Available online at www.sciencedirect.com
ScienceDirect
Neurocomputing 72 (2008) 392–400
NEUROCOMPUTING
www.elsevier.com/locate/neucom

Structural identifiability of generalized constraint neural network models for nonlinear regression ☆

Shuang-Hong Yang[a,*], Bao-Gang Hu[a], Paul-Henry Cournède[b]

[a]NLPR & LIAMA, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China
[b]Laboratory of Applied Mathematics and Systems, École Centrale Paris 92295, France

Received 16 August 2007; received in revised form 30 November 2007; accepted 17 December 2007
Communicated by D. Wang
Available online 31 December 2007

Abstract

Identifiability becomes an essential requirement for learning machines when the models contain physically interpretable parameters. This paper presents two approaches to examining structural identifiability of the *generalized constraint neural network* (GCNN) models by viewing the model from two different perspectives. First, by taking the model as a static deterministic function, a functional framework is established, which can recognize deficient model and at the same time reparameterize it through a pairwise-mode symbolic examination. Second, by viewing the model as the mean function of an isotropic Gaussian conditional distribution, the algebraic approaches [E.A. Catchpole, B.J.T. Morgan, Detecting parameter redundancy, Biometrika 84 (1) (1997) 187–196] are extended to deal with multivariate nonlinear regression models through symbolically checking linear dependence of the *derivative functional vectors*. Examples are presented in which the proposed approaches are applied to GCNN nonlinear regression models that contain coupling physically interpretable parameters.

Keywords: Identifiability; Parameter redundancy; Derivative functional vector; Nonlinear regression; Hybrid neural network

1. Introduction

Parameter estimation (or identification) is an important tool in system modeling. It occurs whenever one wants to model a process using a parameterized model. However, determining identifiability of the model being used should be addressed before any implementation of identifications [26,24,11], because identifiability is closely related to the convergence [17,23,5] of a class of estimates (including the maximum likelihood estimate, MLE). Lack of identifiability gives no guarantee of convergence to the true value of parameters and therefore usually gives rise to confusing results, which is a critical issue especially when some parameters are of practical importance. Besides the ability to detect deficient models in advance, the analysis of identifiability can also bring practical benefits, such as insightful revealing of the relations among inputs, outputs and parameters, which can be very *helpful for model structure designing & selection* [25,24,5] *and numerical estimation* [11]. In particular, the importance of identifiability in machine learning can be recognized in at least threefold:

- *Un-/semi-supervised learning*: Identifiability seems not a big deal in supervised learning, where output–input behavior has dominant importance and lack of identifiability merely means that one obtains an equivalent class of parameter vectors [13]. However, it is of fundamental importance in unsupervised and semi-supervised learning [10,5], where incomplete data or latent variables are usually involved and identifiability is necessary to ensure coherent inference of such latent

☆This work is supported in part by NSFC (#60073007, #60275025).
*Corresponding author at: National Laboratory Pattern Recognition (NLPR) and Sino-French Computer Science Laboratory (LIAMA), Institute of Automation, Chinese Academy of Sciences, P.O. Box 2728, Beijing 100080, China.
E-mail addresses: shyang@nlpr.ia.ac.cn (S.-H. Yang), hubg@nlpr.ia.ac.cn (B.-G. Hu), Paul-Henry.Cournede@mas.ecp.fr (P.-H. Cournède).

0925-2312/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2007.12.013

variables. For instance, in mixture density models, if the parameter cannot be determined uniquely, the mixture cannot be decomposed into its true components. Other examples include Kalman Filter models, Bayesian networks, etc.

- *Physically interpretable* (*sub-*)*models*: Identifiability is also of fundamental importance if we wish to interpret the parameter values discovered by a model [4,26] or if the parameterization of the model is based on physical prior knowledge of the process [12,9]. In both cases, the model parameters are of practical importance, and to identify the true values of such parameters is imperative because nonuniqueness of such parameters not only means nonunique description of the process but also results in severe ill-conditioned identification problems so that estimation gives rise to completely erroneous or misleading results. For example, in hybrid neural network (NN) [20,27], sub-models may contain real-world parameters, whose identification is also a critical task.
- *Reliable neural modeling*: A critical problem with the artificial neural networks (ANNs) lies in that: on the one hand, minor setting changes (e.g., initialization, the number of hidden layers or units, outliers, etc.) could lead to totally different results [15]; on the other hand, when the number of hidden units is large, the training errors may be rather insensitive to those factors [2]. This problem makes the NN technique less reliable in application in comparison with the other existing methods (e.g., methods based on convex optimization such as support vector machines). Recently, a convex NN formulation [2] is proposed to address this problem, but it is rather computational expensive. Theoretically, an obvious alternative to address the unreliability problem of NNs will be to carefully design/select the model structure so as to make sure the parameters can be uniquely determined upon any nontrivial training data set. If the model being used is structurally identifiable, the ill-condition problem might be alleviated [26].

The structural identifiability (*s.i.*) is concerned with the uniqueness of the parameters determined from the input–output data. The term 'structural' means independent of the parameter values [12]. If different parameter values lead to different output throughout the parameter space, the model is said to be structurally global identifiable (*s.g.i.*); if all different parameter values that lead to identical output are isolated from each other, the model is structurally local identifiable (*s.l.i.*), otherwise, it is structurally nonidentifiable (*s.n.i.*).

Structure identifiability is a fundamental prerequisite for implementation of identifications. Indeed, this problem should be addressed, as part of the qualitative experiment design [5] or model selection [4], before any experimental data have been collected because the difficulties associated with identification usually stems from the structure of the model and the method of parameterization rather than inappropriate experiment designing or poor data collection. In other words, if the model is structurally unidentifiable, no matter how carefully we design the experiment or how good the observations are, one will definitely fail to get a reasonable estimation, even when a model selection criterion (e.g., AIC, BIC, etc.) or regularization term is employed to penalize the complexity of the model. Therefore, once a model structure has been chosen (or a set of structures among which one will have to choose), one should test the structure identifiability, as independently of the data as possible, so as to rule out priori unidentifiable models to avoid potential defects.

