A New RBF Neural Network With Boundary Value Constraints

Xia Hong, *Senior Member, IEEE*, and
Sheng Chen, *Senior Member, IEEE*

***Abstract*—We present a novel topology of the radial basis function (RBF) neural network, referred to as the boundary value constraints (BVC)-RBF, which is able to automatically satisfy a set of BVC. Unlike most existing neural networks whereby the model is identified via learning from observational data only, the proposed BVC-RBF offers a generic framework by taking into account both the deterministic prior knowledge and the stochastic data in an intelligent manner. Like a conventional RBF, the proposed BVC-RBF has a linear-in-the-parameter structure, such that it is advantageous that many of the existing algorithms for linear-in-the-parameters models are directly applicable. The BVC satisfaction properties of the proposed BVC-RBF are discussed. Finally, numerical examples based on the combined D-optimality-based orthogonal least squares algorithm are utilized to illustrate the performance of the proposed BVC-RBF for completeness.**

***Index Terms*—Boundary value constraints (BVC), D-optimality, forward regression, radial basis function (RBF), system identification.**

I. Introduction

The radial basis function (RBF) network has been widely studied and applied in system dynamics modeling and prediction [1]–[4]. Most RBF models are constructed to represent a systems' input/output mapping, in which the system output observations are used as the direct target of the model output of RBF networks in training. A fundamental problem in RBF network modeling is to achieve a network with a parsimonious model structure producing good generalization. For general linear in the parameters systems, an orthogonal forward regression (OFR) algorithm based on Gram–Schmidt orthogonal decomposition has extensively been studied [5]–[7]. The OFR algorithm has been a popular tool in associative neural networks such as fuzzy/neurofuzzy systems [8], [9] and wavelets neural networks [10], [11]. The algorithm has also been utilized in a wide range of engineering applications, e.g., aircraft gas turbine modeling [12], fuzzy control of multiple-input–multiple-output nonlinear systems [13], power system control [14], and fault detection [15]. In optimum experimental design [16], D-optimality criterion is regarded as most effective in optimizing the parameter efficiency and model robustness via the maximization of the determinant of the design matrix. In order to achieve a model structure with improved model generalization, the D-optimality-based OFR algorithm is introduced in which D-optimality-based cost function is used in the model searching process [4], [17], [18].

Note that all the aforementioned RBF modeling algorithms are conditioned on that the model is determined based on the observational data only, so that these fit into the statistical learning framework. In many modeling tasks, there are more or less some prior knowledge available. Although any prior knowledge about the system should help to improve the model generalization, in general incorporating the deterministic prior knowledge into a statistically learning paradigm would make the development of modeling algorithms more difficult if not impossible.

In this contribution, we aim to open up a new ground for the RBF by enhancing its capability of automatic constraints satisfaction. We consider a special type of prior knowledge given by a type of boundary value constraints (BVC) and introduce the BVC-RBF as a new topology of RBF neural network that has the capability of automatically satisfying the BVC. The proposed BVC-RBF is constructed and parameterized based on the given BVC. It is shown that the BVC-RBF remains as a linear-in-the-parameter structure just as the conventional RBF does. Therefore, many of the existing modeling algorithms for a conventional RBF are almost directly applicable to the new BVC-RBF without added algorithmic complexity nor computational cost. Consequently, the proposed BVC-RBF effectively lends itself as a single framework in which both the deterministic prior knowledge and stochastic data are fused with ease. For completeness, the combined D-optimality-based orthogonal least squares (OLS) algorithm [4] is used to demonstrate the modeling performance of the proposed BVC-RBF.

II. Problem Formulation

We consider the identification of a semi-unknown system. Defining the system input vector as $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^{\mathrm{T}}$ and the system output as $y(t)$, and given a training data set D_N consisting of N input/output data pairs $\{\mathbf{x}(t), y(t)\}_{t=1}^{N}$, the goal is to find the underlying system dynamics

$$y(t) = f\left(\mathbf{x}(t), \boldsymbol{\theta}\right) + e(t). \tag{1}$$

The underlying function $f : \Re^n \to \Re$ is unknown, and $\boldsymbol{\theta}$ is the vector of associated parameters. $e(t)$ is the noise, which is often assumed to be independent identically distributed with constant variance σ^2. In addition, it is required that the model *strictly* satisfies a set of L BVC given by

$$f(\mathbf{x}_j) = d_j, \qquad j = 1, \ldots, L \tag{2}$$

where $\mathbf{x}_j \in \Re^n$ and $d_j \in \Re$ are known. Note that the information from the given BVC is fundamentally different from that of the observational data set D_N and should differently be treated. The BVC is a deterministic condition, but D_N is subject to observation noise and possesses stochastic characteristics. The BVC may represent the fact that at some critical regions, there is a complete knowledge about the system.

