Available online at www.sciencedirect.com

ScienceDirect

Neurocomputing 71 (2008) 3421–3433

NEUROCOMPUTING

www.elsevier.com/locate/neucom

# Fully complex-valued radial basis function networks: Orthogonal least squares regression and classification

S. Chen[a,*], X. Hong[b], C.J. Harris[a], L. Hanzo[a]

[a] *School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK*
[b] *School of Systems Engineering, University of Reading, Reading RG6 6AY, UK*

Received 29 August 2007; received in revised form 29 November 2007; accepted 3 December 2007
Communicated by J. Zhang
Available online 19 February 2008

## Abstract

We consider a fully complex-valued radial basis function (RBF) network for regression and classification applications. For regression problems, the locally regularised orthogonal least squares (LROLS) algorithm aided with the *D*-optimality experimental design, originally derived for constructing parsimonious real-valued RBF models, is extended to the fully complex-valued RBF (CVRBF) network. Like its real-valued counterpart, the proposed algorithm aims to achieve maximised model robustness and sparsity by combining two effective and complementary approaches. The LROLS algorithm alone is capable of producing a very parsimonious model with excellent generalisation performance while the *D*-optimality design criterion further enhances the model efficiency and robustness. By specifying an appropriate weighting for the *D*-optimality cost in the combined model selecting criterion, the entire model construction procedure becomes automatic. An example of identifying a complex-valued nonlinear channel is used to illustrate the regression application of the proposed fully CVRBF network. The proposed fully CVRBF network is also applied to four-class classification problems that are typically encountered in communication systems. A complex-valued orthogonal forward selection algorithm based on the multi-class Fisher ratio of class separability measure is derived for constructing sparse CVRBF classifiers that generalise well. The effectiveness of the proposed algorithm is demonstrated using the example of nonlinear beamforming for multiple-antenna aided communication systems that employ complex-valued quadrature phase shift keying modulation scheme.

*Keywords:* Complex-valued radial basis function network; Regression; Classification; Orthogonal least squares algorithm; *D*-optimality experimental design; Fisher ratio of class separability measure

## 1. Introduction

Complex-valued artificial neural networks have found wide-ranging applications in processing of complex-valued signals and data [31,24,23,32,20,22]. In this contribution, we re-visit a special class of neural networks, known as the radial basis function (RBF) network. The complex-valued RBF (CVRBF) network of [14] has widely been used in nonlinear signal processing applications that involve complex-valued signals. In this CVRBF network, each RBF node has a real-valued response that can be interpreted as a conditional probability density function. This interpretation makes such a CVRBF network particularly useful in the equalisation application of communication channels with complex-valued signals [15,4,19,17,3]. Because the RBF node's response is real-valued, this CVRBF network is essentially two separate real-valued RBF networks. Various learning methods, such as the orthogonal least squares (OLS) forward selection algorithm [8,9,27,12], can readily be adopted to this CVRBF network for regression and two-class classification applications. This contribution extends the CVRBF network of [14], where each RBF node has a real-valued response, to a fully CVRBF network, where each RBF node has a complex-valued response. The motivation for

[*] Corresponding author.
*E-mail addresses:* sqc@ecs.soton.ac.uk (S. Chen), x.hong@reading.ac.uk (X. Hong), cjh@ecs.soton.ac.uk (C.J. Harris), lh@ecs.soton.ac.uk (L. Hanzo).

0925-2312/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2007.12.003

considering this more general class of fully CVRBF networks is twofold. Firstly, this extension brings the RBF network architecture to the same general level of the multilayer perceptron architecture where the fully complex-valued hidden node has long been proposed [24]. Secondly, this fully CVRBF network arises naturally from detection problems that originate from communication systems employing complex-valued modulation schemes, as will be shown in Section 4 of this contribution. This paper considers this class of fully CVRBF networks for regression and classification, and we develop efficient learning algorithms for constructing sparse fully CVRBF models with excellent generalisation capability.