This paper concerns *s.i* of the generalized constraint neural network (GCNN, [18]) nonlinear regression models with coupling parameters, since some parameters in GCNN (e.g., the parameters in the partially known relationship (PKR) sub-models) are usually of practical interest. Although the problem involving identifiability has been extensively studied and there have already been a host of existing approaches in the literature, none of these are appropriate for this purpose (see Section 3 for details). In this paper, two different approaches are established by viewing the model from two distinct perspectives. First, by viewing the models as static deterministic nonlinear functions, we present a functional framework for this type of models and propose a novel method to test parameter dependence based on this framework. This method also naturally leads to a pairwise-mode reparameterization approach through detecting and eliminating parameter dependent pairs. However, in the current version, it is only workable for single-input-single-output (SISO) models. Trivial as it seems, this method may provide a new perspective to nonidentifiability. Second, by introducing an augmented noise model and treating the GCNN models as mean functions of Gaussian conditional distributions, we enable the problem to be considered in the conventional stochastic framework; and by modifying the definition of derivative matrix (DM) to redefine a derivative functional vector (DFV), we are able to modify the algebraic approach in [6] to test parameter redundancy of the Gaussian conditional distributions in terms of the symbolical linear dependence of DFV. Both of the two approaches have strengths and limitations. While the former provides a natural way for reparameterization, it is relatively complex and only workable for SISO models currently. The latter is simple and efficient, and can also deal with multivariate regression models with isotropic Gaussian noise, however, it tells nothing about reparameterization when redundancy is detected.

The organization of this paper is as follows. Section 2 introduces the GCNN models briefly. In Section 3, We give a concise overview of the literature, and show why none of the existing methods is able to tackle the problem satisfactorily. Section 4 presents the functional framework for static deterministic models and establishes a criterion based on it. Section 5 redefines a DFV and modifies the

approach in [6] to test parameter redundancy for isotropic Gaussian conditional models. And finally, Section 6 summarizes the whole paper.

2. Generalized constraint neural network (GCNN)

In an attempt to enhance the NN technique so that it can evolve from a black-box tool into a semi-analytical one, Hu et al. [18] proposed a GCNN model for nonlinear regression, which impose generalized constraints on ANN models so that available prior knowledge (which may be incomplete, imprecise, nonquantitative, uncertain or even incorrect) can be expressed partially and thus can play an explicit role in modeling highly complex systems. Specifically, the generalized constraints are extended so that they can be any form of PKR knowledge about the process being studied.

The GCNN model basically consists of two sub-models. One is constructed by the standard NN (e.g., RBF, sigmoidal, etc.) technique to approximate the unknown part of the target system. The other is formed from PKR to impose generalized constraints on the whole model explicitly. Fig. 1 shows a schematic configuration of a GCNN model. The upper part of Fig. 1, $g(\boldsymbol{x}, \boldsymbol{\theta}_g)$, represents the PKR sub-model, and the lower part $h(\boldsymbol{x}, \boldsymbol{\theta}_h)$ represents the NN-sub-model. The GCNN model is formulated as $f(\boldsymbol{x}, \boldsymbol{\theta}) = g(\boldsymbol{x}, \boldsymbol{\theta}_g) \circledast h(\boldsymbol{x}, \boldsymbol{\theta}_h)$, where $\circledast$ denotes a specific type of coupling (connection) between the two parts. This configuration provides a flexible representation to various forms of two interacting sub-models. For more details, see [18] and the references therein.

Numerical experiments proved the advantages of the GCNN models [18] in, e.g., improving the generalization capability of NNs, enhancing the interpretability of the resulted models, speeding up the learning phase, extending applications to highly complex system modeling, etc. However, a critical issue was also seen arising from combining two sub-models, that is, the interactive coupling between the two sub-models tends to introducing different parameter values which lead to the same input–output behavior, i.e., the model is usually unidentifiable.

Since the PKR sub-model usually contains physically interpretable parameters (e.g., parameters in a subset of $\boldsymbol{\theta}_g$) whose identification is of fundamental importance to the understanding of the process, the investigation of structural identifiability of the GCNN model, which is exactly the topic of this paper, is imperative.

3. Identifiability: basic concepts and existing methods

In the literature, most studies concerning identifiability can be categorized into two frameworks according to the modeling type, i.e., the stochastic framework for probabilistic models [22,24,6,16] and the noise-free framework for dynamic state-space models [26,3,21,12]. However, since the problem is only solvable for several types of models (e.g., polynomial, rational, etc.) in the noise-free framework and that the GCNN models are mostly static, we will mainly concentrate on the stochastic framework in this section.

The stochastic framework concerns identifiability of the parameters in a certain probability distribution function (PDF) model.

Assume that $\boldsymbol{x}$ is a random vector in $\mathbb{R}^n$, whose distribution function $f(\boldsymbol{x}, \boldsymbol{\theta})$, is controlled by a set of parameters $\boldsymbol{\theta}$. Let the parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p) \in \boldsymbol{S}$, where p is the total number of parameters in the model and $\boldsymbol{S}$ denotes the parameter space. Following [22], we have the following definition.

Definition 1. A parameter point $\boldsymbol{\theta}^{(0)} \in \boldsymbol{S}$ is said to be locally identifiable (*l.i.*) if there exists a neighborhood $\mathscr{U}$ such that any $\boldsymbol{\theta} \in \mathscr{U}$ satisfying $f(\boldsymbol{x}, \boldsymbol{\theta}^{(0)}) = f(\boldsymbol{x}, \boldsymbol{\theta})$ for all $\boldsymbol{x} \in \mathbb{R}^n$ leads to $\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}$; if $\mathscr{U} = \boldsymbol{S}, \boldsymbol{\theta}^{(0)}$ is said to be globally identifiable (*g.i.*); if $\mathscr{U} = \varnothing, \boldsymbol{\theta}^{(0)}$ is said to be nonidentifiable (*n.i.*). If all $\boldsymbol{\theta} \in \boldsymbol{S}$ are *g.i./l.i./n.i.*, $f(\boldsymbol{x}, \boldsymbol{\theta})$ is said to be structurally *g.i./l.i./n.i.*(*s.g.i./s.l.i./s.n.i.*).

Remark 1. Note that *s.i.* presents a necessary not sufficient condition for realizing parameter estimation. An *s.i.* model is in fact only potentially identifiable in the sense that it does not guarantee success of estimation [6,26]. In fact, attempts to identify *s.i.* models may fail because of various reasons, e.g., missing data, contamination noise, imprecise estimators, etc. However, these factors should in no way detract from the necessity of satisfying the prior

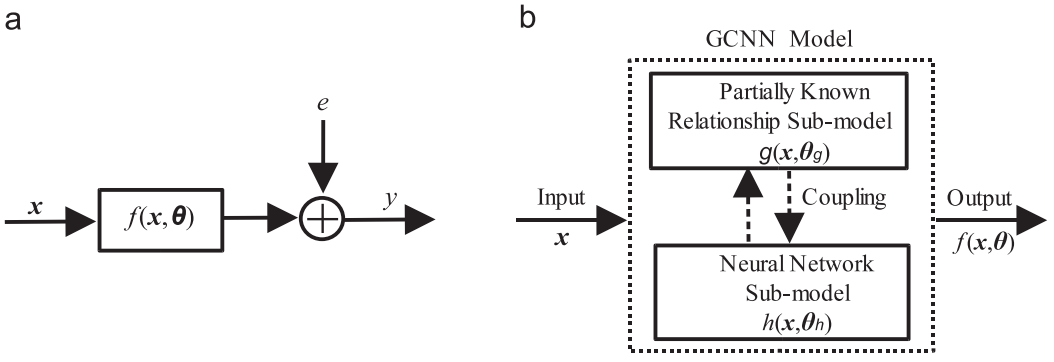


Fig. 1. (a) Schematic diagram of nonlinear regression problems. (b) Architecture of generalized constraint neural network (GCNN) model: composed of partially known relationship (PKR) sub-model and neural network (NN) sub-model, which are interactively coupled.

s.i. requirement, because the most essential reason for nonidentifiability is inherent in the model itself, and bad experiments or poor estimators only makes matter worse. Even with good experiment designs, abundant data and highly precise estimators, it is still impossible to get reasonable estimations for *s.n.i.* models.

