

# Complex-Valued Symmetric Radial Basis Function Network for Beamforming

Sheng Chen  
School of Electronics and Computer Sciences  
University of Southampton  
Southampton SO17 1BJ, United Kingdom  
E-mail: sqc@ecs.soton.ac.uk

## ABSTRACT

The complex-valued radial basis function (RBF) network proposed by Chen *et al.* (1994) has found many applications for processing complex-valued signals, in particular, for communication channel equalisation and signal detection. This complex-valued RBF network, like many other existing RBF modelling methods, constitutes a black-box approach that seeks typically a sparse model representation extracted from the training data. Adopting black-box modelling is appropriate, if no *a priori* information exists regarding the underlying data generating mechanism. However, a fundamental principle in practical data modelling is that if there exists *a priori* information concerning the system to be modelled it should be incorporated in the modelling process. Many complex-valued signal processing problems, particularly those encountered in communication signal detection, have some inherent symmetric properties. This contribution adopts a grey-box approach to complex-valued RBF modelling and develops a complex-valued symmetric RBF (SRBF) network model. The application of this SRBF network is demonstrated using nonlinear beamforming assisted detection for multiple-antenna aided wireless systems that employ complex-valued modulation schemes. Two training algorithms for this complex-valued SRBF network are proposed. The first method is based on a modified version of the cluster-variation enhanced clustering algorithm, while the second method is derived by modifying the orthogonal-forward-selection procedure based on Fisher ratio of class separability measure. The effectiveness of the proposed complex-valued SRBF network and the efficiency of the two training algorithms are demonstrated in nonlinear beamforming application.

## INTRODUCTION

The radial basis function (RBF) network is a popular artificial neural network (ANN) architecture that has found wide-ranging applications in many diverse fields of engineering, see for example, Chen *et al.* (1990); Leonard & Kramer (1991); Chen *et al.* (1993); Caiti & Parisini (1994); Gorinevsky *et al.* (1996); Cha & Kassam (1996); Rosenblum & Davis (1996); Refaee *et al.* (1999); Muraki *et al.* (2001); Mukai *et al.* (2002); Su *et al.* (2002); Li *et al.* (2004); Lee & Choi (2004); Ng *et al.* (2004); Oyang *et al.* (2005); Acir *et al.* (2005); Tan *et al.* (2005). The RBF method is a classical numerical technique for nonlinear functional interpolation with real-valued data (Powell, 1987). A renewed interest in the RBF method coincided with a recent resurgence in the field of ANNs. Connections between the RBF method and the ANN was made and the RBF model was re-interpreted as a one-hidden-layer feedforward network (Broomhead & Lowe, 1988; Poggio & Girosi, 1990). Specifically, by adopting the ANN interpretation, a RBF model can be considered as a processing structure consisting of a hidden layer and an output layer. Each node in the hidden layer has a radially symmetric response around a node parameter vector called a centre, with the hidden node's response shape determined by the chosen basis function as well as a node width parameter, while the output layer is a set of linear combiners with linear connection weights.

The parameters of the RBF network include its centre vectors and variances or covariance matrices of the basis functions as well as the weights that connect the RBF nodes to the network output. All the parameters of a RBF network can be learned together via nonlinear optimisation using the gradient based algorithms (Chen *et al.*, 1990a; An *et al.*, 1993; McLoone *et al.*, 1998; Karayiannis & Randolph-Gips, 2003; Peng *et al.*, 2003), the evolutionary algorithms (Whitehead & Choate, 1994; Whitehead, 1996; Gonzalez *et al.*, 2003) or the expectation-maximisation algorithm (Yang & Chen, 1998; Mak & Kung, 2000). Generally, learning based on such a nonlinear approach is computationally expensive and may encounter the problem of local minima. Additionally, the network structure or the number of RBF nodes has to be determined via other means, typically based on cross validation. Alternatively, clustering algorithms can be applied to find the RBF centre vectors as well as the associated basis function variances (Moody & Darken, 1989; Chen *et al.*, 1992; Chen, 1995; Uykan, 2003). This leaves the RBF weights to be determined by the usual linear least squares solution. Again, the number of the clusters has to be determined via other means, such as cross validation. One of the most popular approaches for constructing RBF networks however is to formulate the problem as a linear learning one by considering the training input data points as candidate RBF centres and employing a common variance for every RBF node. A parsimonious RBF network is then identified using the orthogonal least squares (OLS) algorithm (Chen *et al.*, 1989, 1991, 1999, 2003, 2004a).

Many practical signal processing applications deal with complex-valued signals and data. This motivates the research in complex-valued ANNs, which have found wide-ranging applications in complex-valued signal processing problems (Uncini *et al.*, 1999; Kim & Adali, 2003; Li *et al.*, 2005; Yang & Bose, 2005; Hirose, 2006). A particular complex-

valued ANN architecture proposed by Chen *et al.* (1994) is the complex-valued RBF network, which takes the following form

$$y(k) = \sum_{i=1}^{N_c} w_i \varphi(\mathbf{x}(k); \mathbf{c}_i, \rho^2), \quad (1)$$

where  $y(k) \in \mathcal{C}$  ( $\mathcal{C}$  being the field of complex-valued numbers) and  $\mathbf{x}(k) \in \mathcal{C}^L$  denote the complex-valued RBF network output and input vector, respectively,  $N_c$  denotes the number of RBF units,  $w_i$  are the complex-valued RBF weights,  $\mathbf{c}_i \in \mathcal{C}^L$  are the complex-valued RBF centres,  $\rho^2$  is the positive RBF variance, and  $\varphi(\bullet)$  is the real-valued radial basis function. When  $y(k)$  and  $w_i$  are real-valued, this complex-valued RBF network reduces to the special case of the usual real-valued RBF network (Moody & Darken, 1989; Chen *et al.*, 1991). Each RBF unit in the complex-valued RBF network (1) can be interpreted as some underlying probability density function (Chen *et al.*, 1994). Such a physical interpretation makes this complex-valued RBF network a powerful tool in processing complex-valued signals, particularly, in applications to communication channel equalisation and signal detection (Chen *et al.*, 1994a; Cha & Kassam, 1995; Gan *et al.*, 1999; Deng *et al.*, 2002; Botoca & Budura, 2006).

Like many existing neural network models, this complex-valued RBF network, however, constitutes a black-box approach that seeks a sparse model representation extracted from the training data. Adopting black-box modelling is appropriate, if no *a priori* information exists regarding the underlying data generating mechanism. However, if there exists *a priori* information concerning the system to be modelled, it should be incorporated in the modelling process. The use of available prior knowledge in data modelling often leads to an improved performance. For real-valued signal processing applications, it has been recognised that many real-life phenomena exhibit inherent symmetry and these properties are hard to infer accurately from noisy data with the aid of black-box real-valued neural network models. However, by imposing appropriate symmetry on the model's structure, exploiting the symmetry properties becomes easier and this leads to substantial improvements in the achievable modelling performance. For example, in regression-type applications, how to exploit odd or even symmetry of the underlying system explicitly in both the real-valued RBF network and least squares support vector machine has been demonstrated (Aguirre *et al.*, 2004; Espinoza *et al.*, 2005), while in two-class classification-type applications, a novel real-valued symmetric RBF (SRBF) network has been proposed for communication signal detection (Chen *et al.*, 2006, 2007,a,b), which explicitly utilises odd symmetry of the underlying optimal Bayesian detection solution.

This contribution continues this theme and extend the grey-box approach to complex-valued RBF modelling. Instead of simple odd or even symmetry typically found in real-valued signal processing problems, symmetry properties inherited in many complex-valued signal processing problems are more complicated, and this is demonstrated using the application to nonlinear beamforming assisted detection for multiple-antenna aided wireless systems that employ complex-valued quadrature phase shift keying (QPSK) modulation scheme. This naturally leads to our proposed complex-valued SRBF network. Two training

algorithms for this complex-valued SRBF network are proposed. The first method is based on a modified version of the cluster-variation enhanced clustering algorithm (Chinrungrueng & Séquin, 1995; Chen, 1995; Chen *et al.*, 2007a). The second method is derived by modifying the orthogonal-forward-selection (OFS) procedure based on the Fisher ratio of class separability measure (FRCSM) (Mao, 2002; Chen *et al.*, 2004, 2007b) through changing the two-class FRCSM into the multi-class (four-class) one. The effectiveness of the proposed complex-valued SRBF network and the efficiency of the two training algorithms are demonstrated in nonlinear beamforming application. Although the proposed complex-valued SRBF network is derived in the context of nonlinear beamforming for QPSK wireless systems, it is applicable to the generic complex-valued signal processing problem that exhibits a similar symmetric behaviour.

The remainder of this contribution is organised as follows. We first present the complex-valued signal model for the multiple-antenna aided wireless system that employs the QPSK signalling as well as the optimal Bayesian nonlinear beamforming or detection solution. The inherent symmetric structure of this Bayesian nonlinear beamforming solution is then highlighted. This naturally leads to the proposed complex-valued SRBF network, which can easily be constructed using a cluster-variation enhanced clustering algorithm. While the clustering-based SRBF network developed is a direct ‘‘carbon-copy’’ of the symmetric Bayesian beamformer, a more generic complex-valued SRBF is also proposed and a multi-class FRCSM-based OFS algorithm is derived to construct a parsimonious SRBF network model. Finally, some concluding remarks are offered.

## BACKGROUND

A coherent wireless communication system supports  $M$  users, where each user transmits on the same angular carrier frequency  $\omega$  with a single transmit antenna while the receiver is equipped with a linear antenna array consisting of  $L$  uniformly spaced elements in order to achieve user separation in the angular domain (Paulraj *et al.*, 2003; Tse & Viswanath, 2005). Assume furthermore that the channel is non-dispersive and hence it does not induce intersymbol interference.

### Beamforming Signal Model

The symbol-rate complex-valued received signal vector  $\mathbf{x}(k) = [x_1(k) \ x_2(k) \ \cdots \ x_L(k)]^T$  can be expressed as (Litva & Lo, 1996; Błogh & Hanzo, 2002)

$$\mathbf{x}(k) = \mathbf{P}\mathbf{b}(k) + \mathbf{n}(k) = \bar{\mathbf{x}}(k) + \mathbf{n}(k), \quad (2)$$

where  $\bar{\mathbf{x}}(k)$  denotes the noise-free part of the received signal vector,  $\mathbf{n}(k) = [n_1(k) \ n_2(k) \ \cdots \ n_L(k)]^T$  and  $n_l(k)$  denotes the complex-valued white Gaussian noise associated with the  $l$ -th channel having  $E[|n_l(k)|^2] = 2\sigma_n^2$ , while  $\mathbf{P}$  is the  $L \times M$  complex-valued system’s channel matrix and  $\mathbf{b}(k) = [b_1(k) \ b_2(k) \ \cdots \ b_M(k)]^T$  with  $b_i(k)$  denoting

the  $k$ -th transmitted symbol of user  $i$ . The system's channel matrix  $\mathbf{P}$  is defined by

$$\mathbf{P} = [A_1 \mathbf{s}_1 \ A_2 \mathbf{s}_2 \ \cdots \ A_M \mathbf{s}_M], \quad (3)$$

where  $A_i$  denotes the  $i$ -th complex-valued non-dispersive channel tap and the steering vector of user  $i$  is given by

$$\mathbf{s}_i = [e^{j\omega t_1(\theta_i)} \ e^{j\omega t_2(\theta_i)} \ \cdots \ e^{j\omega t_L(\theta_i)}]^T \quad (4)$$

with  $\theta_i$  and  $t_l(\theta_i)$  denoting the angle of arrival and the relative time delay at array element  $l$  for user  $i$ , respectively. The modulation scheme is assumed to be the QPSK and, therefore,  $b_i(k)$  takes values from the QPSK symbol set

$$\mathcal{B}_{\text{QPSK}} \triangleq \{b^{[1]} = +1 + j, \ b^{[2]} = -1 + j, \ b^{[3]} = -1 - j, \ b^{[4]} = +1 - j\}. \quad (5)$$

Let source  $i$  be the desired user and the rest of the sources be the interfering users. The average signal-to-noise ratio (SNR) of the system is given by

$$\text{SNR} = \left( \frac{1}{M} \sum_{i=1}^M |A_i|^2 \right) \sigma_b^2 / 2\sigma_n^2, \quad (6)$$

where  $\sigma_b^2$  is the QPSK symbol energy, and the desired signal-to-interferer  $q$  ratio (SIR) is defined by  $\text{SIR}_{i,q} = |A_i|^2 / |A_q|^2$ , for  $q \neq i$ .