Among various learning algorithms for regression application of real-valued RBF networks, the local regularisation assisted OLS (LROLS) algorithm combined with the *D*-optimality experimental design criterion [13] is a powerful algorithm for constructing parsimonious real-valued RBF networks that generalise well, because it combines two effective and complementary approaches for modelling, namely, the local regularisation assisted OLS regression [5,6] and the *D*-optimality experimental design [1,21]. By adopting multiple regularisers, the LROLS algorithm is capable of constructing very sparse real-valued RBF models with excellent generalisation capability from noisy data [5,6]. Optimal experimental designs [1] have been used to construct smooth model response surfaces based on the setting of the experimental variables under well controlled experimental conditions. In optimal design, model adequacy is evaluated by design criteria that are statistical measures of goodness of experimental designs by virtue of design efficiency and experimental effort. Quantitatively, model adequacy is measured as function of the eigenvalues of the design matrix, as it is known that the eigenvalues of the design matrix are linked to the covariance matrix of the least squares (LS) parameter estimate. There exist a variety of optimal design criteria based on different aspects of experimental design [1]. The *D*-optimality criterion is most effective in optimising the parameter efficiency and model robustness via the maximisation of the determinant of the design matrix. Combining the *D*-optimality criterion with OLS regression [21] leads to an enhanced construction algorithm, as the coupling effects of the two approaches in the combined algorithm further enhance each other. Moreover, the user only needs to specify a weighting for the *D*-optimality criterion and the model construction process is fully automatic. The value of this weighting does not influence the model selecting procedure critically and it can be chosen with ease from a wide range of values [13]. We extend this combined LROLS algorithm and *D*-optimality experimental design to the fully CVRBF network. An example involving the identification of a complex-valued nonlinear channel is used to demonstrate the effectiveness of the proposed algorithm for constructing sparse fully CVRBF network models for regression application.

The fully CVRBF network is also considered for the application to four-class classification problems that originate from communication systems employing complex-valued quadrature phase shift keying (QPSK) modulation scheme. For the application to two-class classification problems using real-valued RBF networks, the orthogonal forward selection (OFS) based on the two-class Fisher ratio of class separability measure (FRCSM) [27,12] has been demonstrated to be an effective construction algorithm. Because the FRCSM measures the classifier's discriminative power [18], incremental maximisation of the FRCSM leads to a sparse classifier with enhanced generalisation capability. Due to orthogonal decomposition, calculation of the FRCSM along each model basis direction is fast, and this ensures an efficient classifier construction process. We adopt this powerful approach to construct parsimonious fully CVRBF classifiers and derive a complex-valued version of the OFS based on the multi-class (four-class) FRCSM. Application to nonlinear beamforming for multiple-antenna assisted QPSK wireless communication systems [11] is then demonstrated. In general, when to terminate the selection procedure of the CVRBF classifier or the determination of the model size can be decided via cross validation. However, for the particular application to four-class classification problems in communication systems, the number of the underlying channel states [15] is known. Therefore, the construction of a CVRBF classifier can automatically be terminated without the need to apply costly cross validation, when the number of the selected RBF nodes reaches the number of the channel states.

The paper is organised as follows. Section 2 briefly outlines the proposed fully CVRBF network, while Section 3 details the LROLS algorithm with *D*-optimality design for constructing sparse fully CVRBF networks from noisy data as well as presents a case of identifying a complex-valued nonlinear channel using the proposed algorithm. In Section 4 we derive a complex-valued OFS algorithm based on the multi-class FRCSM for constructing parsimonious fully CVRBF classifiers, and this is followed by an application to nonlinear beamforming for multiple-antenna assisted QPSK wireless systems. Our conclusions are offered in Section 5.

## 2. Fully CVRBF network

Consider the modelling of the data set $D_N = \{\mathbf{x}(k), y(k)\}_{k=1}^{N}$, where $N$ is the number of training data, $\mathbf{x}(k) \in \mathscr{C}^m$ is the $k$th complex-valued training input vector, and $y(k)$ is the corresponding complex-valued desired response. More specifically, for regression application, the desired output $y(k) \in \mathscr{C}$. For four-class classification application we adopt the following discrete complex-valued representation of the class label set:

$$\begin{aligned}\mathscr{S}_4 \triangleq \{ & s^{[1]} = +1 + \mathrm{j}, s^{[2]} = -1 + \mathrm{j}, \\ & s^{[3]} = -1 - \mathrm{j}, s^{[4]} = +1 - \mathrm{j}\}\end{aligned} \tag{1}$$

