

# GREY-BOX RADIAL BASIS FUNCTION MODELLING: THE ART OF INCORPORATING PRIOR KNOWLEDGE

Sheng Chen, Chris J. Harris

School of Electronics and Computer Science  
University of Southampton  
Southampton SO17 1BJ, UK

Xia Hong

School of Systems Engineering  
University of Reading  
Reading RG6 6AY, UK

## ABSTRACT

A basic principle in data modelling is to incorporate available *a priori* information regarding the underlying data generating mechanism into the modelling process. We adopt this principle and consider grey-box radial basis function (RBF) modelling capable of incorporating prior knowledge. Specifically, we show how to explicitly incorporate the two types of prior knowledge: the underlying data generating mechanism exhibits known symmetric property and the underlying process obeys a set of given boundary value constraints. The class of orthogonal least squares regression algorithms can readily be applied to construct parsimonious grey-box RBF models with enhanced generalisation capability.

**Index Terms**— Radial basis function network, grey-box modelling, symmetry, boundary value constraint

## 1. INTRODUCTION

The radial basis function (RBF) network has found applications in diverse fields of engineering, and the class of orthogonal least squares (OLS) regression algorithms [1, 2, 3, 4] offers powerful and efficient tools for constructing parsimonious RBF models that generalise well. This approach is equally applicable to the supervised regression [1, 2, 3, 4] and classification [5] as well as the unsupervised density estimation [6]. Like many other data modelling approaches, the RBF model constitutes a black-box data modelling approach. Adopting a black-box modelling is appropriate if no *a priori* information exists regarding the underlying data generating mechanism. However, if there are known prior knowledge concerning the underlying process, they should be incorporated into the model structure explicitly. The use of prior knowledge in data modelling often leads to enhanced modelling performance. A discussion on learning from known prior knowledge or hints is given in [7]. A few works have exploited the symmetric properties of some underlying systems in regression applications [8, 9] as well as in classification problems [10].

How to incorporating prior knowledge to form grey-box RBF model is highly problem dependent. In this contribu-

tion, we specifically consider two types of *a priori* information. In the first type of data modelling problems, the underlying data generating mechanism exhibits known symmetric property and we introduce the symmetric RBF (SRBF) model that guarantees to possess the known symmetry. For the second type of applications, the underlying process obeys a set of given boundary value constraints (BVCs) and we adopt the novel BVC-RBF structure which automatically meets the given BVCs. All the learning algorithms originally derived for the black-box RBF model can be applied to these two grey-box RBF models without the need for any modification. In particular, the class of OLS regression algorithms provides efficient means of building parsimonious grey-box RBF models with improved generalisation performance.

## 2. SYMMETRIC RBF MODELLING

Consider the training data set  $D_K = \{\mathbf{x}(k), y(k)\}_{k=1}^K$  that is generated by the underlying system

$$y(k) = f(\mathbf{x}(k)) + \epsilon(k) \quad (1)$$

where  $\mathbf{x}(k) = [x_1(k) \ x_2(k) \ \dots \ x_n(k)]^T$  is the system input vector,  $y(k)$  the noisy system output, and  $\epsilon(k)$  represents a white observation noise with variance  $\sigma_\epsilon^2$ . The system mapping  $f : \mathcal{R}^n \rightarrow \mathcal{R}$  is unknown. However, the system  $f$  is known to possess the odd symmetry