If the underlying function $f(.)$ is represented by a conventional RBF neural network formulated as

$$\hat{y}(t) = \sum_{k=1}^{M} p_k\left(\mathbf{x}(t)\right)\theta_k \tag{3}$$

where $\hat{y}(t)$ is the output of the RBF model. $p_k(.)$ is a known RBF function, given as

$$p_k\left(\mathbf{x}(t)\right) = \Phi\left(v_k(t), \tau\right) \tag{4}$$

$$v_k(t) = \|\mathbf{x}(t) - \mathbf{c}_k\| \tag{5}$$

where $\| \bullet \|$ denotes the Euclidean norm and τ is a positive scalar called width. $\mathbf{c}_k \in \Re^n$, $1 \le k \le M$ are the RBF centers which should appropriately be chosen and sample the input domain. $\Phi(\| \bullet \|, \tau)$ is

Manuscript received February 27, 2008; revised June 24, 2008 and July 11, 2008. This paper was recommended by Associate Editor X. Zeng.

X. Hong is with the School of Systems Engineering, University of Reading, RG6 6AY Berkshire, U.K. (e-mail: xih@cyber.reading.ac.uk; x.hong@reading.ac.uk).

S. Chen is with the School of Electronics and Computer Science, University of Southampton, SO17 1BJ Southampton, U.K.

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/TSMCB.2008.2005124

a chosen RBF function from $\Re^n \to \Re$, e.g., the Gaussian. Typically, for the identification of *a priori* unknown system using D_N only, the RBF network of (3) is determined using $y(t)$ as the target of the RBF model output $\hat{y}(t)$, via some optimization criterion and often in an unconstrained optimization manner. Note that resultant RBF network cannot meet the BVC given by (2). Clearly, the prior knowledge about the system from BVC helps to improve the model generalization, but equally, this makes the modeling process more difficult, since with constraints we are facing a constrained optimization problem. In this contribution, we introduce a simple yet effective treatment to ease the problem.

III. RBF Neural Networks With BVC

Our design goal is to find a new topology of RBF such that the BVC is automatically satisfied, and, as a consequence, the system identification can be carried out without added algorithmic complexity nor computational cost compared to any modeling algorithm for a conventional RBF. The new topology of RBF, shown in Fig. 1, will be parameterized and dependent upon the given BVC as described below.

Consider the following BVC-RBF model representation:

$$\hat{y}(t) = \sum_{k=1}^{M} p_k\left(\mathbf{x}(t)\right)\theta_k + g\left(\mathbf{x}(t)\right) \tag{6}$$

where the proposed RBF is given by

$$p_k\left(\mathbf{x}(t)\right) = h\left(\mathbf{x}(t)\right)\exp\left(-\frac{\|\mathbf{x}(t)-\mathbf{c}_k\|^2}{\tau_1^2}\right) \tag{7}$$

where $h(\mathbf{x}(t)) = \sqrt[L]{\prod_{j=1}^{L}\|\mathbf{x}(t)-\mathbf{x}_j\|}$ is the geometric mean of the data sample $\mathbf{x}(t)$ to the set of boundary values $\mathbf{x}_j$, $j = 1,\ldots,L$. τ_1 is a positive scalar

$$g\left(\mathbf{x}(t)\right) = \sum_{j=1}^{L}\alpha_j\exp\left(-\frac{\|\mathbf{x}(t)-\mathbf{x}_j\|^2}{\tau_2^2}\right) \tag{8}$$

τ_2 is also a positive scalar. α_j is a set of parameters that is obtained by solving a set of linear equations $g(\mathbf{x}_j) = d_j$, $j = 1,\ldots,L$. That is

$$\boldsymbol{\alpha} = \mathbf{G}^{-1}\mathbf{d} \tag{9}$$

where $\boldsymbol{\alpha} = [\alpha_1,\ldots,\alpha_L]^{\mathrm{T}}$, $\mathbf{d} = [d_1,\ldots,d_L]^{\mathrm{T}}$, and $\mathbf{G}$ is given by

$$\mathbf{G} = \begin{pmatrix} 1 & e^{-\frac{\|\mathbf{x}_1-\mathbf{x}_2\|^2}{\tau_2^2}} & \ldots & e^{-\frac{\|\mathbf{x}_1-\mathbf{x}_L\|^2}{\tau_2^2}} \\ e^{-\frac{\|\mathbf{x}_2-\mathbf{x}_1\|^2}{\tau_2^2}} & 1 & \ldots & e^{-\frac{\|\mathbf{x}_2-\mathbf{x}_L\|^2}{\tau_2^2}} \\ \ldots & \ldots & \ldots & \ldots \\ e^{-\frac{\|\mathbf{x}_L-\mathbf{x}_1\|^2}{\tau_2^2}} & e^{-\frac{\|\mathbf{x}_L-\mathbf{x}_2\|^2}{\tau_2^2}} & \ldots & 1 \end{pmatrix}. \tag{10}$$

In the case of the ill conditioning, the regularization technique is applied to the above solution. It is easy to verify that with the proposed topology of BVC-RBF neural networks, the BVC is automatically satisfied. To elaborate, we use a simple 1-D function based on the following parameter setting. $\tau_1 = \tau_2 = 0.5$, and five centers $c_1 = 0.2$, $c_2 = 0.4$, $c_3 = 0.6$, $c_4 = 0.8$, and $c_5 = 1$. A set of two BVC is given by $f(0.1) = -2$, $f(0.5) = 3$. From (9), we obtain $\alpha_1 = -4.9613$ and $\alpha_2 = 5.6161$. For illustration, we construct the five basis functions $p_k(x)$ using (7) and $g(x)$ using (8), as shown in Fig. 2.

