a simple case that only identifies the first sub-system with $I_f = 2$ and $N_f = 2$, the matrix $\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Theta}$ is singular. Therefore, the least square method is unavailable.

Fig. 2 shows the output spectra, $Y(\omega)$, with and without noise, respectively. It can be observed that there exist multiple harmonics in the frequency range of $Y(\omega)$ (only the first 20 Hz range is shown in Fig. 2 because the sampling time interval is 0.025 s with a data length of 1 s) although only two frequencies are involved in the input (shown in (45), i.e., 1 Hz and 2 Hz), due to the nonlinearities involved in the first and second subsystems (42)–(44). As the output frequency increases, the output spectrum without noise shows a downward tendency. For the noised output case, the output spectrum follows the same descending trend for the lower-order frequency components, before tending to a constant as the output frequency increases. The reason is that the noise in the numerical study is a Gaussian white noise, which has a constant power spectra density.

Fig. 3 shows the identification results of the cascade dynamic nonlinear system with three different initial model parameters. The expectation of the noised output spectrum $E(\tilde{Y}(\omega))$ and $E(\tilde{Y}(\omega_1)\tilde{Y}(\omega_2))$ in the proposed BGTA method are obtained via 100 measurements. It can be seen that after several rounds of negotiations, the estimated coefficients converge to their respective true values, which also indicates that the proposed BGTA converges to the Nash equilibrium in the multi-objective optimization.

Fig. 4 shows the comparison of the convergence performance between the proposed BGTA and the gradient based naive method mentioned in Remark 7. It can be seen that the gradient based naive method has obvious 'jump' phenomenon between two consecutive iterations (shown in the sub-figure), i.e., the estimate error suddenly increases to a large error in the iteration, which deteriorates the convergence performance of the optimization. This is due to the fact that only local information

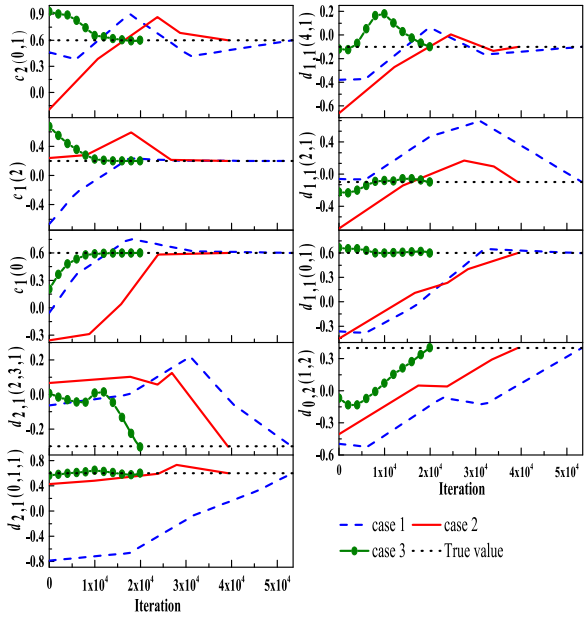

Fig. 3. Identification of the cascade dynamic nonlinear system based on the bargaining game modeling with different initial model parameters.

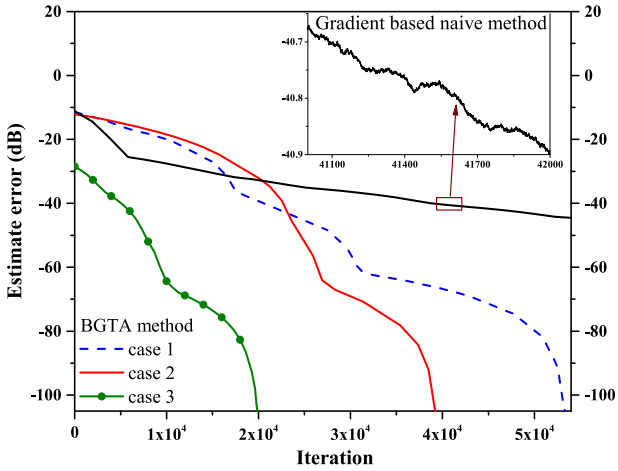

Fig. 4. Comparisons of the convergence performance of the proposed BGTA and the gradient based naive method.