Traditionally, a linear beamformer is adopted to detect the desired user's signal (Litva & Lo, 1996; Blogh & Hanzo, 2002). The linear beamformer for user  $i$  is defined by  $y_{\text{Lin}}(k) = \boldsymbol{\alpha}_i^H \mathbf{x}(k)$ , where  $\boldsymbol{\alpha}_i = [\alpha_{1,i} \ \alpha_{2,i} \ \cdots \ \alpha_{L,i}]^T$  is the complex-valued  $i$ -th linear beamformer's weight vector. The decision regarding the transmitted symbol  $b_i(k)$  is given by  $\hat{b}_i(k) = \text{sgn}(y_{\text{Lin}}(k))$  with

$$\text{sgn}(y) = \begin{cases} b^{[1]} = +1 + j, & y_R \geq 0 \text{ and } y_I \geq 0, \\ b^{[2]} = -1 + j, & y_R < 0 \text{ and } y_I \geq 0, \\ b^{[3]} = -1 - j, & y_R < 0 \text{ and } y_I < 0, \\ b^{[4]} = +1 - j, & y_R \geq 0 \text{ and } y_I < 0, \end{cases} \quad (7)$$

where  $y_R = \Re[y]$  and  $y_I = \Im[y]$  denote the real and imaginary parts of  $y$ , respectively. The optimal weight vector designed for the linear beamformer is known to be the minimum bit error rate (L-MBER) solution (Chen *et al.*, 2005), which directly minimises the bit error rate (BER) of the linear beamformer. However, if one is willing to extend the concept of beamforming from linear to nonlinear, significant performance improvement can be achieved, at the cost of considerably increased complexity (Chen *et al.*, 2008).

### Optimal Bayesian Beamforming Solution

Denote the  $N_b = 4^M$  legitimate combinations of  $\mathbf{b}(k)$  as  $\mathbf{b}_q$ ,  $1 \leq q \leq N_b$ . The noiseless channel output  $\bar{\mathbf{x}}(k)$  takes values from the vector state set

$$\bar{\mathbf{x}}(k) \in \mathcal{X} \triangleq \{\bar{\mathbf{x}}_q = \mathbf{P}\mathbf{b}_q, 1 \leq q \leq N_b\}. \quad (8)$$

The signal state set  $\mathcal{X}$  can be divided into the four subsets conditioned on the value of  $b_i(k)$  as follows

$$\mathcal{X}^{[m]} \triangleq \{\bar{\mathbf{x}}_q^{[m]} \in \mathcal{X}, 1 \leq q \leq N_{sb} : b_i(k) = b^{[m]}\}, \quad (9)$$

for  $1 \leq m \leq 4$ , where the size of  $\mathcal{X}^{[m]}$  is  $N_{sb} = 4^{M-1}$ . Denote the conditional probabilities of receiving  $\mathbf{x}(k)$  given  $b_i(k) = b^{[m]}$  as  $p^{[m]}(\mathbf{x}(k)) = p(\mathbf{x}(k)|b_i(k) = b^{[m]})$ . According to Bayes' decision theory (Duda & Hart, 1973), the optimal detection strategy is

$$\hat{b}_i(k) = b^{[m^*]} \quad (10)$$

where

$$m^* = \arg \max_{1 \leq m \leq 4} p^{[m]}(\mathbf{x}(k)). \quad (11)$$

If we introduce the following complex-valued Bayesian decision variable (Chen *et al.*, 1994a)

$$y_{\text{Bay}}(k) \triangleq \sum_{m=1}^4 b^{[m]} \cdot p^{[m]}(\mathbf{x}(k)), \quad (12)$$

the optimal Bayesian detection rule (10) and (11) is equivalent to  $\hat{b}_i(k) = \text{sgn}(y_{\text{Bay}}(k))$ .

The conditional probability  $p^{[m]}(\mathbf{x}(k))$ ,  $1 \leq m \leq 4$ , can be expressed as

$$p^{[m]}(\mathbf{x}(k)) = \sum_{q=1}^{N_{sb}} \beta_q e^{-\frac{\|\mathbf{x}(k) - \bar{\mathbf{x}}_q^{[m]}\|^2}{2\sigma_n^2}}, \quad (13)$$

where  $\bar{\mathbf{x}}_q^{[m]} \in \mathcal{X}^{[m]}$ , and  $\beta_q$  is proportional to the *a priori* probability of  $\bar{\mathbf{x}}_q^{[m]}$ . Since all the  $\bar{\mathbf{x}}_q^{[m]}$  are equiprobable,  $\beta_q = \beta = \frac{1}{N_{sb}(2\pi\sigma_n^2)^L}$ . It can be seen from (13) that the optimal Bayesian decision variable (12) takes the structure of a complex-valued RBF network (Chen *et al.*, 1994a) with a Gaussian RBF function. In fact, substituting (13) into (12) provides a physical motivation for Chen *et al.* (1994) to derive the complex-valued RBF network (1).

The state subsets  $\mathcal{X}^{[m]}$ ,  $1 \leq m \leq 4$ , are distributed symmetrically with respect to each other as summarised in the following lemma.

**Lemma 1** The four subsets  $\mathcal{X}^{[m]}$ ,  $1 \leq m \leq 4$ , satisfy

$$\mathcal{X}^{[2]} = +j \cdot \mathcal{X}^{[1]}, \quad \mathcal{X}^{[3]} = -1 \cdot \mathcal{X}^{[1]}, \quad \mathcal{X}^{[4]} = -j \cdot \mathcal{X}^{[1]}. \quad (14)$$

*Proof:* Consider any  $\bar{\mathbf{x}}_q^{[1]} = \mathbf{P}\mathbf{b}_q^{[1]} \in \mathcal{X}^{[1]}$ , where the  $i$ -th element of  $\mathbf{b}_q^{[1]}$  is  $b^{[1]} = +1 + j$ . Noting  $+j \cdot b^{[1]} = b^{[2]}$ ,

$$+j \cdot \bar{\mathbf{x}}_q^{[1]} = \mathbf{P} \left( +j \cdot \mathbf{b}_q^{[1]} \right) \in \mathcal{X}^{[2]}. \quad (15)$$

This proves the first relationship in (14). The proofs of the other two relationships are similar. Given this symmetry, the optimal Bayesian solution (12) can alternatively be expressed as

$$y_{\text{Bay}}(k) = \sum_{q=1}^{N_{sb}} \left\{ b^{[1]} \beta \cdot e^{-\frac{\|\mathbf{x}(k) - \bar{\mathbf{x}}_q^{[1]}\|^2}{2\sigma_n^2}} + b^{[2]} \beta \cdot e^{-\frac{\|\mathbf{x}(k) - j \cdot \bar{\mathbf{x}}_q^{[1]}\|^2}{2\sigma_n^2}} \right. \\ \left. + b^{[3]} \beta \cdot e^{-\frac{\|\mathbf{x}(k) + \bar{\mathbf{x}}_q^{[1]}\|^2}{2\sigma_n^2}} + b^{[4]} \beta \cdot e^{-\frac{\|\mathbf{x}(k) + j \cdot \bar{\mathbf{x}}_q^{[1]}\|^2}{2\sigma_n^2}} \right\}, \quad (16)$$

where  $\bar{\mathbf{x}}_q^{[1]} \in \mathcal{X}^{[1]}$ . Note that the symmetric property of the Bayesian detection solution for the QPSK communication system is more complex than the simple odd symmetry of the Bayesian detection solution for the binary communication system derived in (Chen *et al.*, 2006, 2007,a,b). Even more complicated symmetric property exists for the generic higher-order quadrature amplitude modulation (QAM) case (Chen *et al.*, 2006a). Extension to the higher-order QAM case is beyond the scope of this contribution.

## CLUSTERING-BASED SYMMETRIC RBF BEAMFORMING

Consider the problem of realising the optimal Bayesian beamforming solution using a complex-valued RBF network. The symmetry of the Bayesian solution (16) should be explicitly exploited, and one way to guarantee this desired symmetry is to employ the following SRBF network for the detection of user  $i$  data

$$y_{\text{SRBF}}(k) = \sum_{q=1}^{N_c} \beta_q \phi(\mathbf{x}(k); \mathbf{c}_q, \rho^2), \quad (17)$$

with the decision  $\hat{b}_i(k) = \text{sgn}(y_{\text{SRBF}}(k))$ . Unlike the complex-valued RBF network (1), here the RBF weights  $\beta_q$  are real-valued and the RBF nodes' response  $\phi(\mathbf{x}(k); \mathbf{c}_q, \rho^2)$  are complex-valued, defined as

$$\phi(\mathbf{x}; \mathbf{c}, \rho^2) = b^{[1]} \varphi(\mathbf{x}; \mathbf{c}, \rho^2) + b^{[2]} \varphi(\mathbf{x}; j\mathbf{c}, \rho^2) + b^{[3]} \varphi(\mathbf{x}; -\mathbf{c}, \rho^2) + b^{[4]} \varphi(\mathbf{x}; -j\mathbf{c}, \rho^2), \quad (18)$$

where  $\varphi(\bullet; \bullet)$  is the usual real-valued radial basis function. In this study, the Gaussian function of the form

$$\varphi(\mathbf{x}; \mathbf{c}, \rho^2) = e^{-\frac{\|\mathbf{x} - \mathbf{c}\|^2}{2\rho^2}}. \quad (19)$$

is used. Note that the standard complex-valued RBF network (1) does not guarantee to possess the same symmetry property of the optimal Bayesian solution (16), particularly when

the RBF centres  $\mathbf{c}_q$  are obtained directly from the channel-impaired observation data. By contrast, the SRBF network (17) with the symmetric node structure (18) guarantees to have the same symmetry property of the optimal Bayesian solution (16).