An important tool in testing identifiability is the Fisher's information matrix (FIM), i.e.,

$$\mathbf{FIM}: \mathscr{F} = \left(\mathbf{E}\left[\frac{\partial \log f}{\partial \theta_i} \cdot \frac{\partial \log f}{\partial \theta_j}\right]\right)_{p\times p}. \tag{1}$$

Rothenberg [22] studied the identifiability of the general parametric models based on examining the rank of the FIM. They proved that under weak regularity conditions local identifiability is equivalent to nonsingularity of the FIM. They also established criteria to test global identifiability for exponential family PDF models. Hochwald and Nehorai [16] studied the connection between identifiability and regularity of the FIM for Gaussian PDF model and established a tool to check regularity based on holomorphic functions.

Structural identifiability is an intrinsic property of the model. The most obvious cause of *s.n.i.* is over-parameterization, or parameter redundancy. If a model is parameter redundant, the likelihood surface will be maximized along a completely flat ridge or plane for any data set [6]. Therefore, parameter redundant models are certainly *s.n.i.*

Definition 2 (*Parameter redundancy*). The model $f(\boldsymbol{x}, \boldsymbol{\theta})$ is said to be parameter redundant if it can be expressed in terms of a smaller parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_q)$, where $q < p$.

Catchpole and Morgan [6] studied the connections between parameter redundancy and *s.i.* They established necessary and sufficient conditions for parameter redundancy of exponential family PDF models. For a general class of nonlinear models, parameter redundancy is equivalent to the singularity of the FIM and can be tested in terms of rank deficiency of the DM (i.e., Jacobian of the mean vector), which is defined as Eq. (2). They [7] also established an approach to determine which combination of parameters may be estimated when a model is parameter redundant, which requires the calculation of the DM and its null space:

$$\mathbf{DM}: \mathscr{D} = \left(\frac{\partial \mu_i}{\partial \theta_j}\right)_{p\times n}, \tag{2}$$

where $(x_1, x_2, \ldots, x_n)$ is a data vector from the exponential family PDF $f(\boldsymbol{x}, \boldsymbol{\theta})$, $\mu_i = \mathbf{E}[x_i]$ is the mean of x_i.

There also exists computer software for examining identifiability, which is based on either numerical or symbolic computation. Numerical packages, which calculate the FIM at the maximum likelihood parameter estimate and numerically estimate the eigenvector (e.g., by singular value decomposition) to judge singularity by the existence of zero eigenvalues. In contrast, symbolic algebra computer packages [8,9], usually based on the software Maple, test parameter identifiability based on the symbolic algebraic computation, thus are not vulnerable to numerical errors and are irrespective of the extent of the data set.

However, these methods are not appropriate for applications in GCNN models: (i) the GCNN model is static (no time-variation exists in general), thus methods in the noise-free framework for dynamic models cannot be used (e.g., we cannot apply Laplacian transformation [19] to a static model); (ii) as for the GCNN model, we care more about the identifiability of models (i.e., *s.i*) rather than of parameter points, and want to rule out *s.n.i.* models before experiment design and data collection. However, most of the criteria in the stochastic framework (e.g., the methods based on FIM) are based on the data set and are originally established to test identifiability of certain parameter points; (iii) although the algebraic approach [6] does not have the above limitations, it cannot directly be applied either, because it checks parameter redundancy by the row rank deficiency of the DM, which has only one column for the GCNN models so that the rank equals to either one or zero.

4. A functional framework for static deterministic models

In this section, we present a functional framework for static deterministic functional models. In this framework, it is assumed that the model is static and noise free. In other words, the model $y = f(\boldsymbol{x}, \boldsymbol{\theta})$ is simply a parameterized deterministic nonlinear function. The definition of (structural) identifiability is the same as that in Section 2 with an exception that $f(\boldsymbol{x}, \boldsymbol{\theta})$ is not a PDF model.

We concern another cause of *s.n.i*, which we term as *parameter dependence* in the sense that there exists a lower ordered subspace (a manifold) of $\boldsymbol{S}$ such that every parameter point in this subspace is equivalent to each other, i.e., $f(\boldsymbol{x}, \boldsymbol{\theta})$ is invariant of $\boldsymbol{\theta}$ when the (subset of) parameters change their values along the manifold. We will link this new concept of parameter dependence to the commonly used ones in literature, such as parameter redundancy and *s.i.* In addition, we will show that several interesting results and useful tools can be derived from this concept, which may not have been achieved by other existing ones.

Denoting θ_i the ith component of the parameter vector $\boldsymbol{\theta}$ and $S(\theta_i)$ its value space, we first introduce the new concept of parameter dependence as follows.

Definition 3 (*Observational equivalence*). Two model $f(\boldsymbol{x}, \boldsymbol{\theta})$ and $g(\boldsymbol{x}, \boldsymbol{\beta})$ are said to be observationally equivalent if for any parameter point $\boldsymbol{\theta}^{(0)}$, there exists a $\beta^{(0)}$ such that $f(\boldsymbol{x}, \boldsymbol{\theta}) = g(\boldsymbol{x}, \boldsymbol{\beta})$ holds for all $\boldsymbol{x} \in \mathbb{R}^n$.

Definition 4 (*Dependent parameter pairs*). For $y = f(\boldsymbol{x}, \boldsymbol{\theta})$, if for any given $\theta_1^{(1)}$, $\theta_1^{(2)} \in S(\theta_1)$ and $\theta_2^{(1)} \in S(\theta_2)$, there always exists $\theta_2^{(2)} \in S(\theta_2)$ such that $f(\boldsymbol{x}, \boldsymbol{\theta}|\theta_1 = \theta_1^{(1)}, \ \theta_2 = \theta_2^{(1)})$ and $f(\boldsymbol{x}, \boldsymbol{\theta}|\theta_1 = \theta_1^{(2)}, \ \theta_2 = \theta_2^{(2)})$ are observationally

equivalent, then θ_1 is said to be dependent on θ_2. The two parameters θ_1 and θ_2 are defined as dependent parameter pairs, if one of them is dependent on the other.