From Fig. 2, we note the following basic features.

1) As shown in Fig. 2(a), we see that $p_k(x)$, $k = 1,\ldots,5$ have the properties of *zero forcing* at the boundary points $x_1 = 0.1$, $x_2 = 0.5$. Effectively, the *zero forcing* feature extends to the first term in (6). This means that due to the special network topology, the adjustable parameters θ_k have no effects on the first term in (6) at any of the boundary points.
2) Fig. 2(b) shows that the summation term $g(x)$ has the characteristics of passing all the predetermined boundary values (the required offsets). Consequently, we have $f(0.1) = g(0.1) = -2$, $f(0.5) = g(0.5) = 3$. We also note that $g(x)$ is totally parameterized by the BVC, but does not contain any adjustable parameters dependent on D_N. Effectively, $g(x)$ provides as an offset function for any x.
3) Over the input range as distributed by the RBF centers, the set of smooth functions $p_k(x)$ has diverse local responses and has nonzero adjustable contribution toward $f(x)$ via the adjustable parameters θ_k.
4) We note that all the five basis functions $p_k(x)$ and $g(x)$ are bounded and approach to zero as $x \to \infty$.

In general, $p_k(\mathbf{x}(t))$ and $g(\mathbf{x}(t))$ act as building blocks of the BVC-RBF networks in (6), with a novel feature compared to most of the existent neural networks architecture. That is, by resorting to the given boundary conditions, its topology is designed for the boundary constraints satisfaction, or more generally, for incorporating given prior knowledge. Clearly, the boundary constraints satisfaction property is achieved due to the fact of our choosing $h(\mathbf{x}(t))$ as the geometric mean of the data sample $\mathbf{x}(t)$ to the set of boundary values $\mathbf{x}_j$. However, note that there is no reason to limit the geometric mean as the only choice of $h(\mathbf{x}(t))$ as long as the features above can be maintained. We point out that the basic features listed above are not mathematically vigorous, and how to describe the mathematical properties of a general form of $h(\mathbf{x}(t))$ poses as an open problem.

Note that the boundary condition satisfaction via the network topology is an inherent, but often overlooked, feature for any model representation. For example, the autoregressive with exogenous output model automatically satisfies the boundary condition of $f(\mathbf{0}) = 0$, and for the conventional RBF given by (1) together with the Gaussian basis functions, $f(\infty) = 0$. The aim of this contribution is to introduce and exploit the boundary condition satisfaction via the network topology in a controlled manner, so that the modeling performance may be enhanced by incorporating the *a priori* knowledge via boundary conditions satisfaction.

IV. Identification Algorithm

Substituting (6) into (1) and defining an auxiliary output variable $z(t) = y(t) - g(\mathbf{x}(t))$, we have

$$z(t) = \sum_{k=1}^{M} p_k\left(\mathbf{x}(t)\right)\theta_k + e(t). \tag{11}$$

Based on the model representation of (6), as shown Fig. 1, we suggest a very general two stage training procedure for the identification of BVC-RBF.

1) Determine the offset function $g(\mathbf{x}(t))$ using (9).
2) Apply an existent RBF network identification algorithm to the input/output training data set $\{\mathbf{x}(t), z(t)\}$.

The step 2) above is very general, and its up to the users to decide which identification algorithm to use. Hence, in terms of the training procedure, the difference between the BVC-RBF and RBF is just that the additional step 1) above is required for BVC-RBF.