### Clustering-Based Training Algorithm

Given a set of the  $K$  training data  $D_K = \{\mathbf{x}(k), d(k)\}_{k=1}^K$ , where  $d(k) = b_i(k) \in \mathcal{B}_{\text{QPSK}}$ , the task is to construct this SRBF network from the training data set  $D_K$ . Since the number of users is usually known, the number of RBF centres can be set to  $N_c = N_{sb}$ . To further exploit the structure of the optimal Bayesian solution (16), all the real-valued RBF weights can be set to a positive constant  $\beta_q = \beta > 0$ . Specific value of  $\beta$  has no influence to the detection performance. Furthermore, set the RBF variance to  $\rho^2 = \hat{\sigma}_n^2$ , where  $\hat{\sigma}_n^2$  is an estimate of the noise variance. It is worth emphasising that the performance of the SRBF network is not sensitive to the value of  $\rho^2$  used and there exist a wide-range values of  $\rho^2$  which enable the SRBF network to approach the optimal Bayesian performance. This will further be demonstrated in the following simulation study. The insensitiveness of the SRBF network to the value of  $\rho^2$  used is a direct consequence of the insensitiveness of the Bayesian detection solution to the value of the noise variance used. Thus, adaptation of the SRBF network (17) with the node structure (18) becomes the task of finding appropriately the RBF centre vectors  $\mathbf{c}_q$ .

To use the channel-impaired training data  $D_K$  to directly obtain the RBF centre vectors for the SRBF network (17) and hence to approximate the optimal Bayesian solution, we propose to use a modified version of the cluster-variation assisted clustering algorithm (Chinrungrueng & Séquin, 1995; Chen, 1995). Specifically, during training, the RBF centres are adjusted according to

$$\mathbf{c}_l(k) = \mathbf{c}_l(k-1) + \mu_c \mathcal{M}_l(\tilde{\mathbf{x}}(k))(\tilde{\mathbf{x}}(k) - \mathbf{c}_l(k-1)), \quad (20)$$

where

$$\tilde{\mathbf{x}}(k) = \begin{cases} +1 \cdot \mathbf{x}(k), & b_i(k) = b^{[1]}, \\ -j \cdot \mathbf{x}(k), & b_i(k) = b^{[2]}, \\ -1 \cdot \mathbf{x}(k), & b_i(k) = b^{[3]}, \\ +j \cdot \mathbf{x}(k), & b_i(k) = b^{[4]}, \end{cases} \quad (21)$$

$\mu_c$  is the step size and the membership function  $\mathcal{M}_l(\mathbf{x})$  is defined as

$$\mathcal{M}_l(\mathbf{x}) = \begin{cases} 1, & \text{if } \bar{v}_l \|\mathbf{x} - \mathbf{c}_l\|^2 \leq \bar{v}_q \|\mathbf{x} - \mathbf{c}_q\|^2, \forall q \neq l, \\ 0, & \text{otherwise,} \end{cases} \quad (22)$$

with  $\bar{v}_l$  being the variation of the  $l$ -th cluster. In order to estimate the associated variation  $\bar{v}_l$ , the following updating rule (Chinrungrueng & Séquin, 1995; Chen, 1995) is used

$$\bar{v}_l(k) = \mu_v \bar{v}_l(k-1) + (1 - \mu_v) \mathcal{M}_l(\tilde{\mathbf{x}}(k)) \|\tilde{\mathbf{x}}(k) - \mathbf{c}_l(k-1)\|^2, \quad (23)$$

where  $\mu_v$  is a constant slightly less than 1.0. The initial variations  $\bar{v}_l(0)$ ,  $\forall l$ , are set to the same small positive number.



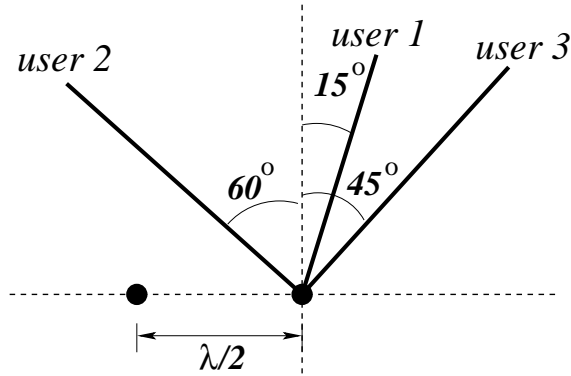


Figure 1: Angular locations of the three QPSK users with respect to the two-element linear antenna array having  $\lambda/2$  spacing, where  $\lambda$  is the wavelength.

It is known that this cluster-variation enhanced clustering algorithm is capable of obtaining the optimal cluster partitioning structure and all the cluster variations converge to the same value (Chinrungrueng & Séquin, 1995). Specifically, in our particular application, the RBF centre vectors  $\mathbf{c}_q$  converge to the noise-free signal states  $\bar{\mathbf{x}}_q^{[1]}$  and all the cluster variations converge to the noise variance  $2\sigma_n^2$ .

### Simulation Study

A simulation study was carried out to investigate performance of the proposed clustering-based SRBF network in nonlinear beamforming application.

**Example One.** The system employed a two-element antenna array to support three QPSK users. Fig. 1 depicts the angular positions of the three users with respect to the antenna array. The simulated narrowband channels were  $A_i = 1 + j0$ ,  $1 \leq i \leq 3$ , and all the three users had an equal power. Thus all the SIRs were 0 dB. First, we demonstrated the performance improvement achievable by the optimal nonlinear beamforming over the optimal linear one, when the system's channel matrix  $\mathbf{P}$  and the noise statistics  $\sigma_n^2$  were known. Fig. 2 compares the BER performance of the Bayesian beamforming and the L-MBER beamforming. As expected, the Bayesian beamforming achieved much better BER performance over the optimal linear beamforming. This performance gain was of course obtained at the cost of an increased complexity. From Fig. 2, it can be seen that the performance of the individual linear beamformer depended on the particular user's angular position as well as the other users' locations. For user 3, the underlying system was not linearly separable and hence the L-MBER beamformer exhibits a high error floor. By contrast, all the three Bayesian beamformers had the similar performance, as a nonlinear beamformer can operate successfully even in the linearly nonseparable scenario. Because of this remarkable robustness property, we only concentrated on the user one when investigating the SRBF beamforming.

The clustering-based SRBF beamforming for user one was then studied. For this example, the number of the subset channel states was  $N_{sb} = 16$ , and we used the first 16 data

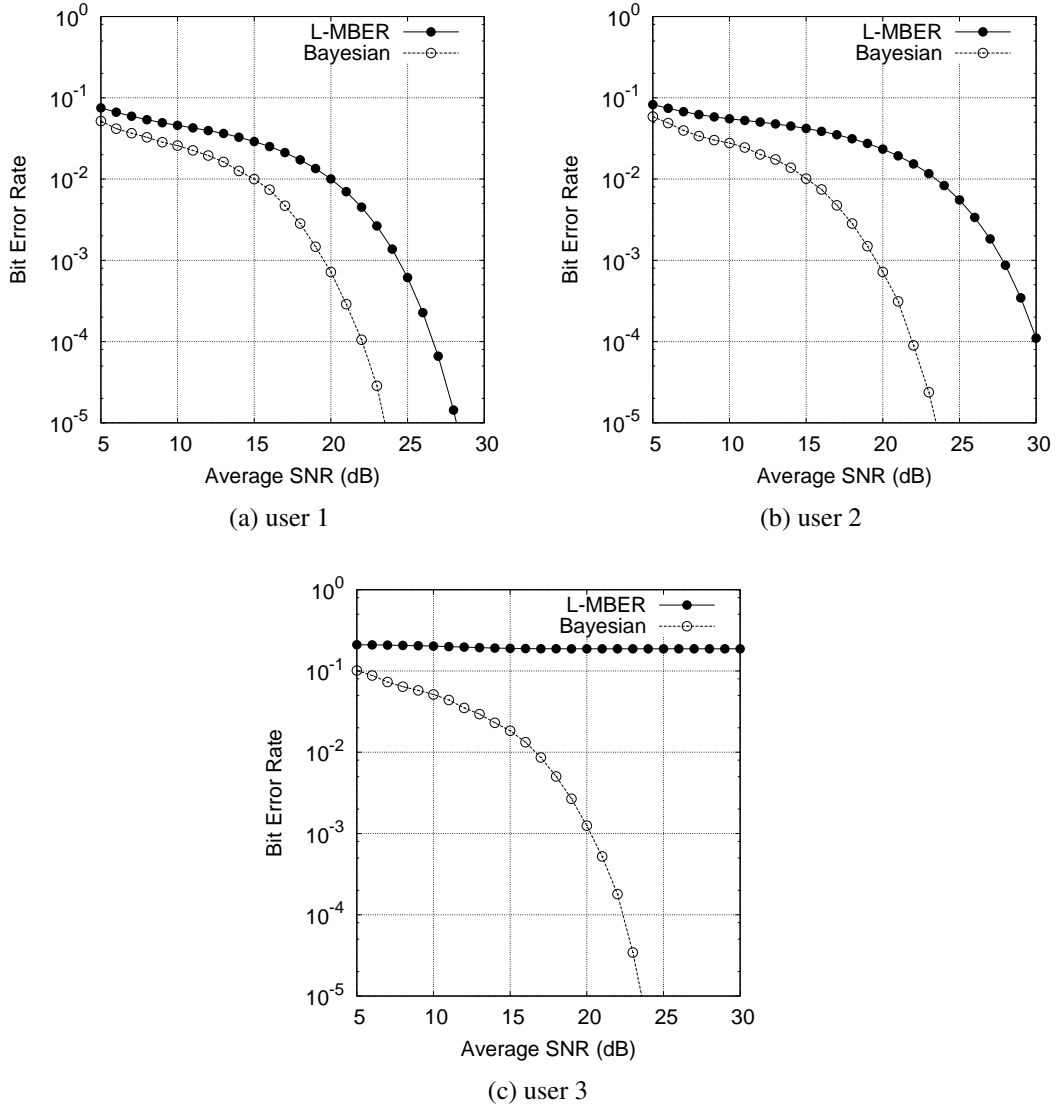


Figure 2: BER performance comparison of the optimal nonlinear beamforming with the optimal linear beamforming, for the two-element array system supporting three QPSK users.

points  $\tilde{\mathbf{x}}(k)$ ,  $1 \leq k \leq 16$ , as the initial RBF centres. The initial cluster variations were set to  $\bar{v}_l(0) = 0.02$  for  $1 \leq l \leq N_{sb}$ , and the adaptive constant for updating the cluster variations was chosen to be  $\mu_v = 0.995$ . Note that the general rule is that all the initial cluster variations  $\bar{v}_l(0)$  should be set to the same small positive number and  $\mu_v$  should be set to a constant slightly less than 1.0. Convergence performance of the cluster-variation enhanced clustering algorithm was assessed in the simulation based on the Euclidean distance between the set of the RBF centres  $\{c_l\}_{l=1}^{N_{sb}}$  and the set of the true subset channel