is considered in the optimization (the gradient is employed). This 'jump' phenomenon will be even worse when the estimate error become smaller. As to the proposed BGTA with different initial coefficients, Fig. 4 shows a much-improved convergence behavior. In fact, the estimate error quickly reduces with the increasing number of the iteration for all cases. Meanwhile, as pointed out in remark 6, the proposed BGTA employs both the local information and the global information in the optimization. Therefore, no 'jump' phenomenon appears, which is also responsible for the improved convergence performance compared with the gradient based naive method. It can also be observed that when the proposed BGTA achieves the optimal strategy, the gradient based naive method is still far from convergence. The effectiveness, efficiency, and the superiority of the proposed game theory algorithm can therefore be demonstrated.

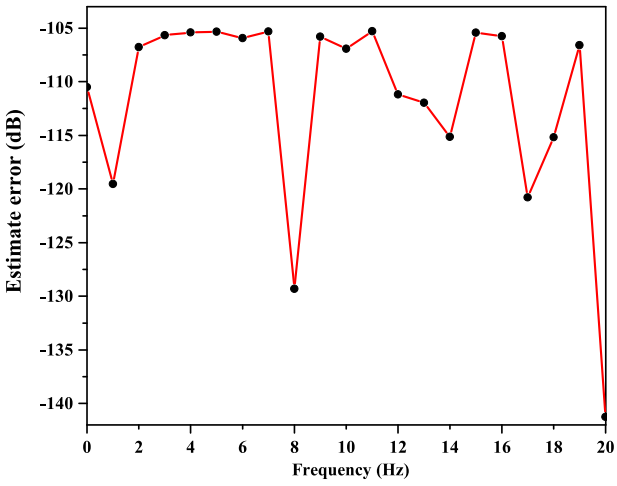

Fig. 5. Estimate error at some frequencies of interest.

It can be observed that the convergence rate (number of iterations to convergence) of the proposed BGTA may depend on the given initial model parameters, e.g., case 3 converges with 2×10^4 iterations, and case 1 with 5.4×10^4 iterations. The given step $\triangle_\theta$ in (36) and the maximum utility Φ_{targeted} in Algorithm 1 (or the maximum estimate error acceptable at the frequencies of interest) will also influence the number of iterations to the convergence. It ranges from twenty to forty minutes for the four numerical simulations in Figs. 3 and 4, which is acceptable as an offline identification algorithm. To improve the efficiency, either the L_1-norm or L_2-norm can be considered to characterize the identification as a single-objective optimization problem, which will only provide a total estimate error in the whole frequency range instead of a tight estimate error at the frequencies of interest that can be obtained by the proposed BGTA method.

Fig. 5 shows the estimate errors at the frequencies of interest. It can be observed that the maximum estimate error is -105 dB at 11 Hz, which is consistent with that shown in Fig. 4. The estimate error at all other frequencies are smaller.

Figs. 6 and 7 show the performance of the proposed BGTA against different noise levels. When the SNR decreases from 40 dB to 30 dB, the estimated model parameters well converge to the true values, demonstrating the robustness of the proposed algorithm. The expectation of the noised output spectrum $E(\tilde{Y}(\omega))$ in the proposed BGTA method are obtained via 100 measurements for all different noise levels. When the SNR is set to be 20 dB, although some of the parameters, e.g., $d_{0,2}(1,2)$, slightly deviate from the true values, most of the parameters still converge to the true values. Therefore, the result is still acceptable. For a SNR lower than 20 dB, more measurements should be used to calculate an accurate expectation of the noised output spectrum $E(\tilde{Y}(\omega))$ to make the estimate results acceptable.