Definition 5 (*Parameter dependence*). A model $y=f(\boldsymbol{x},\boldsymbol{\theta})$ is said to be with dependent parameters if it contains at least one dependent parameter pair.

We now investigate the connections of parameter dependence with *s.i* and parameter redundancy. The relationships can be described by the following two theorems.

Theorem 1. *A model with dependent parameters is s.n.i.*

Proof. Suppose θ_1 and θ_2 are the dependent parameter pairs in $f(\boldsymbol{x},\boldsymbol{\theta})$, denote the parameter vector composed of the remaining parameters as $\boldsymbol{\theta}_R$. Then according to Definition 4, if $(\theta_1^{(1)},\theta_2^{(1)},\boldsymbol{\theta}_R^{(0)})$ is an estimation for $\boldsymbol{\theta}$, so is $(\theta_1^{(2)},\theta_2^{(2)},\boldsymbol{\theta}_R^{(0)})$. Because $\theta_1^{(1)}$, $\theta_2^{(1)}$, $\boldsymbol{\theta}_R^{(0)}$ and $\theta_1^{(2)}$ are arbitrary, $f(\boldsymbol{x},\boldsymbol{\theta})$ is *s.n.i.* □

Theorem 2. *A model with dependent parameters is parameter redundant. Particularly, if θ_1 and θ_2 are two parameters of $f(\boldsymbol{x},\boldsymbol{\theta})$ and θ_1 is dependent on θ_2, then there exists $\theta_1^{(1)}\in S(\theta_1)$ such that $f(\boldsymbol{x},\boldsymbol{\theta}|\theta_1=\theta_1^{(1)})$ is observationally equivalent to $f(\boldsymbol{x},\boldsymbol{\theta})$.*

Proof. Since θ_1 is dependent on θ_2, according to Definition 4, for any given $\theta_1^{(1)}\in S(\theta_1)$, $\theta_2^{(1)}\in S(\theta_2)$ and $\theta_1^{(2)}\in S(\theta_1)$, there always exists a $\theta_2^{(2)}\in S(\theta_2)$ such that $f(x,\boldsymbol{\theta}|\theta_1=\theta_1^{(1)},\ \theta_2=\theta_2^{(1)})$ and $f(x;\boldsymbol{\theta}|\theta_1=\theta_1^{(2)},\ \theta_2=\theta_2^{(2)})$ are observationally equivalent. Informally, let us denote this relationship as a function $g(\cdot)$, i.e., $\theta_2^{(2)}=g(\theta_1^{(1)},\theta_1^{(2)},\theta_2^{(1)})$. Suppose $\boldsymbol{\theta}^{(0)}$ is an arbitrary parameter point with $\theta_1=\theta_1^{(0)}$, $\theta_2=\theta_2^{(0)}$ and $\boldsymbol{\theta}_R=\boldsymbol{\theta}_R^{(0)}$, then obviously $f(x,\boldsymbol{\theta}|\theta_1=\theta_1^{(0)},\ \theta_2=\theta_2^{(0)},\boldsymbol{\theta}_R=\boldsymbol{\theta}_R^{(0)})$ and $f(x;\boldsymbol{\theta}|\theta_1=\theta_1^{(1)},\ \theta_2=\theta_2^{(g)},\boldsymbol{\theta}_R=\boldsymbol{\theta}_R^{(0)})$ are observationally equivalent, where $\theta_2^{(g)}=g(\theta_1^{(0)},\ \theta_1^{(1)},\ \theta_2^{(0)})$. Since, in $g(\theta_1^{(0)},\ \theta_1^{(1)},\ \theta_2^{(0)})$, $\theta_1^{(0)}$ and $\theta_2^{(0)}$ are arbitrary, $\theta_2^{(g)}$ is an arbitrary value, too. Considering $\boldsymbol{\theta}^{(0)}$ is arbitrary, therefore, $f(\boldsymbol{x},\boldsymbol{\theta}|\theta_1=\theta_1^{(1)})$ is observationally equivalent to $f(\boldsymbol{x},\boldsymbol{\theta})$. □

From Theorem 2, we can see that parameter dependence is a sufficient condition for parameter redundancy. In addition, this theorem also provides a feasible way for reparameterization, i.e., how to eliminate redundant parameters if the model contains dependent parameters. This is obviously an advantage of examining parameter dependence rather than parameter redundancy, since testing parameter redundancy tells nothing about how to reparameterize the model, and thus extra procedures, which can be very complicated and time-consuming, usually need to be carried out [6] when redundancy is detected. In contrast, examination of parameter dependence allows us to detect potential deficiency inherent in the model and at the same time provides a natural way to reparameterize it when dependence is detected.

Theorem 1 can also be derived from Theorem 2, since it has been proved [6] that parameter redundancy is a sufficient condition of *s. n. i*. The clear relationships among *dependent parameter pairs*, *parameter dependence*, *parameter redundancy* and *s. n. i* can be seen from Fig. 2.

We now investigate how to detect parameter redundancy and dependence based on the functional framework. In particular, we present a theorem offering a necessary and sufficient condition of parameter redundancy. A corollary following this theorem provides a sufficient condition of parameter dependence.

Theorem 3 (*Examination of parameter redundancy for SISO models*). *Suppose an SISO nonlinear function, denoting by $y=f(x,\boldsymbol{\theta})$ is differentiable with respect to both x and $\boldsymbol{\theta}$. Define a determinant H as Eq.* (3), *we have*:

The model $y=f(x,\boldsymbol{\theta})$ is not parameters redundant if and only if there exists an open neighborhood in which $H\neq 0$.

$$H=\begin{vmatrix} \dfrac{\partial f}{\partial\theta_1} & \dfrac{\partial f}{\partial\theta_2} & \cdots & \dfrac{\partial f}{\partial\theta_p} \\ \dfrac{\partial f^{(1)}}{\partial\theta_1} & \dfrac{\partial f^{(1)}}{\partial\theta_2} & \cdots & \dfrac{\partial f^{(1)}}{\partial\theta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \dfrac{\partial f^{(p-1)}}{\partial\theta_1} & \dfrac{\partial f^{(p-1)}}{\partial\theta_2} & \cdots & \dfrac{\partial f^{(p-1)}}{\partial\theta_p} \end{vmatrix}. \tag{3}$$

Proof. (i) For sufficiency, if the model $f(x,\boldsymbol{\theta})$ is parameter redundant, then it can be equivalently expressed by a subset of the parameters. Following the chain rule for derivatives, we have $H=0$.