with $\mathrm{j} \triangleq \sqrt{-1}$ and, therefore, $y(k)$ takes values from the set $\mathscr{S}_4$. The aim is to use the RBF network of the form

$$\hat{y}(k) = \sum_{i=1}^{M} \theta_i \phi_i(\mathbf{x}(k)) \tag{2}$$

to capture the underlying data generating mechanism that produces the data set $D_N$, where $\hat{y}(k)$ denotes the complex-valued model output, $\theta_i$ are the complex-valued model weights, $M$ is the number of RBF nodes, and $\phi_i(\mathbf{x}(k))$ denote the CVRBF nodes' response. In particular, for regression $\hat{y}(k)$ represents a prediction of $y(k)$ while for classification $\hat{y}(k)$ is used to estimate the true class label $y(k)$ according to the decision rule

$$\tilde{y}(k) = \mathrm{sgn}(\hat{y}(k)) \triangleq \begin{cases} s^{[1]}, & \Re[\hat{y}(k)] \geqslant 0 \text{ and } \Im[\hat{y}(k)] \geqslant 0, \\ s^{[2]}, & \Re[\hat{y}(k)] < 0 \text{ and } \Im[\hat{y}(k)] \geqslant 0, \\ s^{[3]}, & \Re[\hat{y}(k)] < 0 \text{ and } \Im[\hat{y}(k)] < 0, \\ s^{[4]}, & \Re[\hat{y}(k)] \geqslant 0 \text{ and } \Im[\hat{y}(k)] < 0, \end{cases} \tag{3}$$

where $\Re[\cdot]$ and $\Im[\cdot]$ denote the real and imaginary parts, respectively.

Similar to the case of generic complex-valued neural networks where many complex-valued activation functions can be employed [24], there are many ways of specifying the CVRBF node's response function. One such complex-valued response function is defined by

$$\phi_i(\mathbf{x}) = \varphi(\|\Re[\mathbf{x}] - \Re[\mathbf{c}_i]\|/\rho_i) + \mathrm{j}\varphi(\|\Im[\mathbf{x}] - \Im[\mathbf{c}_i]\|/\rho_i), \tag{4}$$

where $\mathbf{c}_i \in \mathscr{C}^m$ is the $i$th CVRBF centre vector, $\rho_i^2 > 0$ is the $i$th RBF variance, and $\varphi(\cdot)$ is the usual real-valued basis function. Two typical basis functions are the thin-plate-spline function

$$\varphi(\chi/1) = \chi^2 \log(\chi) \tag{5}$$

and the Gaussian function

$$\varphi(\chi/\rho) = \mathrm{e}^{-\chi^2/2\rho^2}. \tag{6}$$

The response function (4) will be adopted for regression application. For four-class classification problems that originate from communication application, however, we will adopt the following RBF node's response function:

$$\begin{aligned} \phi_i(\mathbf{x}) = {} & s^{[1]} \cdot \varphi(\|\mathbf{x} - \mathbf{c}_i\|/\rho_i) + s^{[2]} \cdot \varphi(\|\mathbf{x} - \mathrm{j} \cdot \mathbf{c}_i\|/\rho_i) \\ & + s^{[3]} \cdot \varphi(\|\mathbf{x} + \mathbf{c}_i\|/\rho_i) + s^{[4]} \cdot \varphi(\|\mathbf{x} + \mathrm{j} \cdot \mathbf{c}_i\|/\rho_i), \end{aligned} \tag{7}$$

where the real-valued basis function $\varphi(\cdot)$ is typically chosen to be the Gaussian function of (6). This choice of the RBF node's response explicitly incorporates the desired symmetric property of the underlying data generating mechanism [11], which leads to significant enhancement in classification capability. The choice of this RBF node will be further explained in Section 4.

