

# Complex-Valued Symmetric Radial Basis Function Classifier for Quadrature Phase Shift Keying Beamforming Systems

S. Chen, C.J. Harris and L. Hanzo

**Abstract**—We propose a complex-valued symmetric radial basis function (CV-SRBF) network for nonlinear beamforming in multiple-antenna aided communication systems that employ the complex-valued quadrature phase shift keying modulation scheme. The proposed CV-SRBF classifier explicitly exploits the inherent symmetry property of the underlying data generating mechanism, and this significantly enhances the detection accuracy. An orthogonal forward selection (OFS) algorithm based on the multi-class (four-class) Fisher ratio of class separability measure (FRCSM) is derived for constructing parsimonious CV-SRBF classifiers from noisy training data. Effectiveness of the proposed approach is illustrated using simulation, and the results obtained demonstrate that the sparse CV-SRBF classifier constructed by the multi-class FRCSM-based OFS achieves excellent beamforming detection bit error rate performance.

## I. INTRODUCTION

The complex-valued radial basis function (CV-RBF) network [1] has widely been used in nonlinear signal processing applications that involve complex-valued signals. In this CV-RBF network, each radial basis function (RBF) node has a real-valued response that can be interpreted as a conditional probability density function. This interpretation makes such a CV-RBF network particularly useful in the equalisation application of communication channels with complex-valued signals [2], [3], [4], [5], [6]. This CV-RBF modelling method, 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, 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. 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 RBF (RV-RBF) or kernel 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

The authors are with School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK, E-mails: {sqc,cjh,lh}@ecs.soton.ac.uk

RV-RBF network and least squares support vector machine has been demonstrated [7], [8], while in classification-type applications, a novel real-valued symmetric RBF (RV-SRBF) network has been proposed for communication signal detection [9], [10], [11], which explicitly utilising odd symmetry of the underlying optimal Bayesian detection solution.

This contribution continues this theme and extends this 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 inherent 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 the complex-valued quadrature phase shift keying (QPSK) modulation scheme. This naturally leads to our proposed complex-valued symmetric RBF (CV-SRBF) network, which is more general than the CV-RBF of [1], as here both the RBF nodes' response and RBF weights are complex-valued. A novel training algorithm for this CV-SRBF network is derived using the orthogonal-forward-selection (OFS) procedure based on the multi-class (four-class) Fisher ratio of class separability measure (FRCSM), which can be viewed as an extension of the OFS based on the two-class FRCSM [12], [13], [14], originally derived for real-valued two-class classification problems. The effectiveness of the proposed CV-SRBF network and the efficiency of the OFS based on the multi-class FRCSM are demonstrated in nonlinear beamforming for multiple-antenna assisted QPSK wireless systems. Although we apply the proposed CV-SRBF classifier in the context of multiple-antenna aided QPSK beamforming systems, it is equally applicable to other four-class classification problems with similar symmetric properties.

## II. COMPLEX-VALUED RBF NETWORK

First let us recall the CV-RBF network introduced in [1]

$$y(k) = \sum_{i=1}^M \theta_i \phi_i(\mathbf{x}(k)), \quad (1)$$

where  $y(k) \in \mathcal{C}$  is the complex-valued output for the input  $\mathbf{x}(k) \in \mathcal{C}^L$ ,  $\theta_i$  denotes the  $i$ -th complex-valued RBF weight, and the  $i$ -th RBF node's response function is given by

$$\phi_i(\mathbf{x}) = \varphi(\|\mathbf{x} - \mathbf{c}_i\|/\rho_i). \quad (2)$$

Here  $\mathbf{c}_i \in \mathcal{C}^L$  is the  $i$ -th RBF centre,  $\rho_i^2 > 0$  is the  $i$ -th RBF variance, and  $\varphi(\bullet)$  is the usual real-valued basis function. A typical such basis function is the Gaussian function

$$\varphi(\|\mathbf{x} - \mathbf{c}\|/\rho) = e^{-\|\mathbf{x} - \mathbf{c}\|^2/2\rho^2}. \quad (3)$$

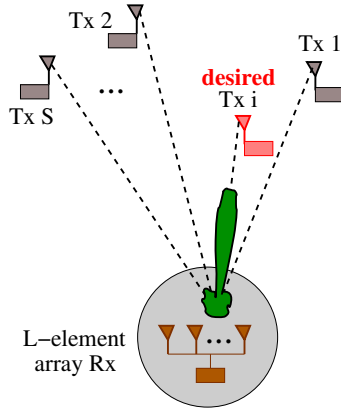


Fig. 1. Beamforming based on multiple-antenna receiver to support multiple users, where  $S$  is the number of users and  $L$  is the number of antenna array elements.

Note that the real-valued node's response (2) has an intrinsic physical interpretation of realising a conditional density function of the underlying data generating mechanism [2]. To illustrate this property, let us consider the nonlinear beamforming for multiple-antenna aided QPSK wireless systems.