$$f(-\mathbf{x}) = -f(\mathbf{x}). \quad (2)$$

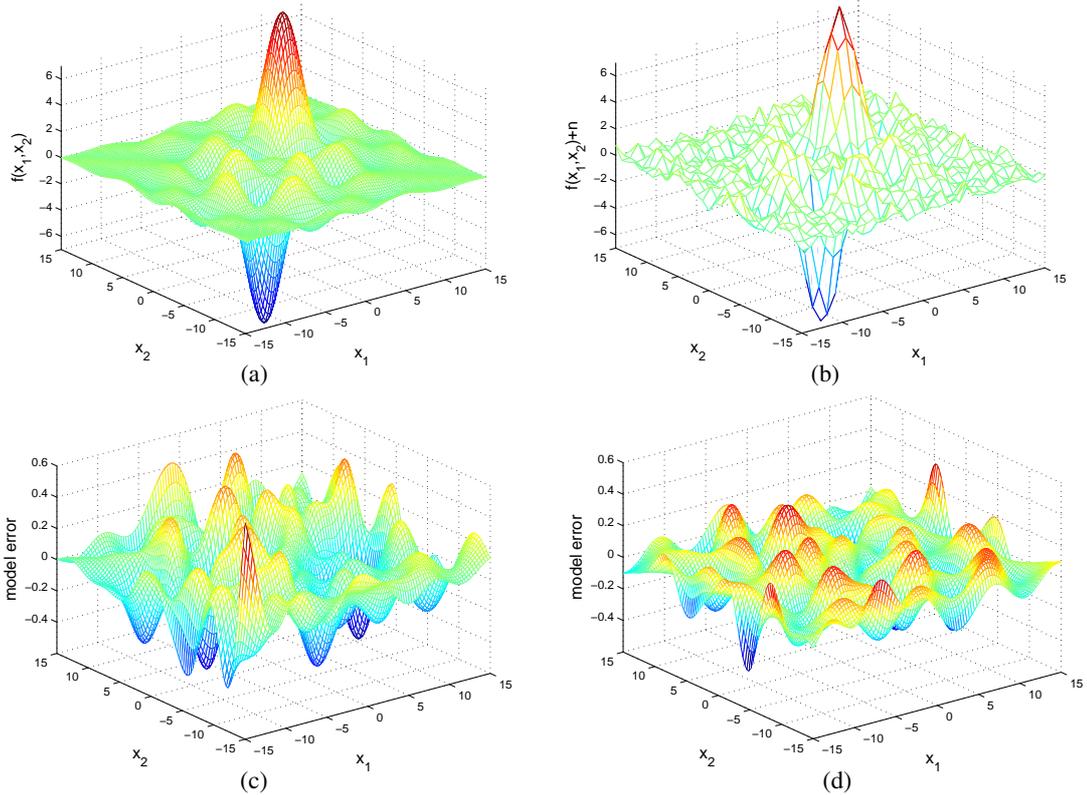
Even symmetry can be treated in a similar way. The goal is to construct the RBF model

$$\hat{y}(k) = \hat{f}(\mathbf{x}(k); \boldsymbol{\theta}) = \sum_{i=1}^M p_i(\mathbf{x}(k); \rho) \theta_i \quad (3)$$

based on  $D_K$  to capture the underlying system  $f$ , where  $\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \dots \ \theta_M]^T$  denotes the RBF weight vector,  $\rho$  is the RBF variance and  $M$  the number of RBF nodes.

To explicitly incorporate the prior knowledge (2), we adopt the following symmetric RBF node

$$p_i(\mathbf{x}; \rho) = \varphi(\|\mathbf{x} - \mathbf{c}_i\|/\rho) - \varphi(\|\mathbf{x} + \mathbf{c}_i\|/\rho), \quad (4)$$



**Fig. 1.** (a) The underlying function  $f(x_1, x_2)$  shown on the grid of 90601 points, (b) 961 noisy training data points, (c) modelling error  $f(x_1, x_2) - \hat{f}(x_1, x_2)$  of the standard RBF model, and (d) modelling error  $f(x_1, x_2) - \hat{f}(x_1, x_2)$  of the SRBF model, for Example One.

where  $\mathbf{c}_i$  is the  $i$ th RBF centre vector and  $\varphi(\bullet)$  the usual RBF basis function. In this study, the Gaussian basis function is used. With this symmetric node structure, the SRBF model has the same odd symmetry as the underlying system. By contrast, the conventional RBF model with the node structure  $p_i(\mathbf{x}; \rho) = \varphi(\|\mathbf{x} - \mathbf{c}_i\|/\rho)$  does not guarantee to have this symmetric property. The grey-box RBF model (3) with the symmetric node structure (4) has the same regression modelling form as the conventional RBF model. Therefore, the OLS regression algorithms [1, 2, 3, 4] can be used to identify a sparse SRBF model based on  $D_K$ .