(ii) For necessary, consider the following equations derived from $y=f(x,\boldsymbol{\theta})$:

$$\begin{cases} y=f(x,\boldsymbol{\theta}), \\ \dfrac{\mathrm{d}y}{\mathrm{d}x}=\dfrac{\mathrm{d}}{\mathrm{d}x}f(x,\boldsymbol{\theta}), \\ \vdots \\ \dfrac{\mathrm{d}^{(p-1)}y}{\mathrm{d}x^{(p-1)}}=\dfrac{\mathrm{d}^{(p-1)}}{\mathrm{d}x^{(p-1)}}f(x,\boldsymbol{\theta}). \end{cases} \tag{4}$$

Take derivatives (Jacobians) with respect to $\boldsymbol{\theta}$ for each question, we have the following first-order differential equations:

$$\mathrm{d}\gamma=M\times\mathrm{d}\boldsymbol{\theta}, \tag{5}$$

where $\boldsymbol{\gamma}=(y,y',\ldots,y^{(p-1)})^{\mathrm{T}}$, $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_p)^{\mathrm{T}}$.

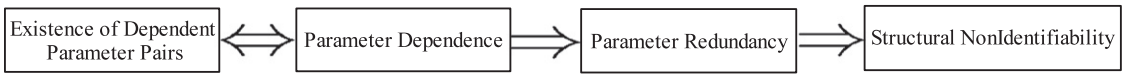


Fig. 2. The relationships among dependent parameter pairs, parameter dependence, parameter redundancy and structural nonidentifiability.

Since there does not exist any neighborhood in which $H \neq 0$, then we can find a series of neighborhoods which cover the whole value space $\boldsymbol{S}$ and in every such neighborhood $H = 0$ holds nontrivially. Therefore, the solution of equation Eq. (5) is either not existing or not unique. Considering that the series of neighborhoods cover $\boldsymbol{S}, \boldsymbol{\theta}$ is *s.n.i.* For each such neighborhood, since $H = 0$, the columns of M are not linearly independent, i.e., (at least) one of the columns can be expressed by the linear combination of others. Hence, the general solution of equation (5) can be expressed by (at most) $p - 1$ parameters, the parameters in $f(x, \boldsymbol{\theta})$ are redundant. □

Lemma 1. *The parameters a and b in the function* $y = f(x, a, b, \mathbf{C})$ (*where* $\mathbf{C}$ *is the set of the rest parameters*) *are independent of one another if and only if there exists a neighborhood of* (x, a, b) *in which the determinant defined in* (9) *does not equal to zero*:

$$H = \begin{vmatrix} \dfrac{\partial f}{\partial a} & \dfrac{\partial f}{\partial b} \\ \dfrac{\partial f'}{\partial a} & \dfrac{\partial f'}{\partial b} \end{vmatrix}. \tag{6}$$

Theorem 3 and Lemma 1 provide a feasible criterion for checking parameter redundancy and the existence of dependent parameter pairs. According to Definition 5, Lemma 1 also provides an approach to detecting parameter dependence, i.e., the existence of dependent parameter pairs is a sufficient condition for parameter dependence. In addition, together with Theorem 2, they enable us to detect deficient model structure and at the same time to reparameterize the model by pairwisely examining and eliminating dependent parameter pairs, that is, to use Lemma 1 to pairwisely detect dependent pairs, and then eliminate them by Theorem 2. Unfortunately, at this point, we has only achieved a partial solution for SISO models.

Example 1. An obvious example of parameter dependence is $y = f(x, u(a, b), \mathbf{C})$, where the parameters a and b present in the model only through an arbitrary function $u(a, b)$, e.g., $y = ae^b x$ and $y = (a + b)x$. Following Lemma 1, this can be easily verified since

$$H = \begin{vmatrix} \dfrac{\partial f}{\partial a} & \dfrac{\partial f}{\partial b} \\ \dfrac{\partial f'}{\partial a} & \dfrac{\partial f'}{\partial b} \end{vmatrix} = \begin{vmatrix} \dfrac{\partial f}{\partial u} \cdot \dfrac{\partial u}{\partial a} & \dfrac{\partial f}{\partial u} \cdot \dfrac{\partial u}{\partial b} \\ \dfrac{\partial f'}{\partial u} \cdot \dfrac{\partial u}{\partial a} & \dfrac{\partial f'}{\partial u} \cdot \dfrac{\partial u}{\partial b} \end{vmatrix} \equiv 0.$$

Example 2. In [18], a hybrid radius basis function (RBF) NN (Fig. 3(a)), which can be formulated as Eq. (7), is applied to a benchmark nonlinear regression problem [1]. In this GCNN model, the sub-model $g(x, \alpha) = e^{-\alpha x}$ is partially known and associated with the RBF NN sub-model $h_1(x, \boldsymbol{\theta}_h)$ in a multiplication configuration, where α is a physically interpretable parameter (i.e., the damping coefficient), which is of interest to estimate since its value reflects the level of the energy dissipation in the target system. All parameters, including the physically based parameter α, were learned simultaneously in this example. Although higher accuracy was obtained by GCNN in comparison with other models due to the utility of prior knowledge, it was also observed, through numerical simulations, that it is impossible to obtain a reasonable estimation for the practically important parameter α. In this example, we will rigorously prove that this model is actually structurally nonidentifiable. Note that the structural identifiability analysis enables us to detect and rule out deficient model structure before any implementation of simulations:

$$y = g(x, \alpha) \times h_1(x, \boldsymbol{\theta}_h) = e^{-\alpha x} \sum_i w_i \exp\left\{-\frac{(x - c_i)^2}{s_i^2}\right\}. \tag{7}$$

For simplicity, let us just consider one term which can be rewritten as below:

$$f(x) = e^{-\alpha x} w e^{-(x-c)^2/s^2} = \text{sign}(w) \cdot e^{\ln|w| - \alpha x - (x-c)^2/s^2}. \tag{8}$$

Note that the exponential term is a quadratic polynomial which can have three independent parameters at most while in Eq. (8) there are four, so intuitively they are redundant. We shall prove this by using Theorem 3.

Proposition 1. *The model in Eq.* (8) *is parameter redundant*.

Proof. We use the Maple package (version 9.5) to prove this proposition. The complete code is given in Table 1.