A significant advantage of the RBF network over other neural networks is that learning can be formulated as a linear-in-the-parameters problem. Specifically, define the modelling residual for $\mathbf{x}(k) \in D_N$ as $e(k) = y(k) - \hat{y}(k)$. Further consider every data points as candidate centres, namely, $M = N$ and $\mathbf{c}_i = \mathbf{x}(i)$ for $1 \leqslant i \leqslant M$. Moreover, set every RBF variance to a given value $\rho_i^2 = \rho^2$. Then we obtain the unified regression model over the data set $D_N$ for both regression and classification problems

$$\mathbf{y} = \mathbf{\Phi}\boldsymbol{\theta} + \mathbf{e}, \tag{8}$$

where $\mathbf{y} = [y(1)\ y(2)\ \cdots\ y(N)]^{\mathrm{T}}$, $\boldsymbol{\theta} = [\theta_1\ \theta_2\ \cdots\ \theta_M]^{\mathrm{T}}$, $\mathbf{e} = [e(1)\ e(2)\ \cdots\ e(N)]^{\mathrm{T}}$ and the complex-valued regression matrix

$$\mathbf{\Phi} = [\boldsymbol{\phi}_1\ \boldsymbol{\phi}_2\ \cdots\ \boldsymbol{\phi}_M] \tag{9}$$

with columns $\boldsymbol{\phi}_i = [\phi_i(\mathbf{x}(1))\ \phi_i(\mathbf{x}(2))\ \cdots\ \phi_i(\mathbf{x}(N))]^{\mathrm{T}}$. Let an orthogonal decomposition of $\mathbf{\Phi}$ be $\mathbf{\Phi} = \mathbf{WA}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & a_{1,2} & \cdots & a_{1,M} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \alpha_{M-1,M} \\ 0 & \cdots & 0 & 1 \end{bmatrix} \tag{10}$$

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

$$\mathbf{W} = [\mathbf{w}_1\ \mathbf{w}_2\ \cdots\ \mathbf{w}_M] = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,M} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,M} \\ \vdots & \vdots & \cdots & \vdots \\ w_{N,1} & w_{N,2} & \cdots & w_{N,M} \end{bmatrix} \tag{11}$$

with columns satisfying $\mathbf{w}_i^{\mathrm{H}} \mathbf{w}_l = 0$, if $i \neq l$. The regression model (8) can alternatively be expressed as

$$\mathbf{y} = \mathbf{Wg} + \mathbf{e}, \tag{12}$$

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

## 3. Locally regularised OLS algorithm with D-optimality design

We first describe the two components of the combined LROLS algorithm and *D*-optimality design. The detailed LROLS algorithm with the *D*-optimality experimental design is then presented, and this is followed by the case of identifying a complex-valued nonlinear channel.

### *3.1. Locally regularised OLS algorithm*

Like the real-valued LROLS algorithm [5,6], the complex-valued version also adopts a similar regularised error criterion defined as

$$J_{\mathrm{R}}(\mathbf{g}, \boldsymbol{\lambda}) = \mathbf{e}^{\mathrm{H}}\mathbf{e} + \sum_{i=1}^{M} \lambda_i |g_i|^2 = \mathbf{e}^{\mathrm{H}}\mathbf{e} + \mathbf{g}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{g}, \tag{13}$$

where $\boldsymbol{\lambda} = [\lambda_1 \ \lambda_2 \ \cdots \ \lambda_M]^{\mathrm{T}}$ is the regularisation parameter vector and $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_M\}$. Similar to the real-valued case [6], with $\mathbf{g}$ set to its optimal value, i.e. at $\partial J_{\mathrm{R}}/\partial \mathbf{g} = 0$, the criterion (13) can be expressed as (see Appendix A)

$$\mathbf{e}^{\mathrm{H}}\mathbf{e} + \mathbf{g}^{\mathrm{H}}\boldsymbol{\Lambda}\mathbf{g} = \mathbf{y}^{\mathrm{H}}\mathbf{y} - \sum_{i=1}^{M}(\mathbf{w}_i^{\mathrm{H}}\mathbf{w}_i + \lambda_i)|g_i|^2. \tag{14}$$

Normalising (14) by $\mathbf{y}^{\mathrm{H}}\mathbf{y}$ yields

$$\frac{(\mathbf{e}^{\mathrm{H}}\mathbf{e} + \mathbf{g}^{\mathrm{H}}\boldsymbol{\Lambda}\mathbf{g})}{\mathbf{y}^{\mathrm{H}}\mathbf{y}} = 1 - \sum_{i=1}^{M}\frac{(\mathbf{w}_i^{\mathrm{H}}\mathbf{w}_i + \lambda_i)|g_i|^2}{\mathbf{y}^{\mathrm{H}}\mathbf{y}}. \tag{15}$$