#### A. Nonlinear QPSK beamforming

The system supports  $S$  users using the same carrier frequency  $\omega$ . 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 [15], [16], as shown in Fig. 1. Assume that the channel is non-dispersive which does not induce intersymbol interference. Then the received signal vector  $\mathbf{x}(k) = [x_1(k) \ x_2(k) \ \dots \ x_L(k)]^T$  at received is expressed as [17], [18]

$$\mathbf{x}(k) = \mathbf{P} \mathbf{b}(k) + \mathbf{n}(k) = \bar{\mathbf{x}}(k) + \mathbf{n}(k), \quad (4)$$

where  $\mathbf{P}$  is the  $L \times S$  complex-valued system's channel matrix,  $\mathbf{n}(k) = [n_1(k) \ n_2(k) \ \dots \ n_L(k)]^T$ ,  $n_l(k)$  is the complex-valued Gaussian white noise associated with the  $l$ -th channel having  $E[|n_l(k)|^2] = 2\sigma_n^2$ ,  $\mathbf{b}(k) = [b_1(k) \ b_2(k) \ \dots \ b_S(k)]^T$ ,  $b_i(k)$  denotes the  $k$ -th transmitted symbol of user  $i$  and  $b_i(k)$  takes the values from the QPSK symbol set

$$\mathcal{B}_4 = \left\{ b^{[1]} = +1 + j, \ b^{[2]} = -1 + j, \right. \\ \left. b^{[3]} = -1 - j, \ b^{[4]} = +1 - j \right\}. \quad (5)$$

The system's channel matrix is defined by

$$\mathbf{P} = [A_1 \mathbf{s}_1 \ A_2 \mathbf{s}_2 \ \dots \ A_S \mathbf{s}_S], \quad (6)$$

where  $A_i$  is the  $i$ -th non-dispersive channel tap coefficient,

$$\mathbf{s}_i = \left[ e^{j\omega t_1(\eta_i)} \ e^{j\omega t_2(\eta_i)} \ \dots \ e^{j\omega t_L(\eta_i)} \right]^T \quad (7)$$

is the steering vector of source  $i$ , with  $\eta_i$  and  $t_l(\eta_i)$  denoting the angle of arrival and the relative time delay at array element  $l$  for user  $i$ , respectively.

Classically, a linear beamforming receiver is adopted to detect the desired user signal [17], [18]. The output of the

linear beamformer for user  $i$  is defined by  $y_{\text{Lin},i}(k) = \mathbf{w}_i^H \mathbf{x}(k)$ , where  $\mathbf{w}_i = [w_{1,i} \ w_{2,i} \ \dots \ w_{L,i}]^T$  denotes the complex-valued  $i$ -th linear beamformer's weight vector. The decision regarding the transmitted symbol  $b_i(k)$  is given by  $\hat{b}_i(k) = \text{sgn}(y_{\text{Lin},i}(k))$  with

$$\text{sgn}(y) = \begin{cases} b^{[1]} = +1 + j, & y_R \geq 0 \text{ and } y_I \geq 0, \\ b^{[2]} = -1 + j, & y_R < 0 \text{ and } y_I \geq 0, \\ b^{[3]} = -1 - j, & y_R < 0 \text{ and } y_I < 0, \\ b^{[4]} = +1 - j, & y_R \geq 0 \text{ and } y_I < 0, \end{cases} \quad (8)$$

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 as the minimum bit error rate (L-MBER) solution [19], [20]. However, if one is willing to extend the beamforming process to nonlinear, substantial improvement in the achievable system's bit error rate (BER) performance and significant enhancement in the user capacity can be achieved at a cost of increased computational complexity [21].

Let us denote the  $N_b = 4^S$  legitimate combinations of  $\mathbf{b}(k)$  as  $\mathbf{b}_q$ ,  $1 \leq q \leq N_b$ . The noiseless channel output  $\bar{\mathbf{x}}(k)$  only takes values from the vector state set  $\mathcal{X} \triangleq \{\bar{\mathbf{x}}_q = \mathbf{P} \mathbf{b}_q, \ 1 \leq q \leq N_b\}$ , which can be divided into the four subsets conditioned on the value of  $b_i(k)$  as follows

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

where the size of  $\mathcal{X}^{[m,i]}$  is  $N_{sb} = N_b/4$ . Denote the conditional probabilities of receiving  $\mathbf{x}(k)$  given  $b_i(k) = b^{[m]}$  as  $p^{[m,i]}(\mathbf{x}(k)) = p(\mathbf{x}(k) | b_i(k) = b^{[m]})$ . According to Bayes' decision theory [22], the optimal detection strategy is