**Example One.** The system to be identified was given by

$$f(x_1, x_2) = 10 \frac{\sin(x_1 - 5) \sin(x_2 - 5)}{(x_1 - 5)(x_2 - 5)} - \frac{\sin(x_1 + 5) \sin(x_2 + 5)}{(x_1 + 5)(x_2 + 5)} \quad (5)$$

This system has odd symmetry and  $f(x_1, x_2)$  is plotted in Fig. 1 (a) using a grid of 90601 points. The training set  $D_K$  contained 961 noisy data points as shown in Fig. 1 (b), where

**Table 1.** Performance comparison between the conventional RBF and SRBF models for Example One.

	model size	Training MSE	Test MSE	MME
RBF	105	0.1543	0.2047	0.0294
SRBF	68	0.1566	0.1839	0.0093

the system noise  $\varepsilon(k)$  was a white Gaussian noise with variance  $\sigma_\varepsilon^2 = 0.16$ . The local regularisation assisted OLS algorithm based on leave-one-out error criterion [4] was used to *automatically* identify both the conventional RBF and SRBF models. The RBF variance  $\rho = 8.0$  was determined separately using cross validation. A separate test data set of 961 noisy data points was also generated to compute the mean square error (MSE) according to

$$\text{MSE} = E[(y(k) - \hat{y}(k))^2] = \frac{1}{K} \sum_{k=1}^K (y(k) - \hat{y}(k))^2. \quad (6)$$

The generalisation performance can be computed with the mean modelling error (MME)

$$\text{MME} = E[(f(x_1, x_2) - \hat{f}(x_1, x_2))^2] \quad (7)$$

by averaging over the grid of 90601 points, where  $\hat{f}(x_1, x_2)$  denotes the identified model mapping. Table 1 compares the performance of the two models obtained. Fig. 1 (c) and (d) show the modelling error  $f(x_1, x_2) - \hat{f}(x_1, x_2)$  on the grid of 90601 points for the two obtained models, respectively.

### 3. BVC-RBF MODELLING

Again consider the identification of the unknown system  $f$  of (1) using the RBF model (3) based on the noisy training

data set  $D_K$ . In addition, the unknown system mapping  $f$  is known to satisfy a set of  $L$  BVCs given by

$$f(\mathbf{x}_j) = d_j, \quad 1 \leq j \leq L, \quad (8)$$

where  $\mathbf{x}_j \in \mathcal{R}^n$  and  $d_j \in \mathcal{R}$  are known. These BVCs may represent the fact that at some critical regions, there is a complete knowledge about the system and any identified model  $\hat{f}$  is required to *strictly* meet these BVCs, that is,

$$\hat{f}(\mathbf{x}_j; \boldsymbol{\theta}) = d_j, \quad 1 \leq j \leq L. \quad (9)$$

It is obvious that a conventional RBF model with the node structure  $p_i(\mathbf{x}; \rho) = \varphi(\|\mathbf{x} - \mathbf{c}_i\|/\rho)$  cannot guarantee to satisfy the known set of BVCs. Conventional way of incorporating the BVCs (9) as a set of equality constraints in learning will complicate the resulting optimisation problem. We propose the novel BVC-RBF network model which has the capacity of satisfying the given BVCs automatically without added algorithmic complexity and computational cost.