As in the case of the original OLS algorithm [8], the regularised error reduction ratio due to $\mathbf{w}_i$ is defined by

$$[\mathrm{rerr}]_i = (\mathbf{w}_i^{\mathrm{H}}\mathbf{w}_i + \lambda_i)|g_i|^2/\mathbf{y}^{\mathrm{H}}\mathbf{y}. \tag{16}$$

Based on this ratio, significant regressors can be selected in a forward-regression procedure, and the selection process is terminated at the $n_{\mathrm{s}}$th stage when

$$1 - \sum_{l=1}^{n_{\mathrm{s}}}[\mathrm{rerr}]_l < \xi \tag{17}$$

is satisfied, where $\xi$ is a chosen tolerance. This produces a sparse model containing $n_{\mathrm{s}}$ ($\ll M$) significant regressors.

The regularisation parameters specify the prior distributions of $\mathbf{g}$. Since initially we do not know the optimal value of $\mathbf{g}$, $\lambda_i$ should be initialised to the same small value, and this corresponds to choose a same flat distribution for each prior of $g_i$ [6]. Similar to the real-valued regression model case [6], applying the evidence procedure [26] will lead to the updating formulas for the regularisation parameters

$$\lambda_i^{\mathrm{new}} = \frac{\gamma_i^{\mathrm{old}}}{N - \gamma^{\mathrm{old}}}\frac{\mathbf{e}^{\mathrm{H}}\mathbf{e}}{|g_i|^2}, \quad 1 \leqslant i \leqslant M, \tag{18}$$

where $g_i$ denotes the current optimal weight solution, and

$$\gamma_i = \frac{\mathbf{w}_i^{\mathrm{H}}\mathbf{w}_i}{\lambda_i + \mathbf{w}_i^{\mathrm{H}}\mathbf{w}_i} \quad \text{and} \quad \gamma = \sum_{i=1}^{M}\gamma_i. \tag{19}$$

Usually a few iterations (typically 10–20) are sufficient to find an optimal $\boldsymbol{\lambda}$.

### 3.2. D-optimality experimental design

Adopting the usual concepts of experimental design, we refer to the matrix $\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}$ as the design matrix. The LS estimate of $\boldsymbol{\theta}$ is given by $\hat{\boldsymbol{\theta}} = (\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{y}$. Assume that (8) represents the true data generating process and $\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}$ is nonsingular. Then, the LS estimate $\hat{\boldsymbol{\theta}}$ is unbiased and the covariance matrix of the estimate is determined by the design matrix

$$\begin{cases} E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}, \\ Cov[\hat{\boldsymbol{\theta}}] \propto (\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi})^{-1}. \end{cases} \tag{20}$$

It is well known that the model based on the pure LS estimate tends to be unsatisfactory for an ill conditioned regression matrix (design matrix). The condition number of the design matrix is given by

$$C = \frac{\max\{\kappa_i, 1 \leqslant i \leqslant M\}}{\min\{\kappa_i, 1 \leqslant i \leqslant M\}}, \tag{21}$$

with $\kappa_i$, $1 \leqslant i \leqslant M$, being the eigenvalues of $\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}$. Too large a condition number will result in unstable LS parameter estimate while a small condition number improves model robustness. The $D$-optimality design criterion maximises the determinant of the design matrix for the constructed model. Specifically, let $\boldsymbol{\Phi}_{n_{\mathrm{s}}}$ be a column subset of $\boldsymbol{\Phi}$ representing a constructed $n_{\mathrm{s}}$-term subset model. According to the $D$-optimality criterion, the selected subset model is the one that maximises $\det(\boldsymbol{\Phi}_{n_{\mathrm{s}}}^{\mathrm{H}}\boldsymbol{\Phi}_{n_{\mathrm{s}}})$. This helps to prevent the selection of an oversized ill-posed model and the problem of high parameter estimate variances.