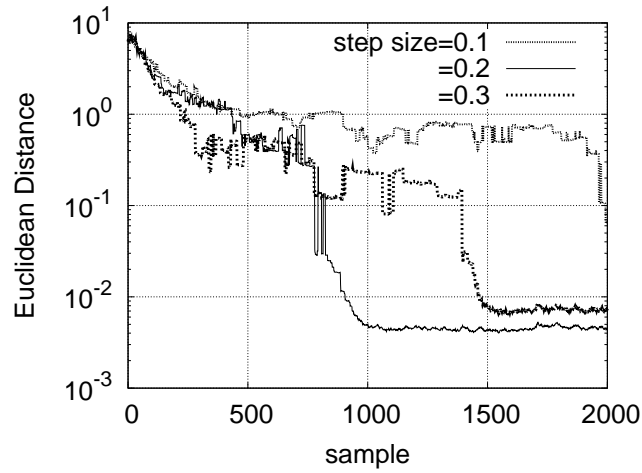


Figure 3: Learning curves of the cluster algorithm for user one of the two-element array system supporting three QPSK users, in terms of Euclidean distance between the RBF centres and true channel states averaged over ten runs, given SNR= 20 dB.

states  $\{\bar{\mathbf{x}}_l^{[1]}\}_{l=1}^{N_{sb}}$  defined as

$$ED(k) = \frac{1}{N_{sb}} \sum_{l=1}^{N_{sb}} \|\mathbf{c}_l(k) - \bar{\mathbf{x}}_l^{[1]}\|^2. \quad (24)$$

Given SNR= 20 dB, Fig. 3 plots the learning curves of the clustering algorithm in terms

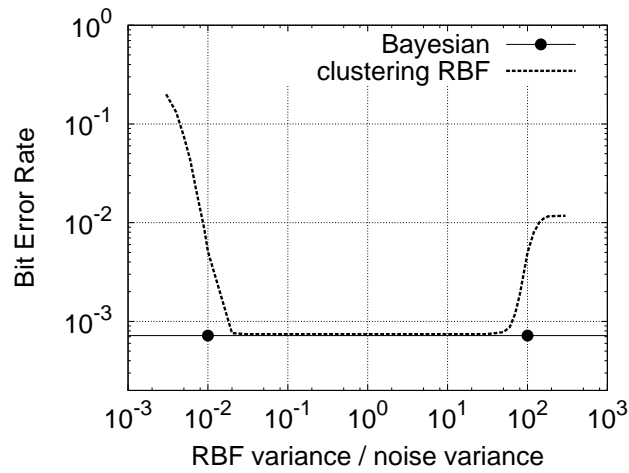


Figure 4: Influence of the RBF variance on the BER performance of the clustering-based SRBF beamformer for user one of the two-element array system supporting three QPSK users, given SNR= 20 dB.

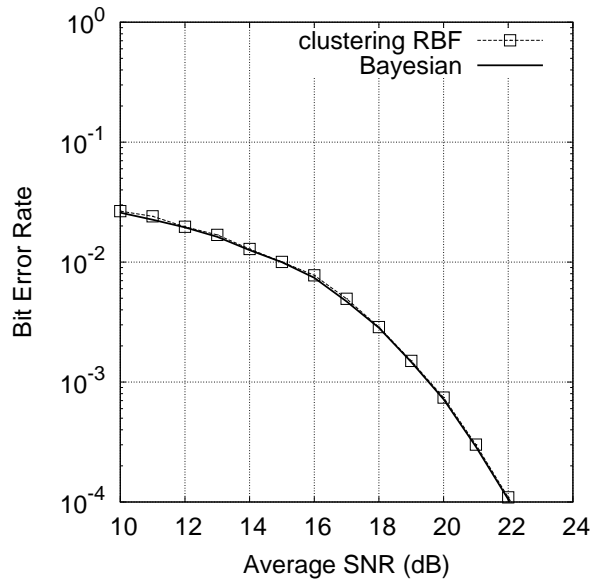


Figure 5: User-one BER performance of the clustering-based SRBF beamformer for the two-element array system supporting three QPSK users, given  $\rho^2 = \sigma_n^2$ , in comparison with the Bayesian beamforming performance.

of the Euclidean distance (24) averaged over 10 different random runs for the three values of the adaptive gain  $\mu_c$ . It is seen from Fig. 3 that for this example the best convergence performance was achieved with  $\mu_c = 0.2$ . The robustness of the clustering-based adaptive SRBF beamforming with respect to the value of the RBF variance  $\rho^2$  used is demonstrated in Fig. 4, where it can be seen that there exist a wide-ranging values of the RBF variance  $\rho^2$  for the clustering-based SRBF network to achieve the Bayesian beamforming performance. Fig. 5 compares the BER performance of the clustering-based adaptive SRBF beamformer for user one after convergence with that of the Bayesian beamformer, given the RBF variance  $\rho^2 = \sigma_n^2$ .

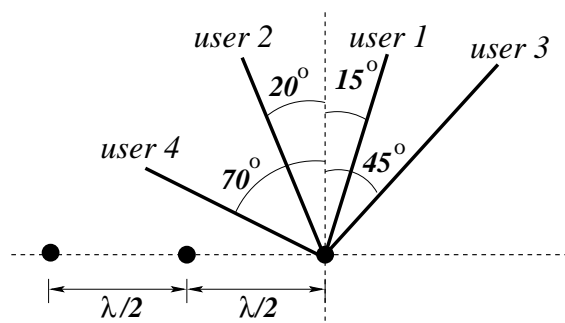


Figure 6: Angular locations of the four QPSK users with respect to the three-element linear antenna array having  $\lambda/2$  spacing, where  $\lambda$  is the wavelength.

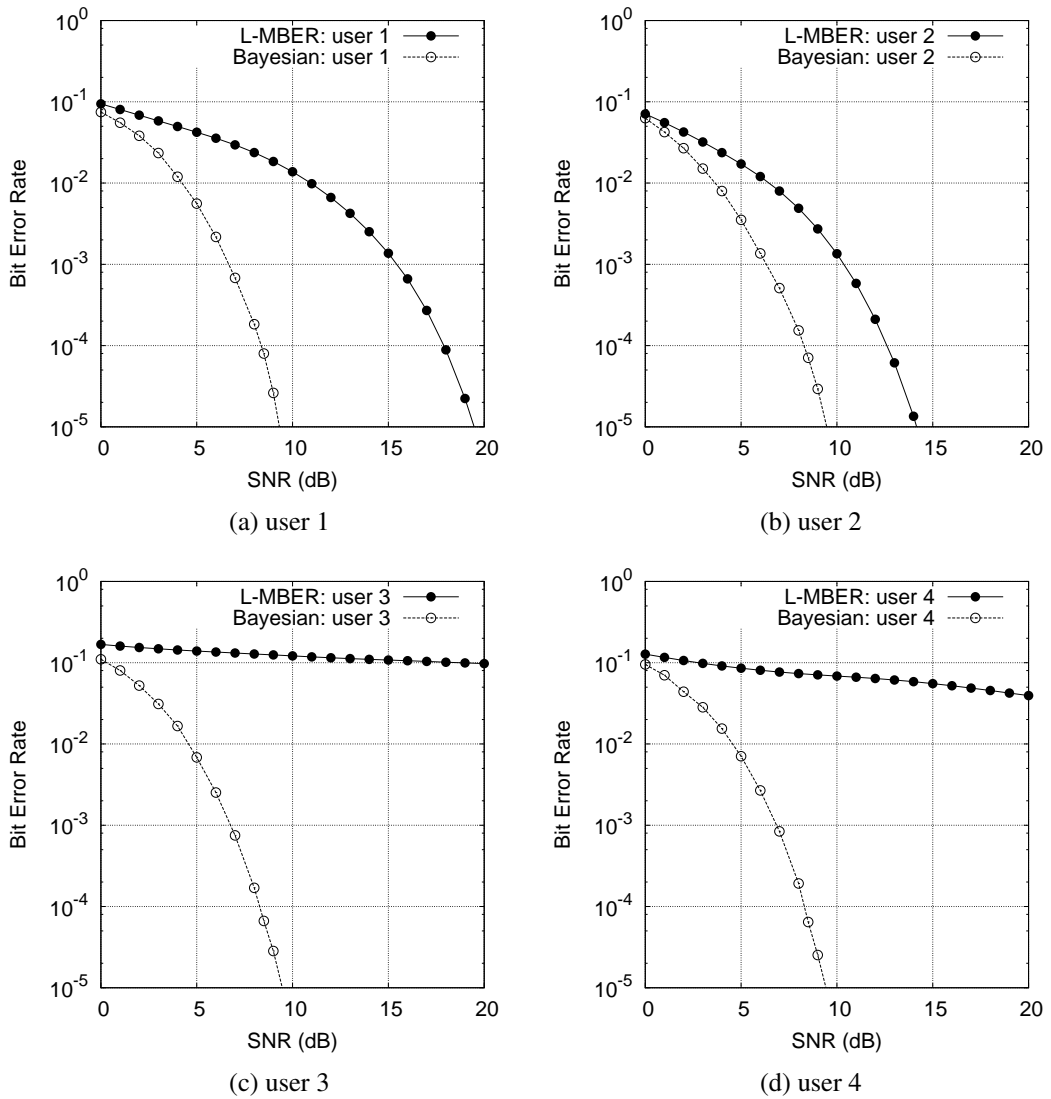


Figure 7: BER performance comparison of the optimal nonlinear beamforming with the optimal linear beamforming, for the three-element array system supporting four QPSK users.

**Example Two.** A three-element antenna array was designed to support four QPSK users. Fig. 6 shows the angular positions of the four users with respect to the antenna array. The simulated narrowband channels were  $A_i = 1 + j0$  for  $1 \leq i \leq 4$ , and all the four users had the same power. Fig. 7 demonstrates the BER performance improvement achievable by the Bayesian beamforming over the L-MBER beamforming. Again, the robustness of the Bayesian beamforming is clearly shown in Fig. 7, where it can be seen that all the four Bayesian beamformers had the similar performance.