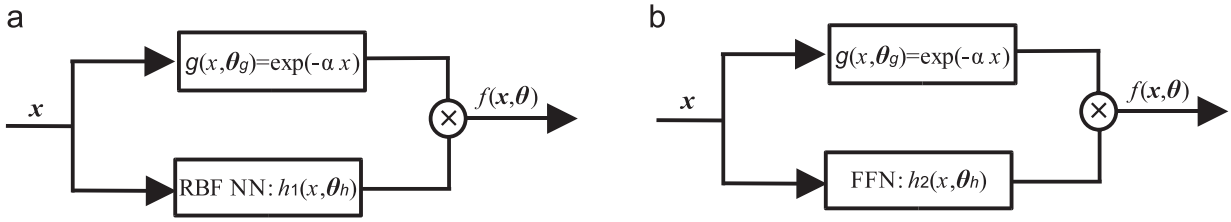


Fig. 3. Improper choice of activation functions for neural networks can result in severe ill-conditioned problems for learning: (a) when radius basis function (RBF) neural network is employed as the sub-model of generalized constraint neural network (GCNN) (Example 2), the hybrid model is structurally nonidentifiable; (b) when feed-forward network (FFN) sub-model is employed (Example 3), the defects are avoided.

Table 1
Maple commands for Propositions 1 and 2

```
Proof for Proposition 1
with(VectorCalculus):
f := w * exp(−1 * a * x − (x − c)^2/(s^2));
Jacobian([f,diff(f,x$1),diff(f,x$2),diff(f,x$3)],[a,w,s,c],'determinant');

Proof for Proposition 2
with(VectorCalculus) :
f:=w*exp(-1*a*x)/(1+exp(-1*r*x-b));
Jacobian([f,diff(f,x$1),diff(f,x$2),diff(f,x$3)],[a,w,r,b], 'determinant');
```

We have

$$H_f = \begin{vmatrix} \frac{\partial f}{\partial \alpha} & \frac{\partial f}{\partial w} & \frac{\partial f}{\partial s} & \frac{\partial f}{\partial c} \\ \frac{\partial f'}{\partial \alpha} & \frac{\partial f'}{\partial w} & \frac{\partial f'}{\partial s} & \frac{\partial f'}{\partial c} \\ \frac{\partial f^{(2)}}{\partial \alpha} & \frac{\partial f^{(2)}}{\partial w} & \frac{\partial f^{(2)}}{\partial s} & \frac{\partial f^{(2)}}{\partial c} \\ \frac{\partial f^{(3)}}{\partial \alpha} & \frac{\partial f^{(3)}}{\partial w} & \frac{\partial f^{(3)}}{\partial s} & \frac{\partial f^{(3)}}{\partial c} \end{vmatrix} = 0 \tag{9}$$

therefore, according to the theorem, the parameters α, w, s, and c in model Eq. (7) are redundant. □

Example 3. Consider the example above again, but instead of using RBF-NN, we now use sigmoid feed-forward perceptrons, $h_2(x, \theta_h)$, as the NN sub-model (see Fig. 3(b)). Then, the corresponding GCNN model is given by

$$y = g(x, \alpha) \times h_2(x, \theta_h) = e^{-\alpha x} \sum_i w_i \frac{1}{1 + e^{-\gamma_i x - b_i}}, \tag{10}$$

where γ_i is the input layer weights and b_i is the bias parameter. We will show that the nonidentifiability issue will be avoided if we use a different type of NN model. Similarly, let us consider the model below for simplicity:

$$f(x) = \frac{we^{-\alpha x}}{1 + e^{-\gamma x - b}}. \tag{11}$$

Proposition 2. *The model in Eq.* (11) *is not parameter redundant.*

Proof. The Maple code is also given in Table 1. The result is

$$H_f = \frac{w^3\gamma^4 \exp(-4\alpha x - 2\gamma x - 2b)}{(1 + \exp(-\gamma x - b))^8}.$$

$H_f = 0$ occurs if and only if $w\gamma = 0$. Therefore, there does not exists any neighborhood for $H_f = 0$. Hence the model in Eq. (11) is not parameter redundant. □

The two examples above illustrate that inappropriate model selection, e.g., improper choice of activation functions for NNs, can result in severe ill-conditioned problems for learning (i.e., the parameters may never be estimated reasonably no matter how much data is given). On the other hand, it indicates that the illposedness can be possibly eliminated by using another type of activation functions when deficiency is detected in the model.

5. Parameter redundancy in isotropic Gaussian conditional distribution models

The GCNN model estimates the parameters by minimizing the mean square error (MSE) on the training set, which is equivalent to MLE of a Gaussian-noise contaminated model [4,13]. This interpretation allows us to treat the GCNN model as the mean function of a Gaussian conditional distribution model in Eq. (12) and to deal with the problem within the stochastic framework:

$$P(y \mid \boldsymbol{x}) = \mathcal{N}(y|f(\boldsymbol{x}, \boldsymbol{\theta}), \sigma^2). \tag{12}$$

However, the existing criteria in this framework cannot be used directly to test *s.i* of this model. For example, the algebraic approach to checking parameter redundancy proposed by [6] is based on the DM. As for the conditional distribution model in Eq. (12), checking parameter redundancy by the row rank deficiency of the DM is not appropriate since it has only one column for this type of model.

According to [22], Eq. (12) is a type of one-dimensional reduced form model in the sense that the distribution depends on the parameters only through a reduced form parameter which completely characterizes the distribution. In exponential family distributions, this reduced form parameter is the sufficient statistic. In particular, since the conditional distribution model in Eq. (12) is isotropic, it depends on the parameters $\boldsymbol{\theta}$ solely through the conditional mean $f(\boldsymbol{x}, \boldsymbol{\theta})$. Therefore, the identification of $\boldsymbol{\theta}$ depends solely on the properties of the mapping $f(\boldsymbol{x}, \boldsymbol{\theta})$. Based on this notion, we define a DFV to tackle the parameter redundancy of the conditional distribution model. The DFV is nothing but a modification of the DM, i.e.,

$$\textbf{DFV}: \boldsymbol{d} = \nabla_{\boldsymbol{\theta}} f = \left(\frac{\partial f}{\partial \theta_i}\right)_{p \times 1}. \tag{13}$$

Accordingly, the symbolic deficiency of the DM is modified to symbolic linear dependence, which is defined as follows.

Definition 6 (*Symbolically linear dependence*). The DVF is said to be symbolically linear dependent if there exists a vector function $\lambda(\boldsymbol{\theta})$, nonzero for all $\boldsymbol{\theta} \in \boldsymbol{S}$, such that $\lambda(\boldsymbol{\theta})^{\mathrm{T}} \boldsymbol{d}(\boldsymbol{\theta}) = 0$.

To check parameter redundancy of the conditional distribution model, we have the following theorems.

Theorem 4. *A model is parameter redundant if and only if its DFV is symbolically linear dependent.*

Proof (*Catchpole and Morgan [6]*). For necessity, suppose $f(\boldsymbol{x}, \boldsymbol{\theta})$ is parameter redundant, i.e., it can be rewritten as $f(\boldsymbol{x}, \boldsymbol{\beta})$, where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_q)$. Since $q < p$, following the chain rule for derivatives, $\boldsymbol{d}(\boldsymbol{\theta})$ is obviously linear dependent.