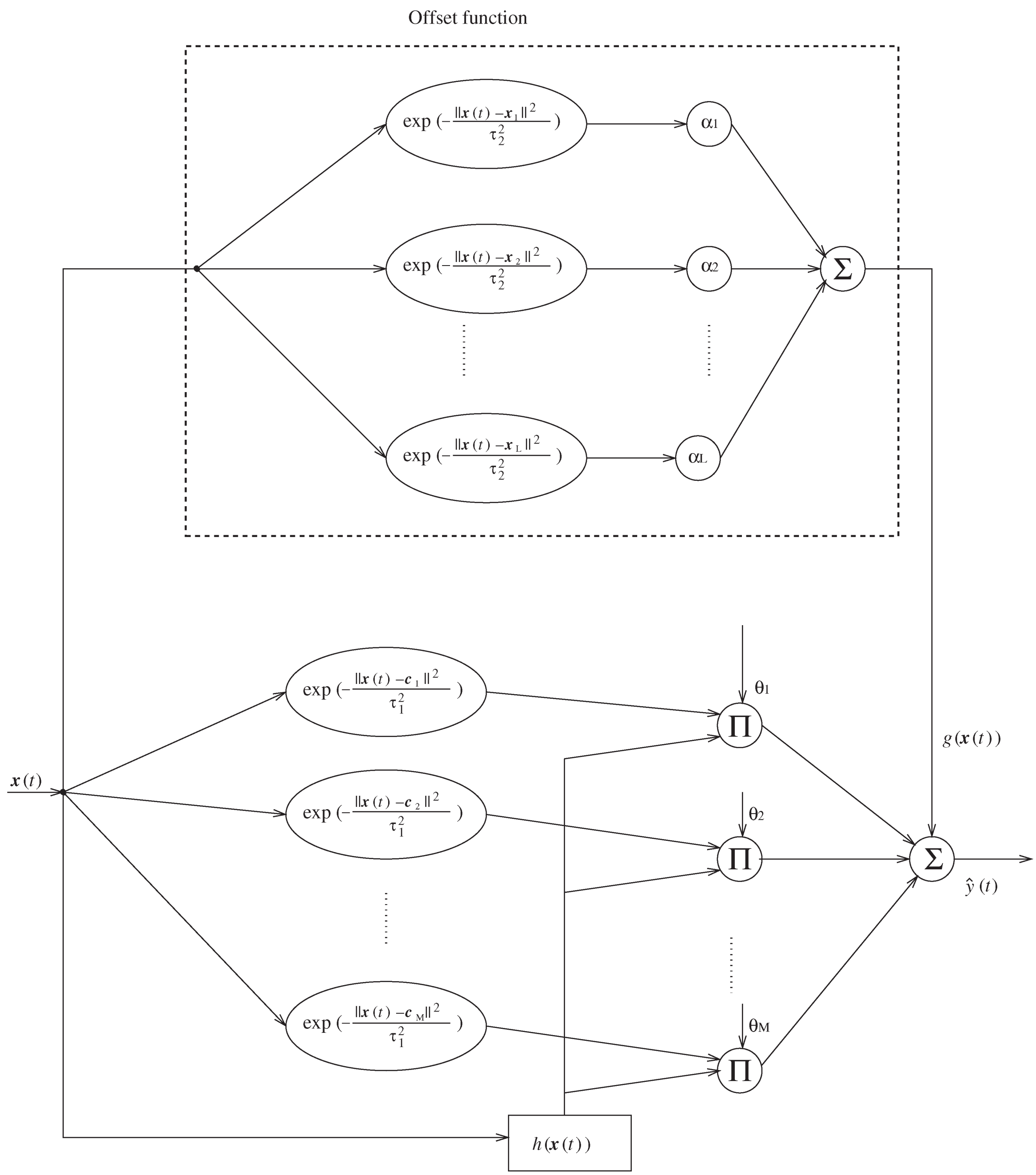

Fig. 1. Graphical illustration of the proposed BVC-RBF neural network.

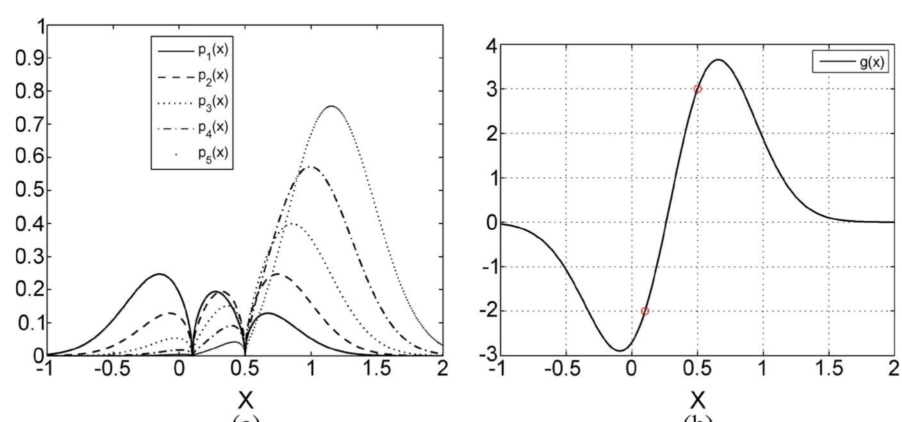

Fig. 2. Illustration of basis functions (a) zero forcing RBFs and (b) offset passing function $g(x)$.