The clustering-based SRBF beamforming for user one was investigated. Note that the

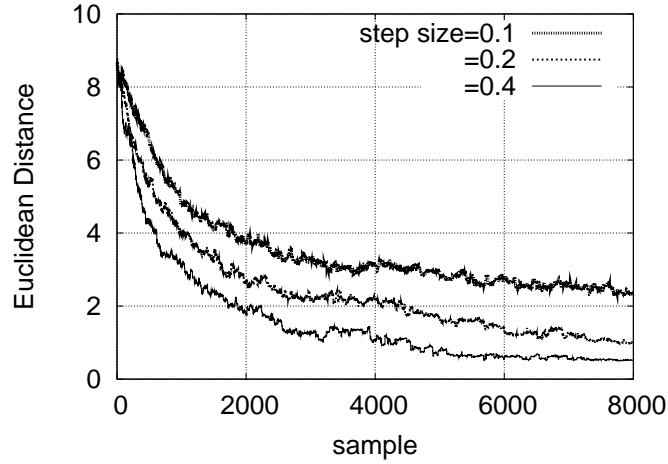


Figure 8: Learning curves of the cluster algorithm for user one of the three-element array system supporting four QPSK users, in terms of Euclidean distance between the RBF centres and true channel states averaged over ten runs, given SNR= 7 dB.

number of the subset channel states in this case was  $N_{sb} = 64$  and, therefore, the first 64 data points  $\check{\mathbf{x}}(k)$ ,  $1 \leq k \leq 64$ , were used as the initial RBF centres. The initial cluster variations were all set to  $\bar{v}_l(0) = 0.1$  for  $1 \leq l \leq N_{sb}$ , and the adaptive constant for updating the cluster variations was set to  $\mu_v = 0.995$ . Fig. 8 depicts the learning curves of the clustering algorithm in terms of the Euclidean distance (24) averaged over 10 different random runs for the three values of the adaptive gain  $\mu_c$ , given SNR= 7 dB. It is seen

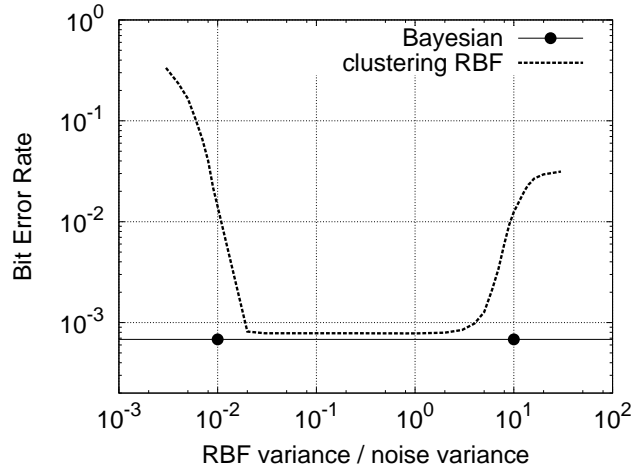


Figure 9: Influence of the RBF variance on the BER performance of the clustering-based SRBF beamformer for user one of the three-element array system supporting four QPSK users, given SNR= 7 dB.

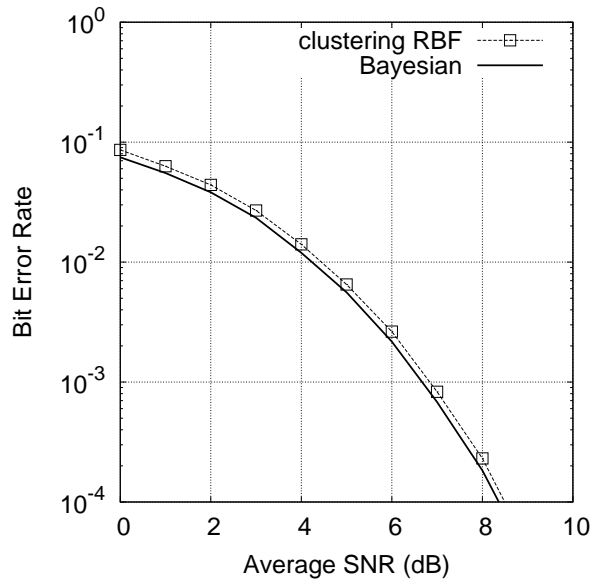


Figure 10: User-one BER performance of the clustering-based SRBF beamformer for the three-element array system supporting four QPSK users, given  $\rho^2 = \sigma_n^2$ , in comparison with the Bayesian beamforming performance.

that for this example the best convergence performance was achieved with  $\mu_c = 0.4$ . Given SNR = 7 dB, Fig. 9 illustrates the influence of the RBF variance  $\rho^2$  to the BER performance of the clustering-based SRBF network, where it was demonstrated again that there existed a wide-ranging values of  $\rho^2$  which enabled the clustering-based SRBF network to approach the Bayesian performance. Finally, Fig. 10 depicts the BER performance of the clustering-based SRBF beamformer for user one after convergence, given the RBF variance  $\rho^2 = \sigma_n^2$ , in comparison with the Bayesian benchmark.

## SYMMETRIC RBF BEAMFORMING CONSTRUCTION BASED ON OFS

The SRBF network (17) with the symmetric node structure (18) is a direct copy of the Bayesian detection solution (16). In general, however, the RBF weights can be complex-valued and we introduce the following generic complex-valued SRBF network

$$y_{\text{SRBF}}(k) = \sum_{q=1}^{N_c} w_q \phi(\mathbf{x}(k); \mathbf{c}_q, \rho^2), \quad (25)$$

where  $w_q$  are complex-valued RBF weights and the complex-valued RBF nodes' response  $\phi(\mathbf{x}(k); \mathbf{c}_q, \rho^2)$  are defined by (18). Thus, this more general complex-valued RBF network also guarantees to process the desired symmetric property. Given the training data set  $D_K = \{\mathbf{x}(k), d(k)\}_{k=1}^K$ , where  $d(k) = b_i(k) \in \mathcal{B}_{\text{QPSK}}$ , we now present an efficient algorithm for constructing this complex-valued SRBF network.

### OFS Based on Fisher Ratio of Class Separability Measure

Consider using every data points as candidate RBF centres, namely, setting  $N_c = K$  and  $\mathbf{c}_q = \mathbf{x}(q)$  for  $1 \leq q \leq K$ . Further assume that the value of the RBF variance  $\rho^2$  is specified. Let us define the modelling residual for  $\mathbf{x}(k) \in D_K$  as  $e(k) = d(k) - y_{\text{SRBF}}(k)$  and introduce the notation  $\phi_{k,q} = \phi(\mathbf{x}(k); \mathbf{c}_q, \rho^2)$ . Then the regression model over the data set  $D_K$  is expressed as

$$\mathbf{d} = \Phi \mathbf{w} + \mathbf{e}, \quad (26)$$

where  $\mathbf{d} = [d(1) d(2) \cdots d(K)]^T$ ,  $\mathbf{w} = [w_1 w_2 \cdots w_K]^T$ ,  $\mathbf{e} = [e(1) e(2) \cdots e(K)]^T$ , and the complex-valued regression matrix

$$\Phi = [\phi_1 \phi_2 \cdots \phi_K] \quad (27)$$

with columns  $\phi_q = [\phi_{1,q} \phi_{2,q} \cdots \phi_{K,q}]^T$ . Let an orthogonal decomposition of  $\Phi$  be  $\Phi = \mathbf{U}\mathbf{A}$ , where

$$\mathbf{A} = \begin{bmatrix} 1 & a_{1,2} & \cdots & a_{1,K} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \alpha_{K-1,K} \\ 0 & \cdots & 0 & 1 \end{bmatrix} \quad (28)$$

with complex-valued  $\alpha_{q,l}$ ,  $1 \leq q < l \leq K$ , and the complex-valued orthogonal matrix

$$\mathbf{U} = [\mathbf{u}_1 \mathbf{u}_2 \cdots \mathbf{u}_K] = \begin{bmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,K} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,K} \\ \vdots & \vdots & \cdots & \vdots \\ u_{K,1} & u_{K,2} & \cdots & u_{K,K} \end{bmatrix} \quad (29)$$

with columns satisfying  $\mathbf{u}_q^H \mathbf{u}_l = 0$ , if  $q \neq l$ . The regression model (26) can alternatively be expressed as

$$\mathbf{d} = \mathbf{U} \mathbf{g} + \mathbf{e}, \quad (30)$$

where the weight vector  $\mathbf{g} = [g_1 g_2 \cdots g_K]^T$  defined in the orthogonal model space satisfies the following triangular system  $\mathbf{A} \mathbf{w} = \mathbf{g}$ .

Recall that the output of the complex-valued SRBF network is used to provide an estimate for  $d(k)$  according to the decision rule  $\hat{d}(k) = \text{sgn}(y_{\text{SRBF}}(k))$ . Since  $d(k) \in \mathcal{B}_{\text{QPSK}}$ , this is a four-class classification problem. Let us first divide the training feature vectors  $\mathbf{X} = \{\mathbf{x}(k)\}_{k=1}^K$  into the four classes

$$\mathbf{X}^{[i]} \triangleq \{\mathbf{x}(k) \in \mathbf{X} : d(k) = b^{[i]}\}, \quad 1 \leq i \leq 4. \quad (31)$$

Assume that the number of samples in  $\mathbf{X}^{[i]}$  is  $K^{[i]}$ . Obviously

$$\sum_{i=1}^4 K^{[i]} = K. \quad (32)$$



Define furthermore the mean and variance of samples belonging to class  $\mathbf{X}^{[i]}$  in the direction of basis  $\mathbf{u}_l$  as  $m_{i,l}$  and  $\sigma_{i,l}^2$ , respectively, which are calculated according to

$$m_{i,l} = \frac{1}{K^{[i]}} \sum_{k=1}^K \delta(d(k) - b^{[i]}) u_{k,l} \quad (33)$$

and

$$\sigma_{i,l}^2 = \frac{1}{K^{[i]}} \sum_{k=1}^K \delta(d(k) - b^{[i]}) (u_{k,l} - m_{i,l})^2, \quad (34)$$

where the indicator function

$$\delta(x) = \begin{cases} 1, & x = 0 + j0, \\ 0, & x \neq 0 + j0. \end{cases} \quad (35)$$

Denote the Fisher ratio of the class separation between classes  $\mathbf{X}^{[i]}$  and  $\mathbf{X}^{[q]}$  in the direction of basis  $\mathbf{u}_l$  as  $F_{i,q,l}$ . Fisher ratio is defined as the ratio of the interclass difference to the intraclass spread (Duda & Hart, 1973)

$$F_{i,q,l} = \frac{(m_{i,l} - m_{q,l})^2}{(\sigma_{i,l}^2 + \sigma_{q,l}^2)}. \quad (36)$$

Fisher ratio provides a good class separability measure because its maximisation leads to the interclass difference being maximised and the intraclass spread being minimised.

Because the problem is a four-class classification one, we define the average Fisher ratio of the class separation in the direction of basis  $\mathbf{u}_l$  as

$$F_l = \frac{1}{6} \sum_{i=1}^3 \sum_{q=i+1}^4 F_{i,q,l}. \quad (37)$$

Based on this average Fisher ratio, significant RBF nodes or regressors can be selected in an OFS procedure, just as in the case of two-class problems (Mao, 2002; Chen *et al.*, 2004). Specifically, at the  $l$ -th stage of the OFS procedure, a regressor is chosen as the  $l$ -th term in the selected complex-valued SRBF network if it produces the largest  $F_l$  among the candidates terms,  $\mathbf{u}_q$ ,  $l \leq q \leq K$ . The procedure is terminated with a sparse  $n_c$ -node network when

$$\frac{F_{n_c}}{\sum_{l=1}^{n_c} F_l} \leq \xi, \quad (38)$$

where the threshold  $\xi$  determines the sparsity of the selected network model. The detailed OFS procedure based on the four-class FRCSM is summarised in Appendix. The least squares (LS) solution for the corresponding sparse model weight vector  $\mathbf{w}_{n_c} = [w_1 \ w_2 \ \cdots \ w_{n_c}]^T$  is readily available from  $\mathbf{A}_{n_c} \mathbf{w}_{n_c} = \mathbf{g}_{n_c}$ , given the LS solution of  $\mathbf{g}_{n_c}$ . In general, a desired value for the threshold  $\xi$  has to be determined via cross validation. However, in our particular application to nonlinear beamforming for multiple-antenna aided communication systems, we can simply set  $n_c = N_{sb}$ . Thus, in this particular application, we do not need to employ costly cross validation to determine the model size.