It is straightforward to verify that maximising $\det(\boldsymbol{\Phi}_{n_{\mathrm{s}}}^{\mathrm{H}}\boldsymbol{\Phi}_{n_{\mathrm{s}}})$ is identical to maximising $\det(\mathbf{W}_{n_{\mathrm{s}}}^{\mathrm{H}}\mathbf{W}_{n_{\mathrm{s}}})$ or, equivalently, minimising $-\log\det(\mathbf{W}_{n_{\mathrm{s}}}^{\mathrm{H}}\mathbf{W}_{n_{\mathrm{s}}})$. In fact,

$$\begin{aligned} \det(\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}) &= \det(\mathbf{A}^{\mathrm{H}})\det(\mathbf{W}^{\mathrm{H}}\mathbf{W})\det(\mathbf{A}) \\ &= \det(\mathbf{W}^{\mathrm{H}}\mathbf{W}) = \prod_{i=1}^{M}\mathbf{w}_i^{\mathrm{H}}\mathbf{w}_i \end{aligned} \tag{22}$$

and

$$\begin{aligned} -\log(\det(\mathbf{W}^{\mathrm{H}}\mathbf{W})) &= \sum_{i=1}^{M} -\log(\mathbf{w}_i^{\mathrm{H}}\mathbf{w}_i) \\ &= \sum_{i=1}^{M} -\log(\kappa_i). \end{aligned} \tag{23}$$

### 3.3. Combined LROLS and D-optimality algorithm

The combined LROLS and $D$-optimality algorithm adopts the following combined criterion:

$$J_{\mathrm{R}D}(\mathbf{g}, \boldsymbol{\lambda}, \beta) = J_{\mathrm{R}}(\mathbf{g}, \boldsymbol{\lambda}) + \beta\sum_{i=1}^{M} -\log(\mathbf{w}_i^{\mathrm{H}}\mathbf{w}_i). \tag{24}$$

In this combined algorithm, the updating of the model weights and regularisation parameters is exactly as in the LROLS algorithm, but the selection is according to the combined regularised error reduction ratio defined as

$$[\mathrm{crerr}]_i = \frac{(\mathbf{w}_i^{\mathrm{H}}\mathbf{w}_i + \lambda_i)|g_i|^2 + \beta\log(\mathbf{w}_i^{\mathrm{H}}\mathbf{w}_i)}{\mathbf{y}^{\mathrm{H}}\mathbf{y}} \tag{25}$$

and the selection is terminated with an $n_{\mathrm{s}}$-term model when

$$[\mathrm{crerr}]_l \leqslant 0 \quad \text{for } n_{\mathrm{s}} + 1 \leqslant l \leqslant M. \tag{26}$$

Note that there always exists a subset model size $n_{\mathrm{s}}$ such that (26) holds [21]. The iterative model selection procedure can now be summarised:

*Initialisation*. Set $\lambda_i$, $1 \leqslant i \leqslant M$, to the same small positive value (e.g. $10^{-6}$), and choose a fixed $\beta$. Set iteration $I = 1$.

*Step* 1: Given the current $\boldsymbol{\lambda}$, use the procedure described in Appendix B to select a subset model with $n_{\mathrm{I}}$ terms.

*Step* 2: Update $\boldsymbol{\lambda}$ using (18) with $M=n_{\mathrm{I}}$. If $\boldsymbol{\lambda}$ remains sufficiently unchanged in two successive iterations or a preset maximum iteration number (e.g. 10) is reached, stop; otherwise set $I+=1$ and go to *Step* 1.