The proposed BVC-RBF model takes the form

$$\hat{y}(k) = \hat{f}(\mathbf{x}(k); \boldsymbol{\theta}) = \sum_{i=1}^M p_i(\mathbf{x}(k); \rho) \theta_i + g(\mathbf{x}(k)) \quad (10)$$

with the novel RBF node structure

$$p_i(\mathbf{x}; \rho) = h(\mathbf{x}) \varphi(\|\mathbf{x} - \mathbf{c}_i\|/\rho), \quad (11)$$

where

$$h(\mathbf{x}) = \sqrt[L]{\prod_{j=1}^L \|\mathbf{x} - \mathbf{x}_j\|} \quad (12)$$

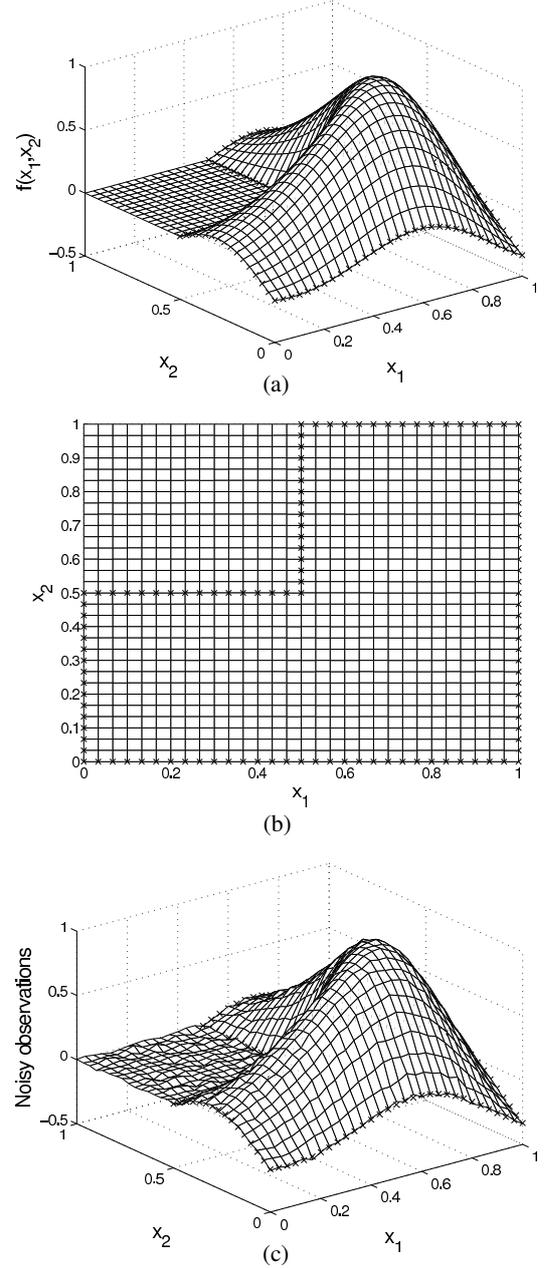
is the geometric mean of the data sample  $\mathbf{x}$  to the set of boundary values  $\mathbf{x}_j$ ,  $1 \leq j \leq L$ ,

$$g(\mathbf{x}) = \sum_{j=1}^L \alpha_j e^{-\frac{\|\mathbf{x} - \mathbf{x}_j\|^2}{\tau}} \quad (13)$$

with  $\tau$  being a positive scalar, and  $\boldsymbol{\alpha} = [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_L]^T$  is a set of parameters that is obtained by solving the set of linear equations  $g(\mathbf{x}_j) = d_j$ ,  $1 \leq j \leq L$ , as follows.  $\boldsymbol{\alpha} = \mathbf{G}^{-1} \mathbf{d}$ , where  $\mathbf{d} = [d_1 \ d_2 \ \cdots \ d_L]^T$  and

$$\mathbf{G} = \begin{bmatrix} 1 & e^{-\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{\tau}} & \cdots & e^{-\frac{\|\mathbf{x}_1 - \mathbf{x}_L\|^2}{\tau}} \\ e^{-\frac{\|\mathbf{x}_2 - \mathbf{x}_1\|^2}{\tau}} & 1 & \cdots & e^{-\frac{\|\mathbf{x}_2 - \mathbf{x}_L\|^2}{\tau}} \\ \vdots & \ddots & \ddots & \vdots \\ e^{-\frac{\|\mathbf{x}_L - \mathbf{x}_1\|^2}{\tau}} & e^{-\frac{\|\mathbf{x}_L - \mathbf{x}_2\|^2}{\tau}} & \cdots & 1 \end{bmatrix}. \quad (14)$$