For sufficiency, suppose there exists a vector function $\boldsymbol{\lambda}(\boldsymbol{\theta})$, nonzero for all $\boldsymbol{\theta} \in \boldsymbol{S}$, such that $\boldsymbol{\lambda}(\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{d}(\boldsymbol{\theta}) = 0$. This implies a linear first order Lagrange partial differential equation with auxiliary equations [14]:

$$\frac{\mathrm{d}\theta_1}{\lambda_1(\boldsymbol{\theta})} = \frac{\mathrm{d}\theta_2}{\lambda_2(\boldsymbol{\theta})} = \cdots = \frac{d\theta_p}{\lambda_p(\boldsymbol{\theta})}. \tag{14}$$

Eq. (14) in general has $p-1$ solutions. Since the distribution is isotropic, which depends on $\boldsymbol{\theta}$ solely through $f(\boldsymbol{x}, \boldsymbol{\theta})$ [22], therefore, the parameters in Eq. (12) are redundant. □

We now give some examples to illustrate the application of Theorem 4 in testing parameter redundancy in GCNN models, and more generally, in regression models.

Example 4. The first example is linear basis function models for regression [4], i.e.,

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{i=1}^{p} \theta_i \phi_i(\boldsymbol{x}), \tag{15}$$

where $\phi_i(\cdot)$ are known as basis functions or feature maps. For this type of models, since the DFV is the basis function vector $(\phi_1(\boldsymbol{x}), \ldots, \phi_p(\boldsymbol{x}))$, the model is not parameter redundant if and only if the basis functions are symbolically linear independent, which is consistent with our intuition. In fact, according to Theorem 4 here and Theorem 5 in [22], it is easy to verify that the model in Eq. (15) is structurally identifiable if and only if its basis functions are independent of each other, which coincides exactly with a result given in [17].

Example 5. This example comes from [23], where

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = \mathrm{e}^{-\theta_2\theta_3 x_1} + \frac{\theta_1}{\theta_2}(1 - \mathrm{e}^{-\theta_2\theta_3 x_1})x_2. \tag{16}$$

It can be easily verified that for any constant $c \neq 0$, $(\theta_1, \theta_2, \theta_3)$ and $(c\theta_1, c\theta_2, \theta_3/c)$ are observationally equivalent.

The DFV for Eq. (16) is

$$\boldsymbol{d} = \begin{pmatrix} \dfrac{1}{\theta_2}(1 - \mathrm{e}^{-\theta_2\theta_3 x_1})x_2 \\ \left(\dfrac{\theta_1}{\theta_2}\theta_3 x_1 x_2 - \theta_3 x_1 - \dfrac{\theta_1}{\theta_2^2}x_2\right)\mathrm{e}^{-\theta_2\theta_3 x_1} - \dfrac{\theta_1}{\theta_2^2}x_2 \\ (\theta_1 x_1 x_2 - \theta_2 x_1)\mathrm{e}^{-\theta_2\theta_3 x_1} \end{pmatrix}. \tag{17}$$

Solving $\boldsymbol{\lambda}(\boldsymbol{\theta})$ out of $\lambda(\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{d}(\boldsymbol{\theta}) = 0$, we have

$$\lambda_1 : \lambda_2 : \lambda_3 = \theta_1 : \theta_2 : \theta_3, \tag{18}$$

that is, any vector function $\boldsymbol{\lambda}(\boldsymbol{\theta})$ satisfying Eq. (18) will lead to $\boldsymbol{\lambda}(\boldsymbol{\theta})^{\mathrm{T}}\,\mathbf{d}(\boldsymbol{\theta}) = 0$, thus, the DFV is symbolically linear dependent, the model in Eq. (16) is parameter redundant.

Example 6. Consider Examples 3 and 4 again. The DFV for Eq. (8) is

$$\boldsymbol{d} = \left(\frac{\partial f}{\partial \alpha}, \frac{\partial f}{\partial w}, \frac{\partial f}{\partial c}, \frac{\partial f}{\partial s}\right)^{\mathrm{T}} = \begin{pmatrix} -wxE \\ E \\ 2(x-c)wE/s^2 \\ 2(x-c)^2 wE/s^3 \end{pmatrix}, \tag{19}$$

where $E = \mathrm{e}^{-(x-c)^2/s^2 - \alpha x}$. Solving $\boldsymbol{\lambda}(\boldsymbol{\theta})$ out of $\boldsymbol{\lambda}(\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{d}(\boldsymbol{\theta}) = 0$, we have

$$\lambda_1 : \lambda_2 : \lambda_3 : \lambda_4 = 1 : cw : s^2/2 : 0. \tag{20}$$

The DFV is symbolically linear dependent, hence the model in Eq. (8) is parameter redundant. In contrast, the DFV for Eq. (11) is

$$\boldsymbol{d} = \left(\frac{\partial f}{\partial \alpha}, \frac{\partial f}{\partial w}, \frac{\partial f}{\partial \gamma}, \frac{\partial f}{\partial b}\right)^{\mathrm{T}} \propto \begin{pmatrix} x\mathrm{e}^{\alpha x}/w \\ -(\mathrm{e}^{\alpha x} + \mathrm{e}^{-\gamma x - b})/w^2 \\ -x\mathrm{e}^{-\gamma x - b}/w \\ -\mathrm{e}^{-\gamma x - b}/w \end{pmatrix}. \tag{21}$$

Solving $\boldsymbol{\lambda}(\boldsymbol{\theta})$ out of $\boldsymbol{\lambda}(\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{d}(\boldsymbol{\theta}) = 0$, we have

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0. \tag{22}$$

The DFV is symbolically linear independent, therefore the model in Eq. (11) is not parameter redundant.

6. Conclusion

Testing nonlinear models for identifiability should be a prerequisite before attempting to decide which model structure is most appropriate and what is the best value for the parameters on the basis of experimental data. Otherwise, no conclusion can be drawn on the value of some physically meaningful parameters.

This paper is a preliminary study on the structure identifiability of the GCNN models. In particular, we have presented a functional framework for static deterministic nonlinear models and established a criteria for testing nonidentifiability based on this framework. In addition, based on the stochastic framework, we have extended the DM-based approach to test parameter redundancy in isotropic Gaussian conditional models in terms of the symbolically linear dependence of the newly defined DFV.

Although this paper is focused on identifiability of the GCNN models, the resulted methods are applicable to more general problems, for example, the functional framework is appropriate for any static deterministic models, and the DFV-based methods are able to examine parameter redundancy for any nonlinear regression models with isotropic Gaussian noises.