It is shown in Section II that the identification of the cascade dynamic nonlinear systems can be transformed into the identification of a double-input single-output Volterra system. Although the identification of a Volterra system was widely studied in the literature, most of the studies focused on a single-input single-output second-order or third-order Volterra model [34],

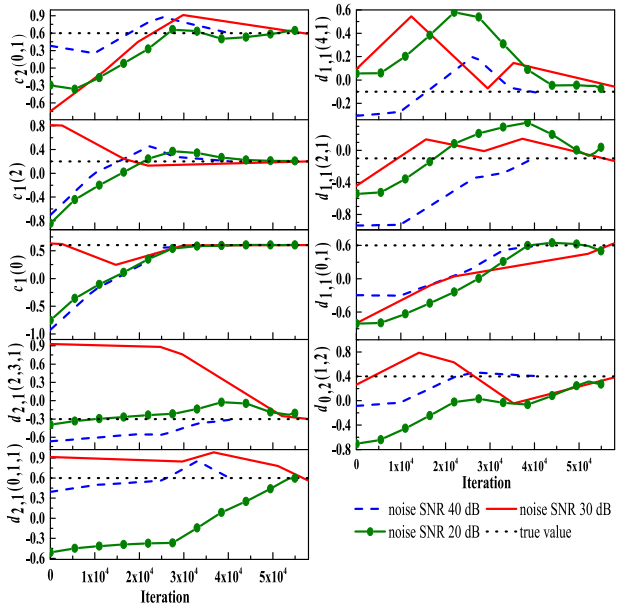

Fig. 6. Identification of the cascade dynamic nonlinear systems at different noise levels.

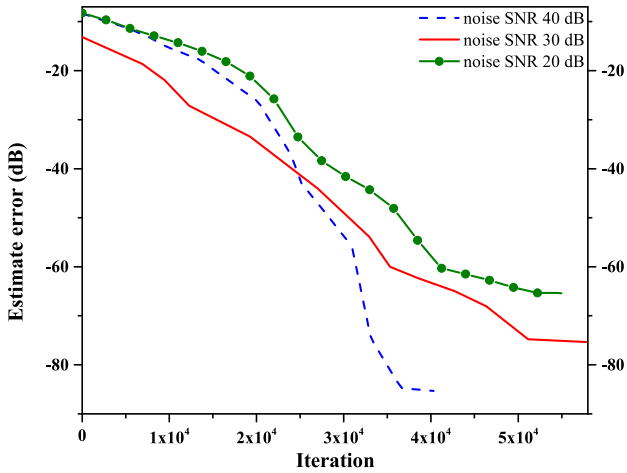

Fig. 7. The convergence performances at different noise levels.

[35], which obviously cannot be directly applied to the present cascade dynamic nonlinear models. In Fig. 8, we compare the present BGTA method with a tensor based method that identifies a multiple-input multiple-output Volterra model [52]. Fig. 8(a) is the identification result of the first sub-system, and Fig. 8(b) is that of the second sub-system. For the second sub-system, the measured output spectrum $\tilde{Y}(\omega)$ and the input spectrum $U(\omega)$ are used to calculate the spectrum of the intermediate variable $|X_{\text{BGTA}}(\omega)|$. Comparing $|X_{\text{BGTA}}(\omega)|$ and $|X_{\text{Tensor}}(\omega)|$ with the true spectrum, we can investigate the effectiveness of different methods on the identification of the first and second subsystems. It can be observed that results from the BGTA method almost overlap with the true spectrum, $|X(\omega)|$, outperforming its tensor based counterpart. Plausible reasons are: 1) the tensor based method is a time domain method, inevitably embracing all the noise in the identification. On the contrary, the BGTA method only takes into account the frequencies shown in Fig. 2