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

where

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

Define the complex-valued Bayesian decision variable [2]

$$y_{\text{Bay},i}(k) \triangleq b^{[1]} \cdot p^{[1,i]}(\mathbf{x}(k)) + b^{[2]} \cdot p^{[2,i]}(\mathbf{x}(k)) \\ + b^{[3]} \cdot p^{[3,i]}(\mathbf{x}(k)) + b^{[4]} \cdot p^{[4,i]}(\mathbf{x}(k)). \quad (12)$$

The Bayesian detection rule (10) and (11) is equivalent to  $\hat{b}_i(k) = \text{sgn}(y_{\text{Bay},i}(k))$ . The conditional probability  $p^{[m,i]}(\mathbf{x}(k))$  can be expressed as

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

where  $\bar{\mathbf{x}}_q^{[m,i]} \in \mathcal{X}^{[m,i]}$ , and  $\beta_q$  is proportional to the *a priori* probability of  $\bar{\mathbf{x}}_q^{[m,i]}$ . Since all the  $\bar{\mathbf{x}}_q^{[m,i]}$  are equiprobable,  $\beta_q = \beta = \frac{1}{N_{sb}(2\pi\sigma_n^2)^L}$ . It can be seen that the Bayesian decision variable (12) takes the structure of the CV-RBF network (1) with a Gaussian RBF basis function.

More is known, however, for the underlying data generating mechanism. The state subsets  $\mathcal{X}^{[m,i]}$ ,  $1 \leq m \leq 4$ , are distributed symmetrically with respect to each other as summarised in the following lemma.

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

$$\left. \begin{aligned} \mathcal{X}^{[2,i]} &= +j \cdot \mathcal{X}^{[1,i]}, \\ \mathcal{X}^{[3,i]} &= -1 \cdot \mathcal{X}^{[1,i]}, \\ \mathcal{X}^{[4,i]} &= -j \cdot \mathcal{X}^{[1,i]}. \end{aligned} \right\} \quad (14)$$

*Proof:* Consider any  $\bar{\mathbf{x}}_q^{[1,i]} = \mathbf{P}\mathbf{b}_q^{[1,i]} \in \mathcal{X}^{[1,i]}$ , where the  $i$ -th element of  $\mathbf{b}_q^{[1,i]}$  is  $b^{[1]} = +1 + j$ . Noting  $j \cdot b^{[1]} = b^{[2]}$ ,  $j \cdot \bar{\mathbf{x}}_q^{[1,i]} = \mathbf{P}(j \cdot \mathbf{b}_q^{[1,i]}) \in \mathcal{X}^{[2,i]}$ . This proves the first relationship. The proofs of the other two relationships are similar. Given this symmetry, the Bayesian solution (12) can alternatively be expressed as

$$y_{\text{Bay},i}(k) = \sum_{q=1}^{N_{sb}} \beta_q \left\{ b^{[1]} \cdot e^{-\frac{\|\mathbf{x}(k) - \bar{\mathbf{x}}_q^{[1,i]}\|^2}{2\sigma_n^2}} \right. \\ \left. + b^{[2]} \cdot e^{-\frac{\|\mathbf{x}(k) - j \cdot \bar{\mathbf{x}}_q^{[1,i]}\|^2}{2\sigma_n^2}} \right. \\ \left. + b^{[3]} \cdot e^{-\frac{\|\mathbf{x}(k) + \bar{\mathbf{x}}_q^{[1,i]}\|^2}{2\sigma_n^2}} + b^{[4]} \cdot e^{-\frac{\|\mathbf{x}(k) + j \cdot \bar{\mathbf{x}}_q^{[1,i]}\|^2}{2\sigma_n^2}} \right\}, \quad (15)$$

where  $\bar{\mathbf{x}}_q^{[1,i]} \in \mathcal{X}^{[1,i]}$ .

### B. The proposed CV-SRBF network

In the light of the above symmetric property, we generalise the CV-RBF network (1) with the real-valued node's response (2) into the proposed CV-SRBF network by adopting the following complex-valued symmetric node

$$\phi_i(\mathbf{x}) \triangleq b^{[1]} \cdot \varphi(\|\mathbf{x} - \mathbf{c}_i\|/\rho_i) + b^{[2]} \cdot \varphi(\|\mathbf{x} - j \cdot \mathbf{c}_i\|/\rho_i) \\ + b^{[3]} \cdot \varphi(\|\mathbf{x} + \mathbf{c}_i\|/\rho_i) + b^{[4]} \cdot \varphi(\|\mathbf{x} + j \cdot \mathbf{c}_i\|/\rho_i). \quad (16)$$

This CV-SRBF network explicitly incorporates the symmetric property of the underlying data generating mechanism, namely, the Bayesian solution (15). However, the proposed CV-SRBF network (1) with the node's response (16) is also more general than the Bayesian detection solution, because it has complex-valued weights while in the Bayesian solution all the weights are equal to a real positive constant. We now turn to the problem of how to construct such a CV-SRBF classifier, when given a block of training data  $D_K = \{\mathbf{x}(k), d(k) = b_i(k)\}_{k=1}^K$ .