The approaches presented in this paper can be easily implemented by symbolic computation packages such as Maple. However, identifiability of nonlinear models is still difficult to test since, whatever the method being used, it is required to solve a system of nonlinear algebraic equations whose complexity increases very fast with the number of

parameters, the dimension of the input vector, the degree of nonlinearity, etc.

Finally, we outline two directions below for future work:

- Although the functional analysis approach enables us to detect parameter dependance, it is only applicable to SISO models in current version. It would be more reliable if we could extend it to the more generic MIMO models. In addition, though this approach naturally offers a way for reparameterization by detecting and eliminating dependent parameter pairs, this pairwise procedure is relatively time-consuming (of complexity $O(p^2)$). It would be highly desirable to seek for reparameterization methods which operate in linear time.
- Another feasible way to make a structurally nonidentifiable model become identifiable is to augment it with more prior knowledge, for example, to associate the model with some constraints from domain knowledge. This is important especially for applications where the *s.n.i.* models being used are based on our incomplete prior knowledge and thus it is impractical to reparameterize the model by simply eliminating redundant parameters. Therefore, it is worthwhile considering problems such as how many, and what types of, constraints are required to produce a unique estimation.

References

[1] V. Basios, A.Y. Bonushkina, V.V. Ivanov, A method for approximating one-dimensional functions, Comput. Math. Appl. 7/8 (1997) 687–693.

[2] Y. Bengio, N.L. Roux, P. Vincent, O. Delalleau, P. Marcotte, Convex neural network, in: 20th International Conference on Advanced in Neural Information Systems (NIPS2006), 2006.

[3] F. Berthier, J.P. Diard, L. Pronzato, E. Walter, Identifiability and distiguishability concepts in electrochemistry, Automatica 32 (7) (1996) 973–984.

[4] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, Berlin, 2006.

[5] G. Casella, R.L. Berger, Statistical Inference (2E), Duxbury, 2002.

[6] E.A. Catchpole, B.J.T. Morgan, Detecting parameter redundancy, Biometrika 84 (1) (1997) 187–196.

[7] E.A. Catchpole, B.J.T. Morgan, S.N. Freeman, Estimation in parameter-redundant models, Biometrika 85 (1) (1998) 462–468.

[8] E.A. Catchpole, B.J.T. Morgan, A. Viallefont, Solving problems in parameter redundancy using computer algebra, Journal of Applied Statistics 29 (2002) 625–636.

[9] J.C. Chapman, K.R. Godfrey, M.J. Chappell, N.D. Evans, Structural identifiability for a class of nonlinear compartmental systems using linea/nonlinear splitting and symbolic computation, Math. Biosci. 183 (2003) 1–14.

[10] A. Corduneanu, T. Jaakkola, Stable mixing of complete and incomplete information, Technical Report AIM-2001-030, MIT AI Memo, 2001.

[11] L. Denis-Vidal, G. Joly-Blanchard, C. Noiret, System identifiability (symbolic computation) and parameter estimation (numerical computation), Numer. Algorithms 34 (2003) 283–292.

[12] H.G.M. Dötsch, P.M.J. van den Hof, Test for local identifiability of high-order nonlinearly parameterized state-space models, Automatica 32 (6) (1996) 875–883.

[13] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley, New York, 2001.

[14] C.H. Edwards, D.E. Penney, Differential Equations and Boundary Value Problems, Computing and Modeling, third ed., Prentice-Hall, Englewood Cliffs, NJ, 2004.

[15] S. Haykin, Neural Networks: A Comprehensive Foundation, second ed., Prentice-Hall/Pearson, Englewood Cliffs, NJ, 1998.

[16] B. Hochwald, A. Nehorai, On identifiability and information-regularity in parametrized normal distributions, Circuit Syst. Signal Process. 16 (1) (1997) 83–89.

[17] K. Hsu, C. Novara, T. Vincent, M. Milanese, K. Polla, Parametric and nonparametric curve fitting, Automatica 42 (11) (2006) 1869–1873.

[18] B.G. Hu, H.B. Qu, Y. Wang, S.H. Yang, Generalized constraint neural network model: associating partially known relationships for nonlinear regressions, NLPR Technical Report, 2007.

[19] L. Ljung, System Identification: Theory for the User, second ed., Prentice-Hall, Englewood Cliffs, NJ, 1999.

[20] K. McGarry, S. Wertmer, J. MacIntyre, Hybrid neural systems: from simple coupling to fully integrated neural networks, Neural Comput. Surveys 2 (1999) 62–93.

[21] R.L.M. Peeters, B.L. Hanzon, Identifiability of homogeneous systems using the state isomorphism approach, Automatica 41 (2005) 513–529.

[22] T.J. Rothenberg, Identification in parametric models, Econometrica 39 (3) (1971) 577–591.

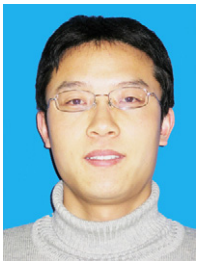
[23] G.A.F. Seber, C.J. Wild, Nonlinear Regression, Wiley, New York, 2003.

[24] E. Tse, J. Anton, On the identifiability of parameters, IEEE Trans. Automat. Control 17 (5) (1972) 637–646.

[25] E. Walter, L. Pronzato, On the identifiabilities and distigushability of nonlinear parametric models, Math. Comput. Simul. 42 (1996) 125–134.

[26] E. Walter, L. Pronzato, Identification of Parametric Models from Experimental Data, Springer, Berlin, 1997.

[27] S. Wermter, R. Sun, Hybrid Neural Systems, Springer, Berlin, 2000.



Shuang-Hong Yang received his B.Sc. degrees from Wuhan University, Wuhan, China, in 2005. Currently he is an M.S. candidate at the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include supervised and unsupervised learning, kernel methods, Bayesian inferences, dimensionality reduction and mathematical modeling.

Bao-Gang Hu received his M.Sc. degree from the University of Science and Technology, Beijing, China in 1983, and his Ph.D. degree from McMaster University, Canada in 1993, all in Mechanical Engineering. From 1994 to 1997, Dr. Hu was a Research Engineer and Senior Research Engineer at C-CORE, Memorial University of Newfoundland, Canada. Currently, he is a Professor with the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Science, Beijing, China. From 2000 to 2005, he was the Chinese Director of the Sino-French Computer Science Laboratory (LIAMA). His main research interests include intelligent systems, pattern recognition, plant growth modeling. He is a Senior Member of IEEE.

Paul-Henry Cournède is Associate Professor in Mathematics at the École Centrale Paris, where he graduated (Ingénieur ECP 1997, Ph.D. 2001). His main research areas are Parametric Estimation, Optimization, Stochastic Processes and Modelling of Plant Growth.