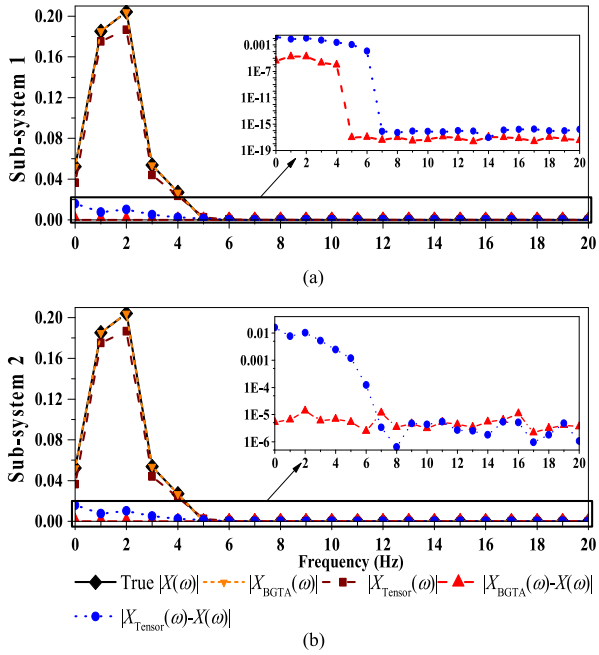

Fig. 8. Comparison with the tensor based method [52]. $|X(\omega)|$ is the magnitude of the true spectrum of intermediate variable $x(k)$, $|X_{\text{BGTA}}(\omega)|$ is the magnitude of the spectrum obtained by the proposed BGTA method, and $|X_{\text{Tensor}}(\omega)|$ is the magnitude of the spectrum obtained by the tensor based method. (a) First sub-system, and (b) second sub-system.

as the frequencies of interest. Therefore, only the noise at those particular frequencies are involved in the identification; 2) a regulation term in the objective function exists in the tensor based method, which may result in an additional error in the identification. Furthermore, the tensor based method can only provide a total estimate error that involves the error at each sampling instance but cannot quantify and guarantee a tight estimate error at the frequencies of interest. BGTA method makes this possible, which is one of distinct contributions of the present work.

V. Conclusions

Cascade dynamic nonlinear systems can be used to model a large class of nonlinear systems in engineering practices, but very few efforts have been devoted to the identification of such systems. The main difficulty comes from the non-convex nature of the identification problem. This paper investigated this problem based on a bargaining game theory model. Firstly, the cascade dynamic nonlinear systems were rearranged and converted into an equivalent DISO system, where the estimate error was considered as the new output, and the system input and measured noised output as the new inputs. By the same token, the identification of the cascade dynamic nonlinear systems was transformed into a convex optimization problem. Secondly, in order to guarantee the estimate error at the frequencies of interest, the DISO identification problem was decomposed into a set of sub-problems defined in the output frequency range. The spectrum of the unbiased estimation error function was then given, which was demonstrated to be a first-order polynomial of the model parameters to be identified and with independent coefficients. This characteristic relationship between the unbiased spectrum and the model parameters to be identified greatly facilitates the optimization in the frequency domain. Reaching this step, the identification problem was transformed into a multi-objective optimization problem. To ensure an effective and efficient multi-objective optimization, the bargaining game theory was employed to model the competitions and interactions among the multiple objectives in the optimization. It was shown that the consideration of both the global and local information in the BGTA greatly improves the convergence performance of the optimization. The proposed formalism allows the identification of the cascade dynamic nonlinear systems to be conducted in the frequency domain such that the nonlinear features can be directly captured with the estimate error being guaranteed at the frequencies of interest, conducive to numerous engineering applications requiring the extraction of nonlinear features related to higher-order harmonics.

References

[1] M. Hong, Z. Su, Q. Wang, L. Cheng, and X. Qing, "Modeling nonlinearities of ultrasonic waves for fatigue damage characterization: Theory, simulation, and experimental validation," *Ultrasonics*, vol. 54, no. 3, pp. 770–778, 2014.