Consider every training data point  $\mathbf{x}(i)$  as a candidate RBF centre. Hence we have  $M = K$  in the CV-SRBF model of (1) and  $\mathbf{c}_i = \mathbf{x}(i)$  for  $1 \leq i \leq K$ , and the all the RBF variances are set to a same constant  $\rho^2$ . Let us now define  $\varepsilon(i) = d(i) - y(i)$  as the modelling residual sequence. Then the model (1) defined over the training data set  $D_K$  can be written in matrix form as

$$\mathbf{d} = \mathbf{\Phi}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \quad (17)$$

where  $\mathbf{d} = [d(1) \ d(2) \ \cdots \ d(K)]^T$ ,  $\boldsymbol{\varepsilon} = [\varepsilon(1) \ \varepsilon(2) \ \cdots \ \varepsilon(K)]^T$ ,  $\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \cdots \ \theta_M]^T$ , and

$$\mathbf{\Phi} = [\phi_1 \ \phi_2 \ \cdots \ \phi_M] \in \mathcal{C}^{K \times M} \quad (18)$$

is the complex-valued regression matrix with the column vectors  $\phi_i = [\phi_i(\mathbf{x}(1)) \ \phi_i(\mathbf{x}(2)) \ \cdots \ \phi_i(\mathbf{x}(K))]^T$ ,  $1 \leq i \leq$

$M$ . Let an orthogonal decomposition of  $\mathbf{\Phi}$  be  $\mathbf{\Phi} = \mathbf{\Omega}\mathbf{A}$ , where

$$\mathbf{A} = \begin{bmatrix} 1 & \alpha_{1,2} & \cdots & \alpha_{1,M} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \alpha_{M-1,M} \\ 0 & \cdots & 0 & 1 \end{bmatrix} \quad (19)$$

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

$$\mathbf{\Omega} = [\boldsymbol{\omega}_1 \ \boldsymbol{\omega}_2 \ \cdots \ \boldsymbol{\omega}_M] \\ = \begin{bmatrix} \omega_{1,1} & \omega_{1,2} & \cdots & \omega_{1,M} \\ \omega_{2,1} & \omega_{2,2} & \cdots & \omega_{2,M} \\ \vdots & \vdots & \vdots & \vdots \\ \omega_{K,1} & \omega_{K,2} & \cdots & \omega_{K,M} \end{bmatrix} \quad (20)$$

with orthogonal columns that satisfy  $\boldsymbol{\omega}_i^H \boldsymbol{\omega}_l = 0$ , if  $i \neq l$ . The model (17) can alternatively be expressed as

$$\mathbf{d} = \mathbf{\Omega}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad (21)$$

where  $\boldsymbol{\gamma} = [\gamma_1 \ \gamma_2 \ \cdots \ \gamma_M]^T = \mathbf{A}\boldsymbol{\theta}$  is the complex-valued weight vector defined in the space spanned by  $\mathbf{\Omega}$ .

### III. MULTI-CLASS FISHER RATIO BASED OFS

First divide the training feature vectors  $\mathbf{X} = \{\mathbf{x}(k)\}_{k=1}^K$  into the  $M_C$  classes ( $M_C = 4$  in our case)

$$\mathbf{X}^{[i]} \triangleq \{\mathbf{x}(k) \in \mathbf{X} : d(k) = b^{[i]}\}, \quad 1 \leq i \leq M_C. \quad (22)$$

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

$$\sum_{i=1}^{M_C} K^{[i]} = K. \quad (23)$$

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

$$m_{i,l} = \frac{1}{K^{[i]}} \sum_{k=1}^{K^{[i]}} \delta(d(k) - b^{[i]}) \omega_{k,l} \quad (24)$$

$$\sigma_{i,l}^2 = \frac{1}{K^{[i]}} \sum_{k=1}^{K^{[i]}} \delta(d(k) - b^{[i]}) (\omega_{k,l} - m_{i,l})^2, \quad (25)$$

where the indicator function

$$\delta(x) = \begin{cases} 1, & x = 0 + j0, \\ 0, & x \neq 0 + j0. \end{cases} \quad (26)$$

Denote the Fisher ratio of the class separation between classes  $\mathbf{X}^{[i]}$  and  $\mathbf{X}^{[q]}$  in the direction of basis  $\boldsymbol{\omega}_l$  as  $F_{i,q,l}$ . Recall that Fisher ratio is defined as the ratio of the interclass difference to the intraclass spread [22], namely,

$$F_{i,q,l} = \frac{(m_{i,l} - m_{q,l})^2}{(\sigma_{i,l}^2 + \sigma_{q,l}^2)}. \quad (27)$$