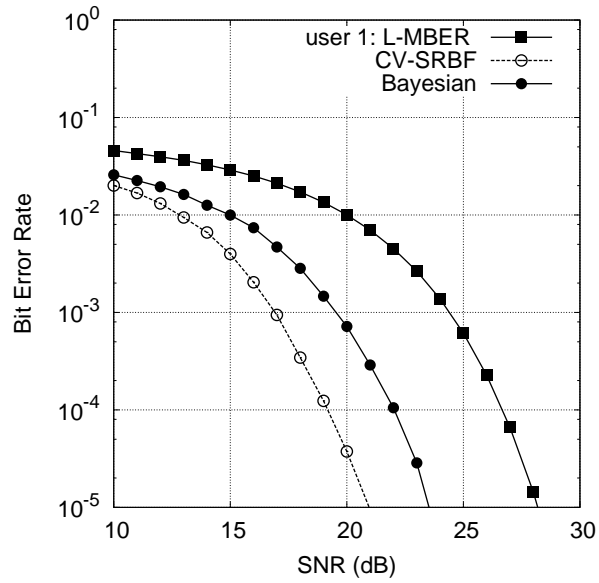


Figure 11: User-one BER performance of the OFS-based SRBF beamformer for the two-element array system supporting three QPSK users, in comparison with the Bayesian and theoretic L-MBER beamforming performance.

### Simulation Study

A simulation study involving the two same examples used in the previous section was performed to investigate the multi-class FRCSM-based OFS algorithm for constructing the

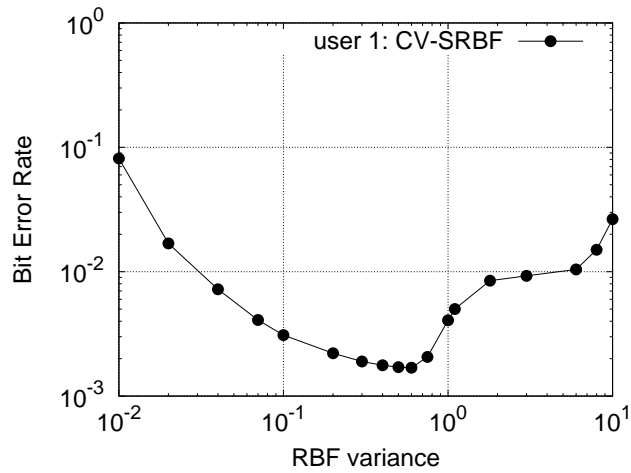


Figure 12: Influence of the RBF variance on the BER performance of the OFS-based SRBF beamformer for user one of the two-element array system supporting three QPSK users, given SNR= 16 dB.

complex-valued SRBF network (25).

**Example One.** This was the same two-element array system supporting three QPSK users with the users' angular positions shown in Fig. 1. Constructing the complex-valued SRBF network (25) for user one was first considered using the multi-class FRCSM-based OFS algorithm. Given each SNR value, a training set of  $K = 200$  samples was generated. For this example,  $N_{sb} = 16$  and, therefore, we stopped the selection procedure after choosing  $n_c = 16$  nodes. Unlike the clustering-based SRBF network, the RBF variance  $\rho^2$  has important influence on the BER performance. The value of the RBF variance  $\rho^2$ , therefore, was determined using cross validation, and appropriate values were found in the range of  $[0.2, 2.0]$  depending on the SNR value and noise realisation in the training data. The BER performance of the 16-term complex-valued SRBF beamformer obtained by the multi-class FRCSM-based OFS algorithm is plotted in Fig. 11, in comparison with the Bayesian and theoretic L-MBER benchmarks. It is surprising to see that the 16-term complex-valued SRBF network outperformed the Bayesian detector. A possible explanation is as follows. The Bayesian solution is derived under the assumption of white noise  $\mathbf{n}(k)$ . In the simulation, the noise was slightly coloured. Note that the weights of the SRBF network (25) are complex-valued. Therefore, a 16-term complex-valued SRBF network has a larger model size than the Bayesian solution (whose weights are real-valued). This doubled model size might have allowed the complex-valued SRBF network to exploit the noise statistics in the training data better. The influence of the RBF variance  $\rho^2$  to the performance of the OFS-based SRBF network is demonstrated in Fig. 12, given SNR= 16 dB.

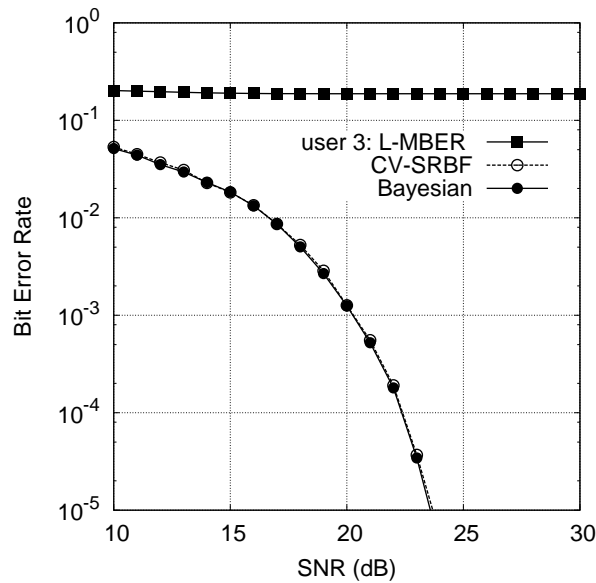


Figure 13: User-three BER performance of the OFS-based SRBF beamformer for the two-element array system supporting three QPSK users, in comparison with the Bayesian and theoretic L-MBER beamforming performance.

The beamforming for detecting the user-three data was also considered. For each SNR value, a training data set consisting of  $K = 200$  samples was used to construct a 16-term complex-valued SRBF network using the multi-class FRCSM-based OFS, and the BER performance of the resulting SRBF beamformer is depicted in Fig. 13, together with the Bayesian and theoretic L-MBER beamforming benchmarks. Again the value of the RBF variance was determined via cross validation, and appropriate values were found in the range of  $[1.6, 2.0]$ , depending on the SNR value. Detection of user-three data was a more difficult task than detection of user-one data as the former was a nonlinearly separable problem. It can be seen from Fig. 13 that the performance of the 16-term complex-valued SRBF network was indistinguishable from that of the Bayesian beamformer.

**Example Two.** The same three-element array system was used to support the same four QPSK users which had the users' angular positions as shown in Fig. 6. Detection of users one and four was considered. For user one, the underlying system was linearly separable, while for user four it was a more difficult nonlinearly separable problem. Given each SNR value, a training set of  $K = 600$  samples was generated to construct the complex-valued SRBF network (25) using the multi-class FRCSM-based OFS algorithm. Because  $N_{sb} = 64$ , we terminated the selection procedure after choosing  $n_c = 64$  nodes. The value of the RBF variance  $\rho^2$  was determined using cross validation.

For user one, appropriate values of  $\rho^2$  were found in the range of  $[0.6, 2.0]$ , depending on the SNR value, and the BER performance of the resulting 64-term complex-valued SRBF beamformer is depicted in Fig. 14, in comparison with the two benchmarks. The

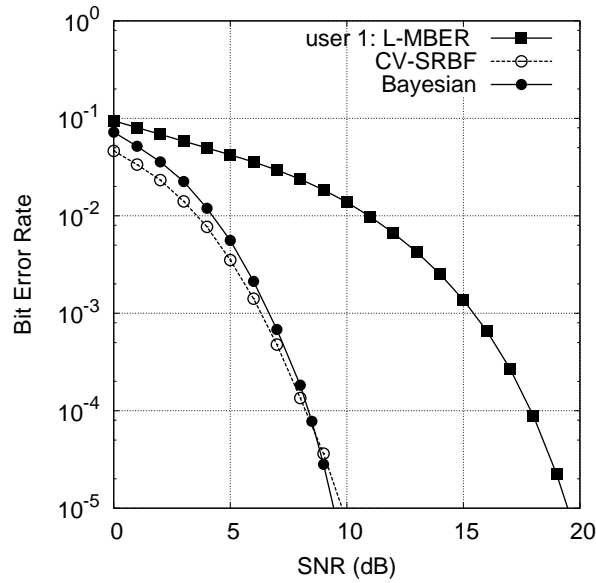


Figure 14: User-one BER performance of the OFS-based SRBF beamformer for the three-element array system supporting four QPSK users, in comparison with the Bayesian and theoretic L-MBER beamforming performance.

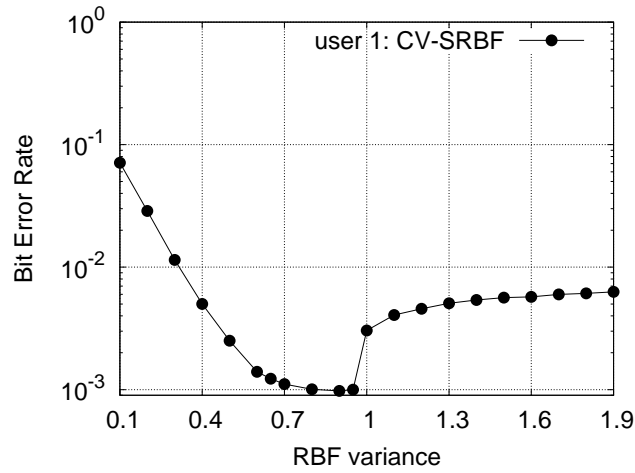


Figure 15: Influence of the RBF variance on the BER performance of the OFS-based SRBF beamformer for user one of the three-element array system supporting four QPSK users, given SNR= 6 dB.

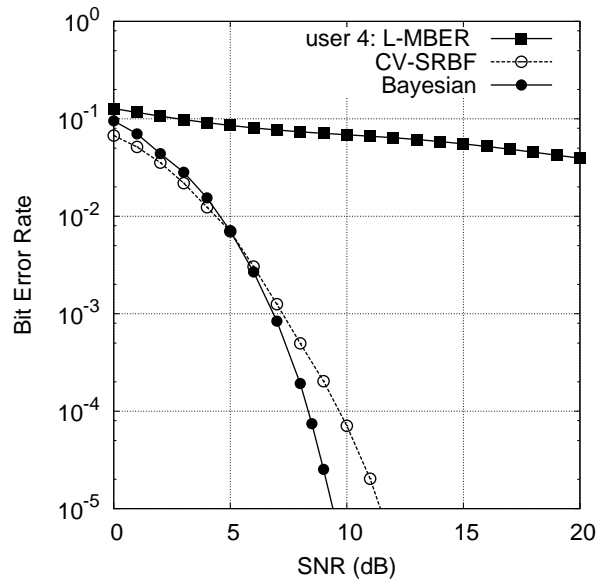


Figure 16: User-four BER performance of the OFS-based SRBF beamformer for the three-element array system supporting four QPSK users, in comparison with the Bayesian and theoretic L-MBER beamforming performance.

influence of the RBF variance  $\rho^2$  to the performance of the 64-term complex-valued SRBF network for user one is illustrated in Fig. 15, given SNR= 6 dB. The beamforming for detecting the user-4 data was demonstrated in Fig. 16, where the BER performance of the

64-term complex-valued SRBF network constructed by the multi-class FRCSM-based OFS was compared with the two benchmarks.