[2] S. Shan, L. Cheng, and P. Li, "Adhesive nonlinearity in lamb-wave-based structural health monitoring systems," *Smart Mater. Struct.*, vol. 26, no. 2, pp. 025019, 2017.

[3] H. Jiang and P. Wilford, "Digital predistortion for power amplifiers using separable functions," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4121–4130, Aug. 2010.

[4] R. Piazza, M. R. B. Shankar, and B. Ottersten, "Data predistortion for multicarrier satellite channels based on direct learning," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5868–5880, Nov. 2014.

[5] E. W. Bai, "Frequency domain identification of Wiener models," *Automatica*, vol. 39, no. 9, pp. 1521–1530, 2003.

[6] S. VanVaerenbergh, J. Vía, and I. Santamaría, "Blind identification of SIMO Wiener systems based on kernel canonical correlation analysis," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2219–2230, May 2013.

[7] G. Li and C. Wen, "Identification of Wiener systems with clipped observations," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3845–3852, Jul. 2012.

[8] L. Vanbeylen, R. Pintelon, and J. Schoukens, "Blind maximum-likelihood identification of Wiener systems," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3017–3029, Aug. 2009.

[9] M. Pawlak, Z. Hasiewicz, and P. Wachel, "On nonparametric identification of Wiener systems," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 482–492, Feb. 2007.

[10] M. Zhang, A. Zhang, and J. Li, "Fast and accurate rank selection methods for multistage Wiener filter," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 973–984, Feb. 2016.

[11] X. Hong, S. Chen, Y. Gong, and C. J. Harris, "Nonlinear equalization of Hammerstein OFDM systems," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5629–5639, Nov. 2014.

[12] C. Yu, C. Zhang, and L. Xie, "A new deterministic identification approach to Hammerstein systems," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 131–140, Jan. 2014.

[13] B. Ni, H. Garnier, and M. Gilson, "A refined instrumental variable method for Hammerstein-Wiener continuous-time model identification," *IFAC Proc. Vol.*, vol. 45, no. 16, pp. 1061–1066, 2012.

[14] B. Q. Mu and H. F. Chen, "Recursive identification of multi-input multi-output errors-in-variables Hammerstein systems," *IEEE Trans. Autom. Control*, vol. 60, no. 3, pp. 843–849, Mar. 2015.

[15] A. Wills, T. B. Schön, L. Ljung, and B. Ninness, "Identification of Hammerstein–Wiener models," *Automatica*, vol. 49, no. 1, pp. 70–81, 2013.

[16] M. Schoukens, E. W. Bai, and Y. Rolain, "Identification of Hammerstein-Wiener systems," *IFAC Proc. Vol.*, vol. 45, no. 16, pp. 274–279, 2012.