In the case of ill-conditioning, a regularisation technique can be applied to the above solution. It is easy to verify that with this proposed topology of BVC-RBF model, the BVCs (8) are automatically satisfied. To elaborate further, we note the following features of the BVC-RBF structure.



**Fig. 2.** (a) The underlying function  $f(x_1, x_2)$  shown on the grid of 961 points, (b) the  $L = 120$  BVCs marked as cross points, and (c) 961 noisy training data points, for Example Two.

1. The BVC-RBF nodes (11) have the property of *zero forcing* at the boundary points  $\mathbf{x}_j$ ,  $1 \leq j \leq L$ , and the adjustable RBF weights  $\theta_i$  have no effects on the first term in (10) at any of the boundary points.
2. The term  $g(\mathbf{x})$  passes all the predetermined boundary values  $f(\mathbf{x}_j) = g(\mathbf{x}_j) = d_j$ ,  $1 \leq j \leq L$ , and it is completely determined by the BVCs but does not contain any adjustable parameters dependent on  $D_K$ .
3. Over the input range, the set of smooth BVC-RBF nodes  $p_i(\mathbf{x}; \rho)$  has diverse local responses, and has non-zero

**Table 2.** Performance comparison between the conventional RBF and BVC-RBF models for Example Two.

	model size	training MSE (inside $D_K$ )	MME (inside boundary)	MME (on boundary)
RBF	91	$1.6894 \times 10^{-4}$	$1.0229 \times 10^{-4}$	$2.1249 \times 10^{-4}$
BVC-RBF	68	$1.0736 \times 10^{-4}$	$4.3787 \times 10^{-5}$	$7.2598 \times 10^{-11}$

adjustable contribution towards modelling  $f(\mathbf{x})$  via the adjustable parameters  $\theta_i$  based on  $D_K$ .

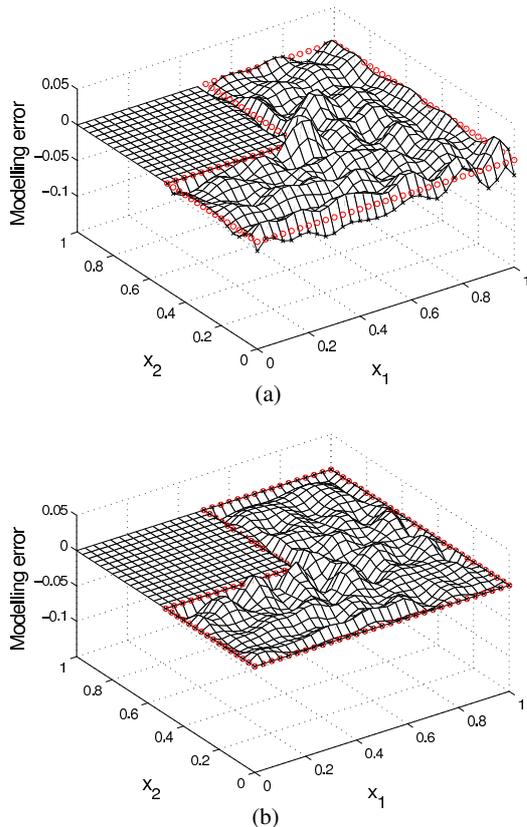
With this BVC-RBF model, no constrained optimisation is needed and the OLS algorithms [1, 2, 3, 4] can be used to identify a sparse BVC-RBF model based on  $D_K$ .