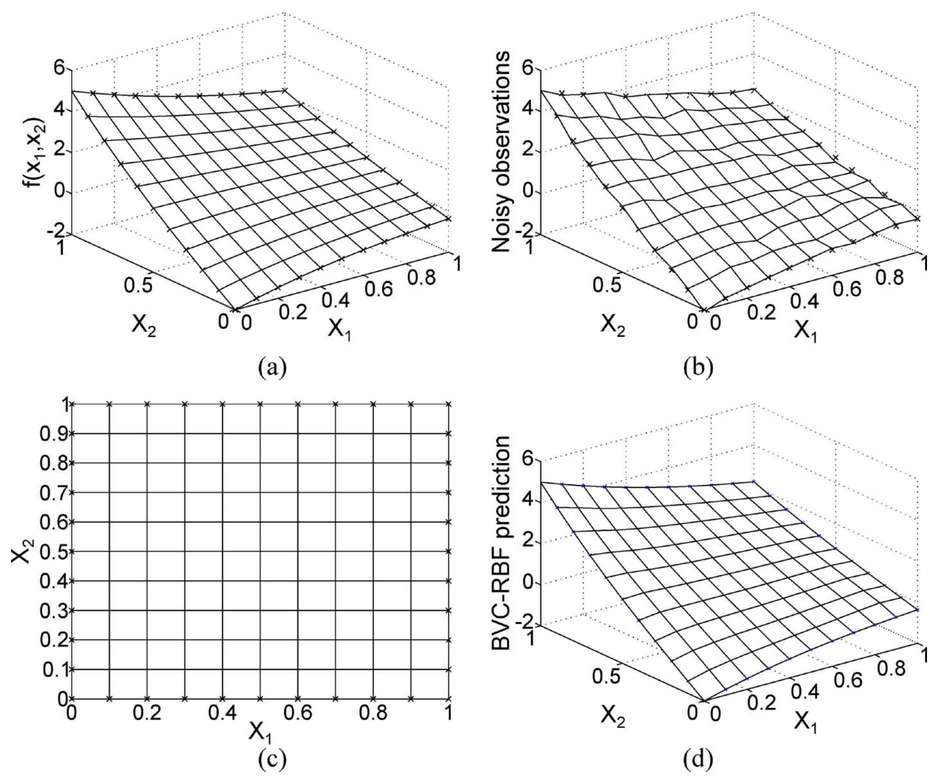

Fig. 3. Example 1. (a) True function $f(x_1, x_2)$. (b) Noisy data $y(x_1, x_2)$. (c) Boundary points. (d) Prediction of the resultant BVC-RBF model.

A practical nonlinear modeling principle is to find the smallest model that generalizes well. Sparse models are preferable in engineering applications since a models' computational complexity scales with its model complexity. Moreover, a sparse model is easier to interpret from the viewpoint of knowledge extraction. Starting with a large candidate regressors set with M regressors, the forward OLS is an efficient nonlinear system identification algorithm [2], [5] which selects regressors in a forward manner by virtue of their contribution to the maximization of the model error reduction ratio. Various forward orthogonal selection algorithms [9], [17], [19]–[22] are directly applicable to the new RBF network with BVC satisfaction without extra computational cost. Clearly, due to the special topology of the new RBF, we note that the formation of data matrices is different from that of the conventional RBF.

Equation (11) can be written in the matrix form as

$$\mathbf{z} = \mathbf{P}\boldsymbol{\theta} + \mathbf{e} \tag{12}$$

where $\mathbf{z} = [z(1), \ldots, z(N)]^{\mathrm{T}}$ is the auxiliary output variable vector, $\theta = [\theta_1, \ldots, \theta_M]^{\mathrm{T}}$ is the parameter vector, $\mathbf{e} = [e(1), \ldots, e(N)]^{\mathrm{T}}$ is the residual vector, and $\mathbf{P}$ is the regression matrix

$$\mathbf{P} = \begin{bmatrix} p_1(\mathbf{x}(1)) & p_2(\mathbf{x}(t)) & \cdots & p_M(\mathbf{x}(1)) \\ p_1(\mathbf{x}(2)) & p_2(\mathbf{x}(t)) & \cdots & p_M(\mathbf{x}(2)) \\ \cdots & \cdots & \cdots & \cdots \\ p_1(\mathbf{x}(N)) & p_2(\mathbf{x}(t)) & \cdots & p_M(\mathbf{x}(N)) \end{bmatrix}.$$

Note that the auxiliary output variable $z(t)$ is used as the target of the first term in (6) (the adjustable part of BVC-RBF). Aiming for improved model robustness, the D-optimality in experimental design [16] has been incorporated in the D-optimality-based model selective criterion [4] to select a set of $n_\theta \ll M$ regressors from M regressors, i.e., to select n_θ columns from $\mathbf{P}$ in a forward regression manner. For completeness, the combined D-optimality-based OLS algorithm [4] is used in the numerical examples.

V. Numerical Examples

Example 1: Consider the partial differential equation given by

$$\left[\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}\right] f(x_1, x_2) = e^{-x_1}\left(x_1 - 2 + x_2^3 + 6x_2\right)$$
$$x_1 \in [0, 1], \qquad x_2 \in [0, 1] \tag{13}$$

with the boundary conditions given by

$$f(0, x_2) = x_2^3 \tag{14}$$
$$f(1, x_2) = \left(1 + x_2^3\right)/e \tag{15}$$
$$f(x_1, 0) = x_1 e^{-x_1} \tag{16}$$
$$f(x_1, 1) = e^{-x_1}(1 + x_1). \tag{17}$$