[17] E. W. Bai, "An optimal two-stage identification algorithm for Hammerstein–Wiener nonlinear systems," *Automatica*, vol. 34, no. 3, pp. 333–338, 1998.
[18] G. Li, C. Wen, W. X. Zheng, and G. Zhao, "Iterative identification of block-oriented nonlinear systems based on biconvex optimization," *Syst. Control Lett.*, vol. 79, pp. 68–75, 2015.
[19] P. Crama and J. Schoukens, "Hammerstein–Wiener system estimator initialization," *Automatica*, vol. 40, no. 9, pp. 1543–1550, 2004.
[20] A. Brouri, F. Giri, F. Ikhouane, F. Z. Chaoui, and O. Amdouri, "Identification of Hammerstein-Wiener systems with backlash input nonlinearity bordered by straight lines," *IFAC Proc. Vol.*, vol. 47, no. 3, pp. 475–480, 2014.
[21] F. Giri, A. Brouri, F. Ikhouane, F. Z. Chaoui, and A. Radouane, "Identification of Hammerstein-Wiener systems including backlash input nonlinearities," *IFAC Proc. Vol.*, vol. 46, no. 11, pp. 360–365, 2013.
[22] Y. Rochdi, F. Giri, and F. Z. Chaoui, "Frequency identification of nonparametric Hammerstein-Wiener systems with output backlash operator," *IFAC Proc. Vol.*, vol. 45, no. 16, pp. 19–24, 2012.
[23] Y. Rochdi, F. Giri, and F. Z. Chaoui, "Frequency identification of nonparametric Hammerstein-Wiener systems," *IFAC Proc. Vol.*, vol. 44, no. 1, pp. 11190–11195, 2011.
[24] A. Wills and B. Ninness, "Generalised Hammerstein–Wiener system estimation and a benchmark application," *Control Eng. Pract.*, vol. 20, no. 11, pp. 1097–1108, 2012.
[25] A. Wills and B. Ninness, "Estimation of generalised Hammerstein-Wiener systems," *IFAC Proc. Vol.*, vol. 42, no. 10, pp. 1104–1109, 2009.
[26] J. Vörös, "Identification of nonlinear dynamic systems with input saturation and output backlash using three-block cascade models," *J. Franklin Inst.*, vol. 351, no. 12, pp. 5455–5466, 2014.
[27] K. Tiels and J. Schoukens, "Wiener system identification with generalized orthonormal basis functions," *Automatica*, vol. 50, no. 12, pp. 3147–3154, 2014.
[28] R. Castro-Garcia, K. Tiels, J. Schoukens, and J. A. K. Suykens, "Incorporating best linear approximation within LS-SVM-based Hammerstein system identification," in *Proc. IEEE 54th Conf. Decis. Control*, 2015, pp. 7392–7397.
[29] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*. New York, NY, USA: Wiley, 2012.
[30] M. Schoukens, R. Pintelon, and Y. Rolain, "Identification of Wiener–Hammerstein systems by a nonparametric separation of the best linear approximation," *Automatica*, vol. 50, no. 2, pp. 628–634, 2014.
[31] J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," Massachusetts Institute of Technology, Cambridge, MA, USA, Tech. Rep. No. 216, 1952.
[32] I. Sandberg, "Expansions for nonlinear systems," *Bell Syst. Tech. J*, vol. 61, no. 2, pp. 159–199, 1982.
[33] S. Boyd and L. Chua, "Fading memory and the problem of approximating nonlinear operators with Volterra series," *IEEE Trans. Circuits Syst.*, vol. 32, no. 11, pp. 1150–1161, Nov. 1985.
[34] W. Rugh, *Nonlinear System Theory*. Baltimore, MD, USA: Johns Hopkins Univ. Press, 1981.
[35] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. Melbourne, FL, USA: Krieger, 2006.
[36] T. Helie and B. Laroche, "Computation of convergence bounds for Volterra series of linear-analytic single-input systems," *IEEE Trans. Autom. Control*, vol. 56, no. 9, pp. 2062–2072, Sep. 2011.
[37] Z. Xiao, X. Jing, and L. Cheng, "Parameterized convergence bounds for Volterra series expansion of NARX models," *IEEE Trans. Signal Process.*, vol. 61, no. 20, pp. 5026–5038, Oct. 2013.
[38] Z. Xiao, X. Jing, and L. Cheng, "Estimation of parametric convergence bounds for Volterra series expansion of nonlinear systems," *Mech. Syst. Signal Process.*, vol. 45, no. 1, pp. 28–48, 2014.
[39] J. Peyton and S. Billings, "Recursive algorithm for computing the frequency response of a class of non-linear difference equation models," *Int. J. Control*, vol. 50, no. 5, pp. 1925–1940, 1989.
[40] S. Billings and J. Peyton, "Mapping non-linear integro-differential equations into the frequency domain," *Int. J. Control*, vol. 52, no. 4, pp. 863–879, 1990.
[41] Z. Lang and S. Billings, "Output frequency characteristics of nonlinear systems," *Int. J. Control*, vol. 64, no. 6, pp. 1049–1067, 1996.
[42] Z. Lang, S. Billings, and R. Yue, "Output frequency response function of nonlinear Volterra systems," *Automatica*, vol. 43, no. 5, pp. 805–816, 2007.
[43] X. Jing, Z. Lang, and S. Billings, "Mapping from parametric characteristics to generalized frequency response functions of non-linear systems," *Int. J. Control*, vol. 81, no. 7, pp. 1071–1088, 2008.
[44] Z. Xiao and X. Jing, "Frequency-domain analysis and design of linear feedback of nonlinear systems and applications in vehicle suspensions," *IEEE/ASME Trans. Mechatronics*, vol. 21, no. 1, pp. 506–517, Feb. 2016.
[45] Z. Xiao and X. Jing, "A novel characteristic parameter approach for analysis and design of linear components in nonlinear systems," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2528–2540, May 2016.
[46] G. Li, C. Wen, W. Zheng, and Y. Chen, "Identification of a class of nonlinear autoregressive models with exogenous inputs based on kernel machines," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2146–2159, May 2011.
[47] E. Bai and K. Chan, "Identification of an additive nonlinear system and its applications in generalized Hammerstein models," *Automatica*, vol. 44, no. 2, pp. 430–436, 2008.
[48] V. Krishnan and K. Chandra, *Probability and Random Processes*. New York, NY, USA: Wiley, 2015.
[49] A. Swain and S. Billings, "Generalized frequency response function matrix for MIMO non-linear systems," *Int. J. Control*, vol. 74, no. 8, pp. 829–844, 2001.
[50] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[51] Q. Wang, M. Hong, and Z. Su, "An in-situ structural health diagnosis technique and its realization via a modularized system," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 4, pp. 873–887, Apr. 2015.
[52] K. Batselier, Z. Chen, and N. Wong, "Tensor network alternating linear scheme for MIMO Volterra system identification," *Automatica*, vol. 84, pp. 26–35, 2017.