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

Define the average Fisher ratio of the class separation in the direction of basis  $\omega_l$  as

$$F_l = \frac{2}{(M_C - 1)M_C} \sum_{i=1}^{M_C-1} \sum_{q=i+1}^{M_C} F_{i,q,l}. \quad (28)$$

Based on this average Fisher ratio, significant RBF nodes can be selected in an OFS procedure, just as in the case of two-class problems [12], [13]. Specifically, at the  $l$ -th stage of the OFS procedure, a node is chosen as the  $l$ -th term in the selected CV-SRBF classifier if it produces the largest  $F_l$  among the candidates terms,  $\omega_i$ ,  $l \leq i \leq M$ . The procedure is terminated with a sparse  $M_{\text{spsa}}$ -term classifier when

$$\frac{F_{M_{\text{spsa}}}}{\sum_{l=1}^{M_{\text{spsa}}} F_l} < \xi, \quad (29)$$

where the threshold  $\xi$  determines the size of the selected classifier. The detailed construction procedure, implemented with a complex-valued version of the Gram-Schmidt orthogonalisation [23], is summarised below.

- 1) At the  $l$ -th step where  $l \geq 1$ , for  $1 \leq q \leq M$ ,  $q \neq q_1, \dots, q \neq q_{l-1}$ , compute

$$\alpha_{i,l}^{(q)} = \begin{cases} \frac{\omega_i^H \phi_q}{\omega_i^H \omega_i}, & 1 \leq i < l, \\ 1, & i = l, \end{cases}$$

$$\omega_l^{(q)} = \begin{cases} \phi_q, & l = 1, \\ \phi_q - \sum_{i=1}^{l-1} \alpha_{i,l}^{(q)} \omega_i, & l > 1, \end{cases}$$

and further for  $1 \leq i \leq M_C$  compute

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

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

where  $\omega_{k,l}^{(q)}$  is the  $k$ -th element of  $\omega_l^{(q)}$ . Then calculate

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

$$F_l^{(q)} = \frac{2}{(M_C - 1)M_C} \sum_{i=1}^{M_C-1} \sum_{p=i+1}^{M_C} F_{i,p,l}^{(q)}.$$

- 2) Find

$$q_l = \arg[\max\{F_l^{(q)}, l \leq q \leq M, q \neq q_1, \dots, q \neq q_{l-1}\}],$$

and select  $F_l = F_l^{(q_l)}$ ,  $\alpha_{i,l} = \alpha_{i,l}^{(q_l)}$  for  $1 \leq i \leq l$  and  $\omega_l = \omega_l^{(q_l)}$ .

- 3) The procedure is monitored and terminated at the index value  $l = M_{\text{spsa}}$ , when for example the condition (29) is satisfied. Otherwise, set  $l = l + 1$ , and go to step 1.

A simple and yet effective mechanism can be built into the selection procedure to automatically avoid any numerical ill-conditioning. If a candidate  $\omega_l^{(q)}$  has too low energy, i.e.

TABLE I  
LOCATIONS OF USERS IN TERMS OF ANGLE OF ARRIVAL (AOA) FOR THE TWO-ELEMENT ANTENNA ARRAY HAVING HALF WAVELENGTH SPACING TO SUPPORT THREE QPSK USERS.

user $i$	1	2	3
AOA $\eta_i$	$15^\circ$	$-60^\circ$	$45^\circ$

$(\omega_l^{(q)})^H \omega_l^{(q)}$  is near zero, it will not be considered. The least squares solution for the weight  $\gamma_l$  is simply

$$\gamma_l = \frac{\omega_l^H \mathbf{d}}{\omega_l^H \omega_l}. \quad (30)$$

In general, a desired value for the threshold  $\xi$  has to be determined via cross validation. However, in our specific application to nonlinear beamforming, the number of subset states  $N_{sb}$  is known. Thus, we can simply terminate the selection procedure when  $M_{\text{spsa}} = N_{sb}$ , without need to employ costly cross validation. The RBF variance  $\rho^2$  is not provided by the construction algorithm, and it may be estimated based on cross validation.