The analytic solution is $f(x_1, x_2) = e^{-x_1}(x_1 + x_2^3)$, by which a 11×11 meshed data set $f(x_1, x_2)$ is generated, as shown in Fig. 3(a). By using $y(x_1, x_2) = f(x_1, x_2) + e(x_1, x_2)$, where $e(x_1, x_2) \sim N(0, 0.1^2)$, the training data set D_N consists of $(N = 121)$ samples of $\{x_1, x_2, y(x_1, x_2)\}$) and is as shown in Fig. 3(b). By sampling data according to (14)–(17), 40 BVC points are produced, and Fig. 3(c) plots the input part as cross points $(L = 40)$. The input part of all the training data set $\{x_1, x_2\}$ is used as the candidate center set $(M = 121)$. $\tau_1 = 0.6$, $\tau_2 = 0.6$ are empirically chosen and used in the candidate BVC-RBF basis functions. Note that ultimate goal is to find a model that is closest to the unknown function. It is difficult to define the analytic form of the ultimate goal with respect to τ_1 and τ_2. However, the technique of cross validation and the grid search may be used to choose τ_1 and τ_2. In conventional RBF model, it is known the modeling performance is not sensitive to τ in a range of suitable values. This means that a coarse search will be sufficient. The combined D-optimality-based OLS algorithm was applied [4] to identify a sparse BVC-RBF model, in which the adjustable parameter in the D-optimality-based cost function (β in [4]) was set as 10^{-4}. The model prediction of the resultant BVC-RBF model is shown in Fig. 3(d). For comparison, a sparse conventional

TABLE I
COMPARISON BETWEEN THE CONVENTIONAL RBF AND THE PROPOSED BVC-RBF MODEL FOR EXAMPLE 1

	Model size n_θ	MSE $\frac{1}{N}\sum(\hat{y}-f)^2$	MSE $\frac{1}{N}\sum(\hat{y}-y)^2$	MSE (boundary) $\frac{1}{L}\sum_j(\hat{y}(\mathbf{x}_j)-d_j)^2$
BVC-RBF	5	7.1665×10^{-4}	0.0078	2.9943×10^{-10}
RBF	10	0.0039	0.0102	0.006

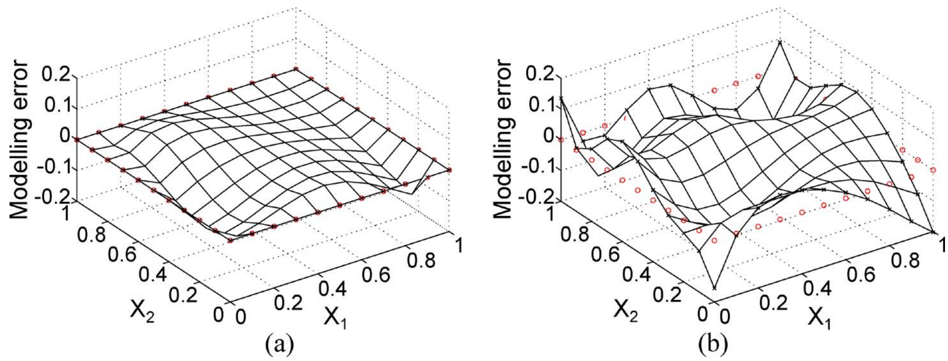

Fig. 4. Modeling error between the true function and the model prediction $(\hat{y}(x_1,x_2)-f(x_1,x_2))$ for Example 1. (a) BVC-RBF model. (b) RBF model.

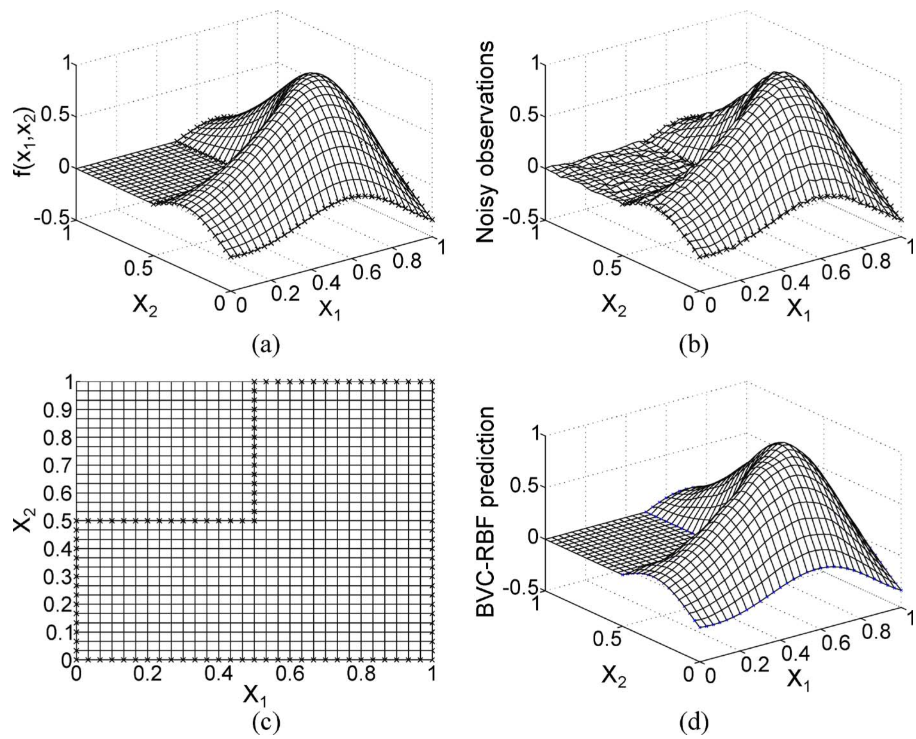

Fig. 5. Example 2. (a) True function $f(x_1,x_2)$. (b) Noisy data $y(x_1,x_2)$. (c) Boundary points. (d) Prediction of the resultant BVC-RBF model.