## CONCLUSION

A grey-box approach has been adopted to complex-valued RBF modelling in this contribution. By exploiting the inherent symmetry of the Bayesian beamforming solution for multiple-antenna aided QPSK wireless systems, a complex-valued SRBF network has been proposed for adaptive nonlinear beamforming. Two SRBF network structures have been proposed. The first SRBF network architecture has real-valued RBF weights and the desired complex-valued symmetric nodes' response, just as the Bayesian detection solution, and therefore it is a direct carbon-copy of the Bayesian solution. A modified version of the cluster-variation enhanced clustering algorithm has been derived to implement this SRBF network. The second SRBF network architecture is more general. Unlike the first SRBF network structure, it has complex-valued RBF weights, while maintaining the desired complex-valued symmetric nodes' response. A multi-class FRCSM-based OFS algorithm has been proposed to construct the sparse model for this complex-valued SRBF network.

Although the proposed complex-valued SRBF network is presented in the context of nonlinear detection in QPSK communication systems, it is generically applicable to other classification problems having similar symmetric properties. An important message from this study is that one should always incorporate available *a priori* information in data modelling applications.

## FUTURE RESEARCH DIRECTIONS

Like other complex-valued neural network models, the standard complex-valued RBF network is a black-box model that does not exploit any *a priori* information exists regarding the underlying data generating mechanism. Grey-box approach that explicitly incorporates known knowledge of the system to be modelled is much desired, as it is capable of substantially improving modelling performance. How to adopt *a priori* information into the structure or architecture of complex-valued RBF network, however, is highly problem dependent. In this contribution, we have demonstrated how to modify the architecture of complex-valued RBF network so that the data modelling process guarantees capturing the known symmetric properties of the underlying data generating mechanism. Although we have developed this novel symmetric complex-valued RBF structure in the context of classification applications, the same idea is equally applicable to regression application. Consider for example modelling a system or function from noisy observation data. Let us assume that the function is known to have some symmetric property. A black-box neural network modelled from data may not capture this symmetric property well, while a grey-box model that explicitly incorporates this symmetric property in its topology will guarantee to capture this symmetric property, despite of the presence of noise in data. Another common *a priori*

information existed in many data modelling problems is manifested as boundary value constraints. Standard neural network models have difficulties in this type of data modelling problems. For real-valued data modelling problems, Hong and Chen (2008) have recently developed a novel topology of RBF network, which is able to automatically satisfy a set of boundary value constraints and yet maintains the efficiency of the original RBF modelling procedure. It is of great practical interests to extend this novel RBF topology to the complex-valued domain. Another interesting extension of the present work is to investigate how to adopting *a priori* information in other complex-valued neural network architectures.

## Appendix

Like the real-valued modified Gram-Schmidt orthogonalisation procedure (Chen *et al.*, 1989), the complex-valued version of the modified Gram-Schmidt orthogonalisation procedure calculates the complex-valued  $\mathbf{A}$  matrix row by row and orthogonalises the complex-valued regression matrix  $\Phi$  as follows: at the  $l$ -th stage make the columns  $\phi_i$ ,  $l+1 \leq i \leq K$ , orthogonal to the  $l$ -th column and repeat the operation for  $1 \leq l \leq K-1$ . Specifically, denoting  $\phi_i^{(0)} = \phi_i$ ,  $1 \leq i \leq K$ , then for  $l = 1, 2, \dots, K-1$

$$\left. \begin{aligned} \mathbf{u}_l &= \phi_l^{(l-1)}, \\ a_{l,i} &= \mathbf{u}_l^H \phi_i^{(l-1)} / \mathbf{u}_l^H \mathbf{u}_l, \quad l+1 \leq i \leq K, \\ \phi_i^{(l)} &= \phi_i^{(l-1)} - a_{l,i} \mathbf{u}_l, \quad l+1 \leq i \leq K. \end{aligned} \right\} \quad (39)$$

The last stage of the procedure is simply  $\mathbf{u}_K = \phi_K^{(K-1)}$ . The elements of  $\mathbf{g}$  are computed by transforming  $\mathbf{d}^{(0)} = \mathbf{d}$  in a similar way

$$\left. \begin{aligned} g_l &= \mathbf{u}_l^H \mathbf{d}^{(l-1)} / \mathbf{u}_l^H \mathbf{u}_l, \\ \mathbf{d}^{(l)} &= \mathbf{d}^{(l-1)} - g_l \mathbf{u}_l, \end{aligned} \right\} \quad 1 \leq l \leq K. \quad (40)$$

This orthogonalisation scheme can be used to derive a simple and efficient algorithm for selecting subset models in a forward-regression manner, just as in the real-valued case. First define

$$\Phi^{(l-1)} = [\mathbf{u}_1 \cdots \mathbf{u}_{l-1} \phi_l^{(l-1)} \cdots \phi_K^{(l-1)}]. \quad (41)$$

If some of the columns  $\phi_l^{(l-1)}, \dots, \phi_K^{(l-1)}$  in  $\Phi^{(l-1)}$  have been interchanged, this will still be referred to as  $\Phi^{(l-1)}$  for notational convenience. Recall the notation  $\phi_q^{(l-1)} = [\phi_{1,q}^{(l-1)} \phi_{2,q}^{(l-1)} \cdots \phi_{K,q}^{(l-1)}]^T$ . Given a very small positive number  $T_z$ , which specifies the zero threshold, the  $l$ -th stage of the OFS procedure is given as follows.

*Step 1.* For  $l \leq q \leq K$ :

**Test**—Conditioning number check. If  $(\phi_q^{(l-1)})^H \phi_q^{(l-1)} < T_z$ , the  $q$ -th candidate is not considered.

Compute for  $1 \leq i \leq 4$

$$m_{i,l}^{(q)} = \frac{1}{K^{[i]}} \sum_{k=1}^K \delta(d(k) - b^{[i]}) \phi_{k,q}^{(l-1)}$$

and

$$\left(\sigma_{i,l}^{(q)}\right)^2 = \frac{1}{K^{[i]}} \sum_{k=1}^K \delta(d(k) - b^{[i]}) \left(\phi_{k,q}^{(l-1)} - m_{i,l}^{(q)}\right)^2.$$

Then calculate

$$F_{i,p,l}^{(q)} = \frac{\left(m_{i,l}^{(q)} - m_{p,l}^{(q)}\right)^2}{\left(\left(\sigma_{i,l}^{(q)}\right)^2 + \left(\sigma_{p,l}^{(q)}\right)^2\right)}, \quad 1 \leq i < p \leq 4,$$

and

$$F_l^{(q)} = \frac{1}{6} \sum_{i=1}^3 \sum_{p=i+1}^4 F_{i,p,l}^{(q)}.$$

Let the index set  $\mathcal{J}_q$  be

$$\mathcal{J}_q = \{l \leq q \leq K \text{ and } q \text{ passes **Test**\}.$$

*Step 2.* Find:

$$F_l = F_l^{(q_l)} = \max\{F_l^{(q)}, q \in \mathcal{J}_q\}.$$

Then the  $q_l$ -th column of  $\Phi^{(l-1)}$  is interchanged with the  $l$ -th column of  $\Phi^{(l-1)}$ , and the  $q_l$ -th column of  $\mathbf{A}$  is interchanged with the  $l$ -th column of  $\mathbf{A}$  up to the  $(l-1)$ -th row. This selects the  $q_l$ -th candidate as the  $l$ -th term in the subset model.

*Step 3.* Perform the orthogonalisation as indicated in (39) to derive the  $l$ -th row of  $\mathbf{A}$  and to transform  $\Phi^{(l-1)}$  into  $\Phi^{(l)}$ . Calculate  $g_l$  and update  $\mathbf{d}^{(l-1)}$  into  $\mathbf{d}^{(l)}$  in the way shown in (40).

## References

- Acir, N., Oztura, I., Kuntalp, M., Baklan, B., & Guzelis, C. (2005). Automatic detection of epileptiform events in EEG by a three-stage procedure based on artificial neural networks *IEEE Trans. Biomedical Engineering*, **52**(1), 30–40.
- Aguirre, L. A., Lopes, R. A. M., Amaral, G. F. V., & Letellier, C. (2004). Constraining the topology of neural networks to ensure dynamics with symmetry properties. *Physical Review E*, **69**, 026701-1–026701-11.
- An, P. E., Brown, M., Chen, S., & Harris, C. J. (1993). Comparative aspects of neural network algorithms for on-line modelling of dynamic processes. *Proc. I. MECH. E., Pt. I, J. Systems and Control Eng.*, **207**, 223–241.



- Blogh, J. S., & Hanzo, L. (2002). *Third Generation Systems and Intelligent Wireless Networking – Smart Antennas and Adaptive Modulation*. Chichester, U.K.: Wiley.
- Botoca, C., & Budura, G. (2006). Symbol decision equalizer using a radial basis functions neural network. In *Proc. 7th WSEAS Int. Conf. Neural Networks* (Cavtat, Croatia), June 12-14, 2006, pp.79–84.
- Broomhead, D. S., & Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems*, **2**, 321–355.
- Caiti, A., & Parisini, T. (1994). Mapping ocean sediments by RBF networks. *IEEE J. Oceanic Engineering*, **19**(4), 577–582.
- Cha, I., & Kassam, S. A. (1995). Channel equalization using adaptive complex radial basis function networks. *IEEE J. Selected Areas in Communications*, **13**(1), 122–131.
- Cha, I., & Kassam, S. A. (1996). RBFN restoration of nonlinearly degraded images. *IEEE Trans. Image Processing*, **5**(6), 964–975.
- Chen, S., Billings, S. A., & Luo, W. (1989). Orthogonal least squares methods and their application to non-linear system identification. *Int. J. Control*, **50**(5), 1873–1896.
- Chen, S., Billings, S. A., Cowan, C. F. N., & Grant, P. M. (1990). Non-linear systems identification using radial basis functions. *Int. J. Systems Sci.*, **21**(12), 2513–2539.
- Chen, S., Cowan, C. F. N., Billings, S. A., & Grant, P. M. (1990a). Parallel recursive prediction error algorithm for training layered neural networks. *Int. J. Control*, **51**(6), 1215–1228.
- Chen, S., Cowan, C. F. N., & Grant, P. M. (1991). Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Trans. Neural Networks*, **2**(2), 302–309.
- Chen, S., Billings, S. A., & Grant, P. M. (1992). Recursive hybrid algorithm for non-linear system identification using radial basis function networks. *Int. J. Control*, **55**(5), 1051–1070.
- Chen, S., Mulgrew, B., & Grant, P. M. (1993). A clustering technique for digital communications channel equalization using radial basis function networks. *IEEE Trans. Neural Networks*, **4**(4), 570–579.
- Chen, S., McLaughlin, S., & Mulgrew, B. (1994). Complex-valued radial basis function network, Part I: network architecture and learning algorithms. *Signal Processing*, **35**(1), 19–31.
- Chen, S., McLaughlin, S., & Mulgrew, B. (1994a). Complex-valued radial basis function network, Part II: application to digital communications channel equalisation. *Signal Processing*, **36**(2), 175–188.