#### IV. SIMULATION STUDY

**Example 1.** The example consisted of a two-element antenna array with half wavelength spacing to support three QPSK users. Table I lists the angular locations of the three users with respect to the antenna array. The simulated channel conditions were  $A_i = 1 + j0$ ,  $1 \leq i \leq 3$ . All the three users had an equal signal power. Fig. 2 portrays the BER performance of both the theoretical L-MBER beamformer and the Bayesian detector for the desired user 1. For each signal to noise ratio (SNR) value,  $K = 200$  training samples  $D_K = \{\mathbf{x}(k), b_1(k)\}_{k=1}^K$  were used to construct the CV-SRBF classifier employing the four-class FRCSM-based OFS algorithm. The RBF variance, determined empirically via cross validation, was in the range of 0.2 to 2.0 depending

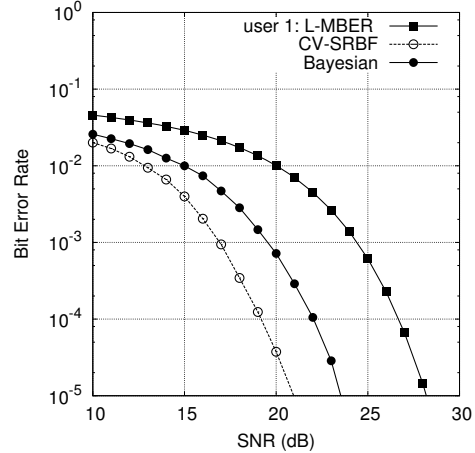


Fig. 2. User-one bit error rate performance in the context of three beamforming detectors for the system of Table I. The CV-SRBF classifier, constructed from 200 noisy training samples using the four-class FRCSM-based OFS, has  $M_{\text{spsa}} = 16$  symmetric RBF nodes. The RBF variance  $\rho^2$  is in the range of 0.2 to 2.0, depending on the SNR value.

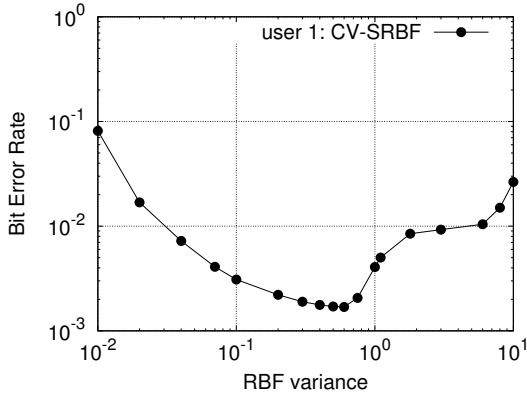


Fig. 3. The influence of the RBF variance on the bit error rate performance of the user-one CV-SRBF classifier for the system of Table I, where SNR=16 dB, the training data length  $K = 200$  and the model size  $M_{spa} = 16$ .

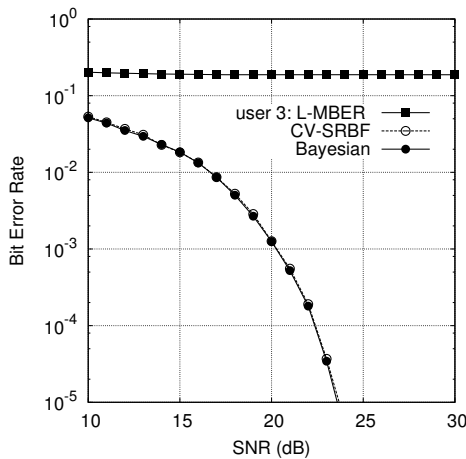


Fig. 4. User-three bit error rate performance in the context of three beamforming detectors for the system of Table I. The CV-SRBF classifier, constructed from 200 noisy training samples using the four-class FRCSM-based OFS, has  $M_{spa} = 16$  symmetric RBF nodes. The RBF variance  $\rho^2$  is in the range of 1.6 to 2.0, depending on the SNR value.

on the SNR value. As the size of the Bayesian detector was  $N_{sb} = 16$ , we terminated the classifier construction at  $M_{spa} = 16$ . The BER performance of the 16-term CV-SRBF detector is also depicted in Fig. 2. It was surprising to see that the CV-SRBF classifier outperformed the Bayesian detector in this case. A possible explanation is that the Bayesian solution is derived under the assumption of white Gaussian channel noise but in the simulation the noise was slightly coloured. Note that the weights of the CV-SRBF network are complex-valued while the weights of the Bayesian beamforming solution are real-valued. Therefore, a 16-term CV-SRBF classifier has a larger model size than the 16-term Bayesian solution. This larger modelling capacity might enable the CV-SRBF classifier to better exploit the statistics of the noise realisation in the simulation. The influence of the RBF variance to the BER performance of the CV-SRBF network constructed by the four-class FRCSM-

TABLE II  
LOCATIONS OF USERS IN TERMS OF ANGLE OF ARRIVAL (AOA) FOR THE THREE-ELEMENT ANTENNA ARRAY HAVING HALF WAVELENGTH SPACING TO SUPPORT FOUR QPSK USERS.

user $i$	1	2	3	4
AOA $\eta_i$	$15^\circ$	$-20^\circ$	$45^\circ$	$-70^\circ$

based OFS was next investigated. Given the SNR value of 16 dB and the number of training samples  $K = 200$ , the BER performance of the 16-terms CV-SRBF classifier as a function of  $\rho^2$  is plotted in Fig. 3. The experiment was repeated for user three, and the BER performance of the three beamformers for detecting user three are shown in Fig. 4, where the training data length for the CV-SRBF network was again  $K = 200$ .