The introduction of the $D$-optimality cost into the algorithm further enhances the efficiency and robustness of the selected subset model and, as a consequence, the combined algorithm can often produce sparser models with equally good generalisation properties, compared with the LROLS algorithm alone. An additional advantage is that it simplifies the selection procedure. Note that it is no longer necessary to specify the tolerance $\xi$ and the algorithm automatically terminates when condition (26) is met. The value of weighting $\beta$ does not critically influence the performance of this combined LROLS and $D$-optimality algorithm. This is because the LROLS algorithm alone is capable of producing a very sparse model and the selected model terms are most likely to have large values of $\mathbf{w}_i^{\mathrm{H}}\mathbf{w}_i$. Using the OLS algorithm without local regularisation, this is not necessarily the case, as model terms with small $\mathbf{w}_i^{\mathrm{H}}\mathbf{w}_i$ can have very large $|g_i|^2$ (over-fitted) and consequently will be chosen. Note that with regularisation such over-fitting will not occur. The $D$-optimality design also favours the model terms with large $\mathbf{w}_i^{\mathrm{H}}\mathbf{w}_i$ and therefore the two component criteria in the combined criterion (25) are not in conflict. Thus, the two methods enhance each other. Consequently, the value of $\beta$ is not critical in arriving a desired sparse model, and the suitable weighting $\beta$ can be chosen with ease from a large range of values [13]. It should also be emphasised that the computational complexity of this algorithm is not significantly more than that of the OLS algorithm. This is simply because after the 1st iteration, which has a complexity of the OLS algorithm, the model set contains only $n_1(\ll M)$ terms, and the complexity of the subsequent iteration decreases dramatically. Typically, after a few iterations, the model set will converge to a constant size of very small $n_{\mathrm{s}}$.

### 3.4. A modelling example

Modelling capabilities of the fully CVRBF network and the efficiency of the combined LROLS and $D$-optimality algorithm is illustrated using an example of modelling a complex-valued nonlinear communication channel. Fig. 1 depicts the schematic of this nonlinear channel. The transmitted data symbols $s(k)=s_{\mathrm{R}}(k)+\mathrm{j}s_{\mathrm{I}}(k)$, where $s_{\mathrm{R}}(k)=\Re[s(k)]$ and $s_{\mathrm{I}}(k)=\Im[s(k)]$, take values from the $Q$-quadrature amplitude modulation (QAM) constellation defined by

$$\mathcal{S}_Q \triangleq \{s_{i,l}=(2i-\sqrt{Q}-1)+\mathrm{j}(2l-\sqrt{Q}-1),\quad 1\leqslant i,l\leqslant\sqrt{Q}\}. \tag{27}$$

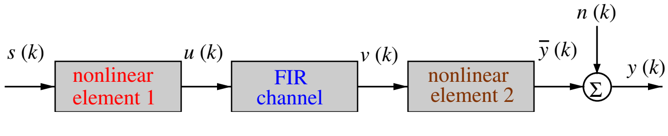


Fig. 1. Baseband discrete-time model of a nonlinear channel.

For $Q=4$, the 4-QAM modulation scheme is equivalent to the QPSK scheme of (1). The first nonlinear element, representing the nonlinear high power amplifier in the transmitter [29], is modelled by the static nonlinearity

$$u(k)=f_{\mathrm{amp}}(s(k))=\frac{2s(k)}{1+|s(k)|^2}\,\mathrm{e}^{\mathrm{j}(\pi/3)|s(k)|^2/(1+|s(k)|^2)}. \tag{28}$$

The time-dispersive transmission medium is modelled as a finite-duration impulse response (FIR) filter whose $z$-transfer function is defined by

$$A(z)=\frac{V(z)}{U(z)}=(0.3725+\mathrm{j}0.2172)\times(1-(0.35+\mathrm{j}0.7)z^{-1})(1-(0.5+\mathrm{j})z^{-1}). \tag{29}$$

The second static nonlinear element is a third-order complex-valued Volterra nonlinearity specified by

$$\bar{y}(k)=f_{\mathrm{Vol}}(v(k))=v(k)+0.2v^2(k)-0.1v^3(k). \tag{30}$$

The additive noise $n(k)=n_{\mathrm{R}}(k)+\mathrm{j}n_{\mathrm{I}}(k)$, where both $n_{\mathrm{R}}(k)$ and $n_{\mathrm{I}}(k)$ are white Gaussian processes having a same variance $\sigma_n^2$. This nonlinear channel thus is characterised by the complex-valued nonlinear model

$$y(k)=\bar{y}(k)+n(k)=f(\mathbf{x}(k))+n(k), \tag{31}$$

where $\mathbf{x}(k)=[s(k)\ \ s(k-1)\ \ s(k-2)]^{\mathrm{T}}$ and $f(\cdot)$ denotes the complex-valued mapping that specifies this nonlinear channel.