RBF model was identified using the combined D-optimality-based OLS algorithm [4]. The Gaussian basis function with an empirically set $\tau=0.6$ was used. Candidate basis functions are generated using the same training data set and the same candidate center set. The adjustable parameter in the D-optimality-based cost function was also set as 10^{-4}. The comparative results are as shown in Table I and Fig. 4. The BVC-RBF has much better performance in terms of the modeling errors to the true function, as a result of making use of BVC. Note that Fig. 4 has shown that the BVC cannot be satisfied by the conventional RBF, but the proposed BVC-RBF model inherently satisfies the BVC via the topology.

Example 2: The Matlab logo was generated by the first eigenfunction of the L-shaped membrane. A 31 × 31 meshed data set $f(x_1,x_2)$ is generated by using Matlab command *membrane.m*, which is defined over a unit square input region $x_1\in[0,1]$ and $x_2\in[0,1]$. The data set $y(x_1,x_2)=f(x_1,x_2)+e(x_1,x_2)$ is then generated by adding a noise term $e(x_1,x_2)\sim N(0,0.01^2)$. The true function $f(x_1,x_2)$ is shown in Fig. 5(a), and the noisy data set $y(x_1,x_2)$ is shown in Fig. 5(b). In Fig. 5(c), the BVC is marked as cross point, and there are $L=120$ boundary points, given by the coordinates of $\{x_1,x_2,f(x_1,x_2)\}$. We use all the data points within the boundary as the training data set D_N consisting of the set of $\{x_1,x_2,y(x_1,x_2)\}$ coordinates ($N=721$). The input part of all the training data set $\{x_1,x_2\}$ is used as the candidate center set ($M=721$). $\tau_1=0.1$, $\tau_2=0.2$ were predetermined in order to generate the candidate BVC-RBF basis functions. The combined D-optimality-based OLS algorithm was applied [4] to identify a sparse model, in which the D-optimality-based cost function's adjustable parameter (denoted as

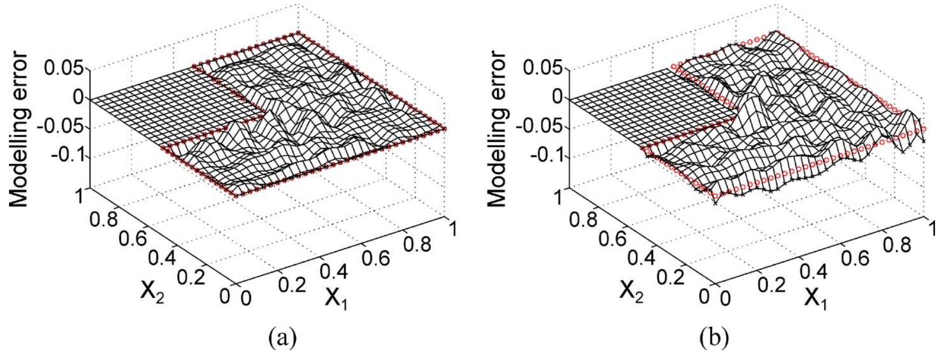

Fig. 6. Modeling error between the true function and the model prediction $(\hat{y}(x_1, x_2) - f(x_1, x_2))$ for Example 2. (a) BVC-RBF model. (b) RBF model.

TABLE II
COMPARISON BETWEEN THE CONVENTIONAL RBF AND THE PROPOSED BVC-RBF MODEL FOR EXAMPLE 2

	Model size n_θ	MSE $\frac{1}{N}\sum(\hat{y}-f)^2$	MSE $\frac{1}{N}\sum(\hat{y}-y)^2$	MSE (boundary) $\frac{1}{L}\sum_j(\hat{y}(\mathbf{x}_j)-d_j)^2$
BVC-RBF	68	4.3787×10^{-5}	1.0736×10^{-4}	7.2598×10^{-11}
RBF	91	1.0229×10^{-4}	1.6894×10^{-4}	2.1249×10^{-4}

β in [4]) was set as 10^{-6}. Fig. 5(d) shows the excellent performance of the resultant BVC-RBF model. For comparison, the combined D-optimality-based OLS algorithm was applied [4] to identify a sparse conventional RBF model. The Gaussian basis function with a predetermined $\tau = 0.1$ was used to generate candidate basis functions from the same training data set and the same candidate center set. The adjustable parameter in the D-optimality-based cost function was also set as 10^{-6}. The comparative results are shown in both Fig. 6 and Table II. It is shown that the BVC-RBF can achieve significant improvement over the RBF in terms of the modeling performance to the true function. In particular, we note that the BVC can be satisfied with the proposed BVC-RBF model, but not by the conventional RBF, as clearly shown in Fig. 6.