**Example 2.** The system consisted of a three-element antenna array having half wavelength spacing to support four QPSK users, and the angular locations of the four users are summarised in Table II. The simulated channel conditions were  $A_i = 1 + j0$ ,  $1 \leq i \leq 4$ , and all the four users had an equal signal power. Fig. 5 compares the BER performance of three beamformers for user one. For this example, the size of the Bayesian detector was  $N_{sb} = 64$ . The training data length was  $K = 600$  and the 64-term CV-SRBF classifier was constructed using the four-class FRCSM-based OFS. The RBF variance used was in the range of 0.6 to 2.0, depending of the SNR value. Fig. 6 illustrates the influence of the RBF variance to the BER performance of the 64-term CV-SRBF classifier for user one, given the SNR value of 6 dB. Beamforming for user four was also investigated, and the BER performance of the three user-four beamformers are compared in Fig. 7, where the training data length for the CV-SRBF network was again  $K = 600$ .

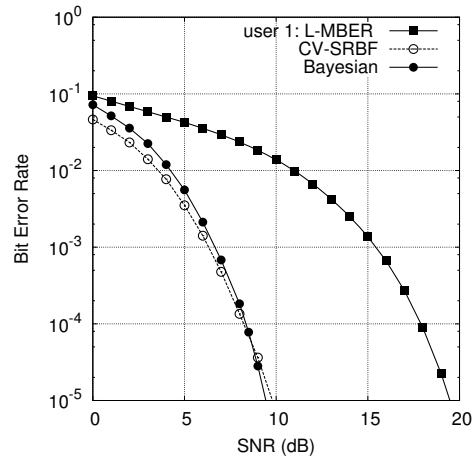


Fig. 5. User-one bit error rate performance in the context of three beamforming detectors for the system of Table II. The CV-SRBF classifier, constructed from 600 noisy training samples using the four-class FRCSM-based OFS, has  $M_{spa} = 64$  symmetric RBF nodes. The RBF variance  $\rho^2$  is in the range of 0.6 to 2.0, depending on the SNR value.

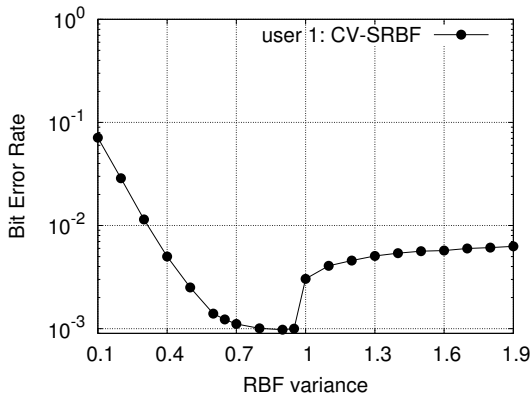


Fig. 6. The influence of the RBF variance on the bit error rate performance of the user-one CV-SRBF classifier for the system of Table II, when SNR=6 dB, the training data length  $K = 600$  and the model size  $M_{spa} = 64$ .

## V. CONCLUSIONS

A complex-valued symmetric RBF classifier has been proposed for nonlinear detection for multiple antenna aided wireless systems that employ the complex-valued QPSK modulation scheme. The proposed model explicitly incorporates the known symmetric property of the underlying data generating mechanism, which substantially enhances the classifier's generalisation capability for nonlinear beamforming application. The orthogonal forward selection procedure based on the multi-class (four-class) Fisher ratio of class separability measure has been derived for constructing sparse CV-SRBF detectors from noisy training data. The proposed approach is fast and results in excellent test error rate performance, as is demonstrated by the simulation results. Although we have presented this sparse CV-SRBF classifier in the context of nonlinear detection in wireless communication systems, it is generically applicable to any four-class

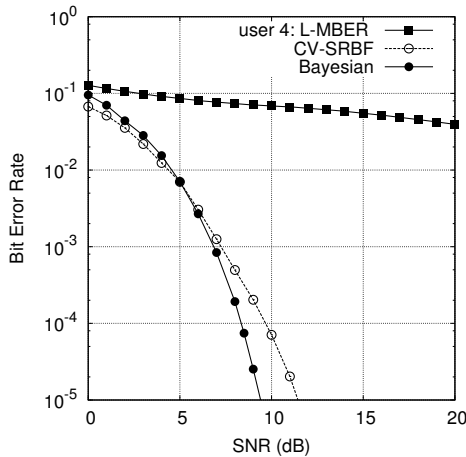


Fig. 7. User-four bit error rate performance in the context of three beamforming detectors for the system of Table II. The CV-SRBF classifier, constructed from 600 noisy training samples using the four-class FRCM-based OFS, has  $M_{spa} = 64$  symmetric RBF nodes. The RBF variance  $\rho^2$  is in the range of 0.2 to 4.0, depending on the SNR value.

classification problem exhibiting a similar symmetry.