**Example Two.** A  $31 \times 31$  meshed data set  $f(x_1, x_2)$ , as depicted in Fig. 2 (a), was generated by using Matlab command *membrane.m*. In Fig. 2 (b), the required  $L = 120$  BVCs, given by the coordinates of  $\{(x_1, x_2), f(x_1, x_2)\}$ , are marked as cross points. The noisy training data set  $D_K$  was generated by adding a white Gaussian noise of variance  $\sigma_\varepsilon^2 = 0.01^2$  to  $f(x_1, x_2)$ , and  $D_K$  is plotted in Fig. 2 (c). We used all the data points of  $D_K$  that were inside the boundary as training samples and applied the OLS regression algorithm based on the combined training MSE and  $D$ -optimality criterion [3] to construct both the RBF and BVC-RBF models. The RBF variance  $\rho = 0.01$  was determined separately based on cross validation. For the offset function (13),  $\tau = 0.04$  was found

to be appropriate. The weighting for the combined cost function of the training MSE and  $D$ -optimality [3] was chosen to be  $10^{-6}$ . Table 2 compares the performance of the conventional RBF model obtained with that of the novel BVC-RBF model constructed, where the sizes of the two models were *automatically* determined by the learning algorithm. Fig. 3 (a) and (b) depict the modelling error  $f(x_1, x_2) - \hat{f}(x_1, x_2)$  of the two obtained models, respectively, where  $\hat{f}$  denotes the model mapping identified.

#### 4. REFERENCES

- [1] 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.
- [2] S. Chen, C. F. N. Cowan and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Networks*, vol. 2, pp. 302–309, March 1991.
- [3] S. Chen, X. Hong and C. J. Harris, "Sparse kernel regression modelling using combined locally regularized orthogonal least squares and  $D$ -optimality experimental design," *IEEE Trans. Automatic Control*, vol. 48, pp. 1029–1036, June 2003.
- [4] S. Chen, X. Hong, C. J. Harris and P. M. Sharkey, "Sparse modelling using orthogonal forward regression with PRESS statistic and regularization," *IEEE Trans. Systems, Man and Cybernetics, Part B*, vol. 34, pp. 898–911, April 2004.
- [5] X. Hong, S. Chen and C. J. Harris, "A fast linear-in-the-parameters classifier construction algorithm using orthogonal forward selection to minimize leave-one-out misclassification rate," *Int. J. Systems Sci.*, vol. 39, pp. 119–125, 2008.
- [6] S. Chen, X. Hong and C. J. Harris, "An orthogonal forward regression technique for sparse kernel density estimation," *Neurocomputing*, vol. 71, pp. 931–943, January 2008.
- [7] Y. S. Abu-Mostafa, "An algorithm for learning from hints," in *Proc. IJCNN'93*, Nagoya, Japan, Oct.25-29, 1993, vol. 2, pp. 1653–1656.
- [8] L. A. Aguirre, R. A. M. Lopes, G. F. V. Amaral and C. Letellier, "Constraining the topology of neural networks to ensure dynamics with symmetry properties," *Physical Review E*, vol. 69, pp. 026701–1–026701–11, 2004.
- [9] M. Espinoza, J. A. K. Suykens and B. De Moor, "Imposing symmetry in least squares support vector machines regression," in *Proc. Joint 44th IEEE CDC and European Control Conf. 2005*, Seville, Spain, Dec.12-15, 2005, pp. 5716–5721.
- [10] S. Chen, A. Wolfgang, C. J. Harris and L. Hanzo, "Symmetric RBF classifier for nonlinear detection in multiple-antenna aided systems," *IEEE Trans. Neural Networks*, vol. 19, pp. 737–745, May 2008.



**Fig. 3.** (a) Modelling error  $f(x_1, x_2) - \hat{f}(x_1, x_2)$  of the standard RBF model, and (b) modelling error  $f(x_1, x_2) - \hat{f}(x_1, x_2)$  of the BVC-RBF model, for Example Two.