- Chen, S. (1995). Nonlinear time series modelling and prediction using Gaussian RBF networks with enhanced clustering and RLS learning. *Electronics Letters*, **31**(2), 117–118.
- Chen, S., Wu, Y., & Luk, B. L. (1999). Combined genetic algorithm optimisation and regularised orthogonal least squares learning for radial basis function networks. *IEEE Trans. Neural Networks*, **10**(5), 1239–1243.
- Chen, S., Hong, X., & Harris, C. J. (2003). Sparse kernel regression modelling using combined locally regularized orthogonal least squares and D-optimality experimental design. *IEEE Trans. Automatic Control*, **48**(6), 1029–1036.
- Chen, S., Hanzo, L., & Wolfgang, A. (2004). Kernel-based nonlinear beamforming construction using orthogonal forward selection with Fisher ratio class separability measure. *IEEE Signal Processing Letters*, **11**(5), 478–481.
- Chen, S., Hong, X., Harris, C. J., & Sharkey, P. M. (2004a). Sparse modelling using orthogonal forward regression with PRESS statistic and regularization. *IEEE Trans. Systems, Man and Cybernetics, Part B*, **34**(2), 898–911.
- Chen, S., Hanzo, L., Ahmad, N. N., & Wolfgang, A. (2005). Adaptive minimum bit error rate beamforming assisted receiver for QPSK wireless communication. *Digital Signal Processing*, **15**(6), 545–567.
- Chen, S., Wolfgang, A., Benedetto, S., Dubamet, P., & Hanzo, L. (2006). Symmetric radial basis function network equaliser. In *Proc. NEWCOM-ACoRN Joint Workshop* (Vienna, Austria), Sept.20-22, 2006, 5 pages.
- Chen, S., Du, H.-Q., & Hanzo, L. (2006a). Adaptive minimum symbol error rate beamforming assisted receiver for quadrature amplitude modulation systems. In *Proc. VTC2006-Spring* (Melbourne, Australia), May 7-10, 2006, pp.2236-2240.
- Chen, S., Labib, K., Kang, R., & Hanzo, L. (2007). Adaptive radial basis function detector for beamforming. In *Proc. ICC 2007* (Glasgow, Scotland), June 24-28, 2007, pp.2967–2972.
- Chen, S., Labib, K., & Hanzo, L. (2007a). Clustering-based symmetric radial basis function beamforming. *IEEE Signal Processing Letters*, **14**(9), 589–592.
- Chen, S., Wolfgang, A., Harris, C. J., & Hanzo, L. (2007b). Symmetric kernel detector for multiple-antenna aided beamforming systems. In *Proc. IJCNN 2007* (Orlando, USA), August 12-17, 2007, pp.2486–2491.
- Chen, S., Hanzo, L., & Tan, S. (2008). Nonlinear beamforming for multiple-antenna assisted QPSK wireless systems. to be presented at *ICC 2008* (Beijing, China), May 19-23, 2008.

- Chinrungrueng, C., & Séquin, C. H. (1995). Optimal adaptive  $\kappa$ -means algorithm with dynamic adjustment of learning rate. *IEEE Trans. Neural Networks*, **6**(1), 1873–1896.
- Deng, J., Sundararajan, N., & Saratchandran, P. (2002). Communication channel equalization using complex-valued minimal radial basis function neural networks. *IEEE Trans. Neural Networks*, **13**(3), 687–696.
- Duda, R. O., & Hart, P. E. (1973). *Pattern Classification and Scene Analysis*. New York: Wiley.
- Espinoza, M., Suykens, J. A. K., & De Moor, B. (2005). Imposing symmetry in least squares support vector machines regression. In *Proc. Joint 44th IEEE Conf. Decision and Control, and European Control Conf. 2005* (Seville, Spain), Dec.12-15, 2005, pp.5716–5721.
- Gan, Q., Saratchandran, P., Sundararajan, N., & Subramanian, K. R. (1999). A complex valued radial basis function network for equalization of fast time varying channels. *IEEE Trans. Neural Networks*, **10**(4), 958–960.
- Gonzalez, J., Rojas, I., Ortega, J., Pomares, H., Fernandez, F. J., & Diaz, A. F. (2003). Multiobjective evolutionary optimization of the size, shape, and position parameters of radial basis function networks for function approximation. *IEEE Trans. Neural Networks*, **14**(6), 1478–1495.
- Gorinevsky, D., Kapitanovsky, A., & Goldenberg, A. (1996). Radial basis function network architecture for nonholonomic motion planning and control of free-flying manipulators. *IEEE Trans. Robotics and Automation*, **12**(3), 491–496.
- Hirose, A. (2006). *Complex-Valued Neural Networks*. Berlin: Springer.
- Karayiannis, N. B., & Randolph-Gips, M. M. (2003). On the construction and training of reformulated radial basis function neural networks. *IEEE Trans. Neural Networks*, **14**(4), 835–846.
- Kim, T. & Adali, T. (2003). Approximation by fully complex multilayer perceptrons. *Neural Computation*, **15**(7), 1641–1666.
- Lee, M.-J., & Choi, Y.-K. (2004). An adaptive neurocontroller using RBFN for robot manipulators. *IEEE Trans. Industrial Electronics*, **51**(3), 711–717.
- Leonard, J. A., & Kramer, M. A. (1991). Radial basis function networks for classifying process faults. *IEEE Control Systems Magazine*, **11**(3), 31–38.
- Li, M.-B., Huang, G.-B., Saratchandran, P., & Sundararajan, N. (2005). Fully complex extreme learning machine. *Neurocomputing*, **68**, 306–314.
- Li, Y., Sundararajan, N., Saratchandran, P., & Wang, Z. (2004). Robust neuro- $H_\infty$  controller design for aircraft auto-landing. *IEEE Trans. Aerospace and Electronic Systems*, **40**(1), 158–167.

- Litva, J., & Lo, T. K. Y. (1996). *Digital Beamforming in Wireless Communications*. London: Artech House.
- Mak, M.-W., & Kung, S.-Y. (2000). Estimation of elliptical basis function parameters by the EM algorithm with application to speaker verification. *IEEE Trans. Neural Networks*, **11**(4), 961–969.
- Mao, K. Z. (2002). RBF neural network center selection based on Fisher ratio class separability measure. *IEEE Trans. Neural Networks*, **13**(5), 1211–1217.
- McLoone, S., Brown, M. D., Irwin, G., & Lightbody, A. (1998). A hybrid linear/nonlinear training algorithm for feedforward neural networks. *IEEE Trans. Neural Networks*, **9**(4), 669–684.
- Moody, J., & Darken, C. J. (1989). Fast-learning in networks of locally-tuned processing units. *Neural Computation*, **1**(2), 281–294.
- Mukai, R., Vlnrotter, V. A., Arabshahi, P., & Jamnejad, V. (2002). Adaptive acquisition and tracking for deep space array feed antennas. *IEEE Trans. Neural Networks*, **13**(5), 1149–1162.
- Muraki, S., Nakai, T., Kita, Y., & Tsuda, K. (2001). An attempt for coloring multichannel MR imaging data. *IEEE Trans. Visualization and Computer Graphics*, **7**(3), 265–274.
- Ng, S. X., Yee, M.-S., & Hanzo, L. (2004). Coded modulation assisted radial basis function aided turbo equalization for dispersive Rayleigh-fading channels. *IEEE Trans. Wireless Communications*, **3**(6), 2198–2206.
- Oyang, Y.-J., Hwang, S.-C., Ou, Y.-Y., Chen, C.-Y., & Chen, Z. W. (2005). Data classification with radial basis function networks based on a novel kernel density estimation algorithm. *IEEE Trans. Neural Networks*, **16**(1), 225–236.
- Paulraj, A., Nabar, R., & Gore, D. (2003). *Introduction to Space-Time Wireless Communications*. Cambridge, U.K.: Cambridge University Press.
- Peng, H., Ozaki, T., Haggan-Ozaki, V., & Toyoda, Y. (2003). A parameter optimization method for radial basis function type models. *IEEE Trans. Neural Networks*, **14**(2), 432–438.
- Poggio, T., & Girosi, F. (1990). Networks for approximation and learning. *Proc. IEEE*, **78**(9), 1481–1497.
- Powell, M. J. D. (1987). Radial basis functions for multivariable interpolation: a review. In J. C. Mason & M. G. Cox (Eds.), *Algorithms for Approximations* (pp.143–167). Oxford.
- Refaee, J. A., Mohandes, M., & Maghrabi, H. (1999). Radial basis function networks for contingency analysis of bulk power systems. *IEEE Trans. Power Systems*, **14**(2), 772–778.

- Rosenblum, M., & Davis, L. S. (1996). An improved radial basis function network for visual autonomous road following. *IEEE Trans. Neural Networks*, **7**(5), 1111–1120.
- Su, C.-T., Yang, T., & Ke, C.-M. (2002). A neural-network approach for semiconductor wafer post-sawing inspection. *IEEE Trans. Semiconductor Manufacturing*, **15**(2), 260–266.
- Tan, K. K., Zhao, S., & Huang, S. (2005). Iterative reference adjustment for high-precision and repetitive motion control applications. *IEEE Trans. Control Systems Technology*, **13**(1), 85–97.
- Tse, D., & Viswanath, P. (2005). *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge University Press.
- Uncini, A., Vecci, L., Campolucci, P., & Piazza, F. (1999). Complex-valued neural networks with adaptive spline activation function for digital ratio links nonlinear equalization. *IEEE Trans. Signal Processing*, **47**(2), 505–514.
- Uykan, Z. (2003). Clustering-based algorithms for single-hidden-layer sigmoid perceptron. *IEEE Trans. Neural Networks*, **14**(3), 708–715.
- Whitehead, B. A., & Choate, T. D. (1994). Evolving space-filling curves to distribute radial basis functions over an input space. *IEEE Trans. Neural Networks*, **5**(1), 15–23.
- Whitehead, B. A. (1996). Genetic evolution of radial basis function coverage using orthogonal niches. *IEEE Trans. Neural Networks*, **7**(6), 1525–1528.
- Yang, C.-C., & Bose, N. K. (2005). Landmine detection and classification with complex-valued hybrid neural network using scattering parameters dataset. *IEEE Trans. Neural Networks*, **16**(3), 743–753.
- Yang, Z. R., & Chen, S. (1998). Robust maximum likelihood training of heteroscedastic probabilistic neural networks. *Neural Networks*, **11**(4), 739–747.

## Additional Reading

- Chen, S., Livingstone, A., Du, H.-Q., & Hanzo, L. (2008). Adaptive minimum symbol error rate beamforming assisted detection for quadrature amplitude modulation. *IEEE Trans. Wireless Communications*, **7**(4), 1140–1145.
- Hong, X., & Chen, S. (2008). A new RBF neural network with boundary value constraints. submitted for publication.
- Nitta, T. (1997). An extension of the back-propagation algorithm to complex numbers. *Neural Networks*, **10**(8), 1391–1415.

- Nitta, T. (2004). Orthogonality of decision boundaries in complex-valued neural networks. *Neural Computation*, **16**(1), 73–97.
- Pande, A., Thakur, A.K., & Roy, S. (2008). Complex-valued neural network in signal processing: A study on the effectiveness of complex valued generalized mean neuron model. *Proc. World Academy of Science, Engineering and Technology*, **27**(2), 240–245.