## REFERENCES

- [1] S. Chen, S. McLaughlin and B. Mulgrew, "Complex-valued radial basis function network, Part I: Network architecture and learning algorithms," *Signal Processing*, vol.35, pp.19–31, 1994.
- [2] S. Chen, S. McLaughlin and B. Mulgrew, "Complex-valued radial basis function network, Part II: Application to digital communications channel equalisation," *Signal Processing*, vol.36, pp.175–188, 1994.
- [3] I. Cha and S.A. Kassam, "Channel equalization using adaptive complex radial basis function networks," *IEEE J. Selected Areas in Communications*, vol.13, no.1, pp.122–131, 1995.
- [4] Q. Gan, P. Saratchandran, N. Sundararajan and K.R. Subramanian, "A complex valued radial basis function network for equalization of fast time varying channels," *IEEE Trans. Neural Networks*, vol.10, no.4, pp.958–960, 1999.
- [5] J. Deng, N. Sundararajan and P. Saratchandran, "Communication channel equalization using complex-valued minimal radial basis function neural networks," *IEEE Trans. Neural Networks*, vol.13, no.3, pp.687–696, 2002.
- [6] C. Botoca and G. Budura, "Symbol decision equalizer using a radial basis functions neural network," in *Proc. 7th WSEAS Int. Conf. Neural Networks* (Cavta, Croatia), June 12–14, 2006, pp.79–84.
- [7] 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.
- [8] M. Espinoza, J.A.K. Suykens and B. De Moor, "Imposing symmetry in least squares support vector machines regression," in *Proc. Joint 44th IEEE Conf. Decision and Control, and the European Control Conf. 2005* (Seville, Spain), Dec.12–15, 2005, pp.5716–5721.
- [9] S. Chen, A. Wolfgang, S. Benedetto, P. Dubamet and L. Hanzo, "Symmetric radial basis function network equaliser," in *Proc. NEWCOM-ACoRN Joint Workshop* (Vienna, Austria), Sept.20–22, 2006, 5 pages.
- [10] S. Chen, K. Labib, R. Kang and L. Hanzo, "Adaptive radial basis function detector for beamforming," in *Proc. ICC 2007* (Glasgow, Scotland), June 24–28, 2007, pp.2967–2972.
- [11] S. Chen, K. Labib and L. Hanzo, "Clustering-based symmetric radial basis function beamforming," *IEEE Signal Processing Letters*, vol.14, no.9, pp.589–592, 2007.
- [12] K.Z. Mao, "RBF neural network center selection based on Fisher ratio class separability measure," *IEEE Trans. Neural Networks*, vol.13, no.5, pp.1211–1217, 2002.
- [13] S. Chen, L. Hanzo and A. Wolfgang, "Kernel-based nonlinear beamforming construction using orthogonal forward selection with Fisher ratio class separability measure," *IEEE Signal Processing Letters*, vol.11, no.5, pp.478–481, 2004.
- [14] S. Chen, A. Wolfgang, C.J. Harris and L. Hanzo, "Symmetric kernel detector for multiple-antenna aided beamforming systems," in *Proc. IJCNN 2007* (Orlando, Florida), Aug.12–17, 2007, pp.2486–2491.
- [15] A. Paulraj, R. Nabar and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press: Cambridge, UK, 2003.
- [16] D. Tse, and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press: Cambridge, UK, 2005.
- [17] J. Litva and T. K.Y. Lo, *Digital Beamforming in Wireless Communications*. London: Artech House, 1996.
- [18] J.S. Blogh and L. Hanzo, *Third Generation Systems and Intelligent Wireless Networking – Smart Antenna and Adaptive Modulation*. Chichester: John Wiley, 2002.
- [19] S. Chen, N.N. Ahmad and L. Hanzo, "Adaptive minimum bit error rate beamforming," *IEEE Trans. Wireless Communications*, vol.4, no.2, pp.341–348, 2005.
- [20] S. Chen, L. Hanzo, N.N. Ahmad and A. Wolfgang, "Adaptive minimum bit error rate beamforming assisted receiver for QPSK wireless communication," *Digital Signal Processing*, vol.15, no.6, pp.545–567, 2005.
- [21] S. Chen, L. Hanzo and S. Tan, "Nonlinear beamforming for multiple-antenna assisted QPSK wireless systems," to be presented at *ICC 2008* (Beijing, China), May 19–23, 2008.
- [22] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- [23] S. Chen, S.A. Billings and W. Luo, "Orthogonal least squares methods and their applications to non-linear system identification," *Int. J. Control*, vol.50, no.5, pp.1873–1896, 1989.