VI. CONCLUSION

A new topology of RBF neural network has been introduced for a type of modeling problems in which a set of BVC is given in addition to an observational data set. A significant advantage of the proposed BVC-RBF is that the BVC satisfaction is taken into account by the network architecture, rather than by the learning algorithm. Consequently, the resultant model maintains a linear-in-the-parameter structure such that many of the existing linear-in-the-parameters learning algorithms are readily applicable. Future work will investigate other RBF topology for other types of BVC.

REFERENCES

[1] J. E. Moody and C. J. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Comput.*, vol. 1, no. 2, pp. 281–294, 1989.
[2] S. Chen, C. F. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Netw.*, vol. 2, no. 2, pp. 302–309, Mar. 1991.
[3] W. Pedrycz, "Conditional fuzzy clustering in the design of radial basis function neural networks," *IEEE Trans. Neural Netw.*, vol. 9, no. 4, pp. 601–612, Jul. 1998.
[4] X. Hong and C. J. Harris, "Experimental design and model construction algorithms for radial basis function networks," *Int. J. Syst. Sci.*, vol. 34, no. 14/15, pp. 733–745, 2003.
[5] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Control*, vol. 50, pp. 1873–1896, 1989.
[6] S. Chen, Y. Wu, and B. L. Luk, "Combined genetic algorithm optimization and regularized orthogonal least squares learning for radial basis function networks," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1239–1243, Sep. 1999.
[7] M. J. L. Orr, "Regularisation in the selection of radial basis function centers," *Neural Comput.*, vol. 7, no. 3, pp. 954–975, May 1995.
[8] L. Wang and J. M. Mendel, "Fuzzy basis functions, universal approximation, and orthogonal least-squares learning," *IEEE Trans. Neural Netw.*, vol. 5, no. 5, pp. 807–814, Sep. 1992.
[9] X. Hong and C. J. Harris, "Neurofuzzy design and model construction of nonlinear dynamical processes from data," *Proc. Inst. Elect. Eng.—Control Theory Appl.*, vol. 148, no. 6, pp. 530–538, Nov. 2001.
[10] Q. Zhang, "Using wavelet network in nonparametric estimation," *IEEE Trans. Neural Netw.*, vol. 8, no. 2, pp. 227–236, Mar. 1993.
[11] S. A. Billings and H. L. Wei, "The wavelet-NARMAX representation: A hybrid model structure combining polynomial models with multiresolution wavelet decompositions," *Int. J. Syst. Sci.*, vol. 36, no. 3, pp. 137–152, Feb. 2005.
[12] N. Chiras, C. Evans, and D. Rees, "Nonlinear gas turbine modeling using NARMAX structures," *IEEE Trans. Instrum. Meas.*, vol. 50, no. 4, pp. 893–898, Aug. 2001.
[13] Y. Gao and M. J. Er, "Online adaptive fuzzy neural identification and control of a class of MIMO nonlinear systems," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 4, pp. 462–477, Aug. 2003.
[14] K. M. Tsang and W. L. Chan, "Adaptive control of power factor correction converter using nonlinear system identification," *Proc. Inst. Elect. Eng.—Elect. Power Appl.*, vol. 152, no. 3, pp. 627–633, May 2005.
[15] G. C. Luh and W. C. Cheng, "Identification of immune models for fault detection," in *Proc. Inst. Mech. Eng., Part I, J. Syst. Control Eng.*, 2004, vol. 218, pp. 353–367.
[16] A. C. Atkinson and A. N. Donev, *Optimum Experimental Designs*. Oxford, U.K.: Clarendon, 1992.
[17] X. Hong and C. J. Harris, "Nonlinear model structure design and construction using orthogonal least squares and D-optimality design," *IEEE Trans. Neural Netw.*, vol. 13, no. 5, pp. 1245–1250, Sep. 2002.
[18] S. Chen, "Locally regularised orthogonal least squares algorithm for the construction of sparse kernel regression models," in *Proc. 6th Int. Conf. Signal Process.*, Beijing, China, 2002, pp. 1229–1232.
[19] X. Hong and J. Harris, "Nonlinear model structure detection using optimum experimental design and orthogonal least squares," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 435–439, Mar. 2001.
[20] S. Chen, X. Hong, and C. J. Harris, "Sparse kernel regression modeling using combined locally regularized orthogonal least squares and D-optimality experimental design," *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 1029–1036, Jun. 2003.
[21] S. Chen, X. Hong, and C. J. Harris, "Sparse multioutput radial basis function network construction using combined locally regularised orthogonal least square and D-optimality experimental design," *Proc. Inst. Elect. Eng.—Control Theory Appl.*, vol. 150, no. 2, pp. 139–146, Mar. 2003.
[22] S. Chen, X. Hong, C. J. Harris, and P. M. Sharkey, "Sparse modeling using orthogonal forward regression with PRESS statistic and regularization," *IEEE Trans. Syst., Man Cybern. B, Cybern.*, vol. 34, no. 2, pp. 898–911, Apr. 2004.