Zhenlong Xiao (M'16) received the B.S. degree from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2008, the M.S. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2011, and the Ph.D. degree from Hong Kong Polytechnic University, Hong Kong, China, in 2015. He is currently an Assistant Professor with the Department of Communication Engineering, School of Information Science and Engineering, Xiamen University, Xiamen, China. His research interests include the analysis and design of nonlinear system, nonlinear signal processing, and system identification.

Shengbo Shan received the B. Eng. and M. Eng. degrees from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2012 and 2015, respectively. He is currently a Ph.D. candidate with the Department of Mechanical Engineering, Hong Kong Polytechnic University, Hong Kong.

His research interests include structural health monitoring, guided wave theory, and signal processing.

Li Cheng received the B.Sc. degree from Xi'an Jiaotong University, Xi'an, China, the DEA and Ph.D. degrees from the Institut National des Sciences Appliquées de Lyon (INSA-Lyon), Villeurbanne, France. He is currently a Chair Professor of mechanical engineering and the Director of the Consortium for Sound and Vibration Research (CSVR), Hong Kong Polytechnic University, Hong Kong. He started his academic career at Laval University, Quebec, Canada, in 1992, rising from an Assistant Professor to an Associate/Full Professor, before coming to Hong Kong in 2000, where he was promoted to a Chair Professor in 2005 and was the Head of the Department from 2011 to 2014. He works in the field of sound and vibration, structural health monitoring, and smart structure and control. He is an elected Fellow of the Acoustical Society of America, Acoustical Society of China, IMechE, Hong Kong Institution of Engineers, and Hong Kong Institute of Acoustics. He currently serves as the Deputy Editor-in-Chief and Receiving Editor for the *Journal of Sound and Vibration* and an Associate Editor for the *Journal of Acoustical Society of America*.