For this example, the input vector $\mathbf{x}(k)$ only takes values from the input state set defined by

$$\mathcal{X}=\{\bar{\mathbf{x}}_l, 1\leqslant l\leqslant N_{\mathrm{st}}\}, \tag{32}$$

where $N_{\mathrm{st}}=Q^3$ is the number of input states. Therefore, the noise-free part of the channel output, $\bar{y}(k)$, only takes values from the output state set specified by

$$\bar{\mathcal{Y}}=\{\bar{y}_l=f(\bar{\mathbf{x}}_l), 1\leqslant l\leqslant N_{\mathrm{st}}\}. \tag{33}$$

Similarly, the model output $\hat{y}(k)=\hat{f}(\mathbf{x}(k))$, where $\hat{f}(\cdot)$ denotes the RBF model mapping, over the input set $\mathcal{X}$ is defined by

$$\hat{\mathcal{Y}}=\{\hat{y}_l=\hat{f}(\bar{\mathbf{x}}_l), 1\leqslant l\leqslant N_{\mathrm{st}}\}. \tag{34}$$

The mean state error of the model $\hat{y}(k)=\hat{f}(\mathbf{x}(k))$ is then defined by

$$\text{Mean State Error}=\frac{1}{2N_{\mathrm{st}}}\sum_{l=1}^{N_{\mathrm{st}}}|\bar{y}_l-\hat{y}_l|^2. \tag{35}$$

In the simulation, the energy of the transmitted data symbol $s(k)$ is normalised to $E[|s(k)|^2]=1.0$. Two sets of data $\{\mathbf{x}(k),y(k)\}_{k=1}^N$, each having $N$ points, are generated for the training and testing purposes, respectively. The mean square error (MSE) over a data set $D_N$ is defined by

$$\mathrm{MSE}=\frac{1}{2N}\sum_{k=1}^{N}|y(k)-\hat{y}(k)|^2 \tag{36}$$

with $\hat{y}(k)$ denoting the model output for the input $\mathbf{x}(k)$.

First, we considered identifying the 4-QAM nonlinear channel (31), and in this case the number of input states was $N_{\mathrm{st}}=64$. Given $\sigma_n^2=0.1$, both the training and testing

Table 1
Modelling performance for the 4-QAM nonlinear channel

| Basis function | *D*-weighting | RBF variance | Number of RBFs | MSE for training | MSE for testing | Mean state error |
|---|---|---|---|---|---|---|
| Gaussian | $10^2$–$10^{-6}$ | 0.5 | 15 | 0.120016 | 0.129401 | 0.027739 |
| Thin-plate-spline | $10^2$–$10^{-6}$ | NA | 15 | 0.120895 | 0.128526 | 0.027029 |

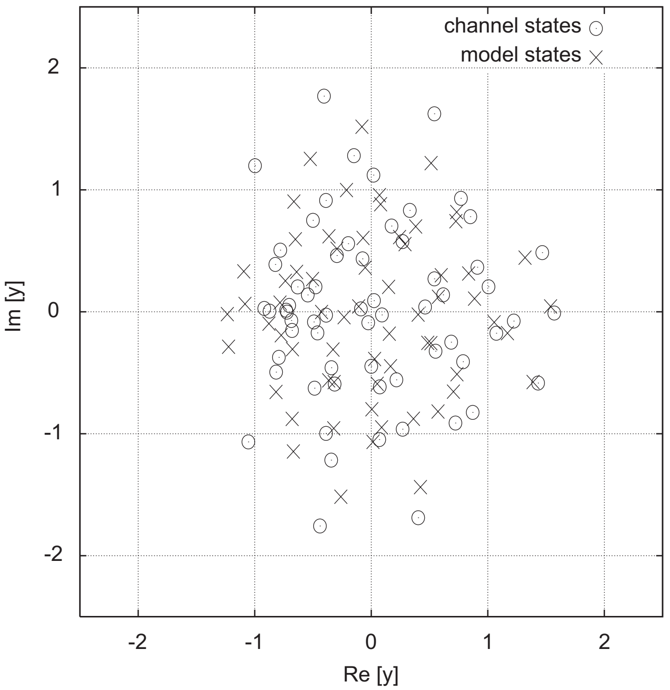


Fig. 2. State constellation for the 4-QAM nonlinear channel, where circles indicate channel states while crosses indicate Gaussian RBF model states.

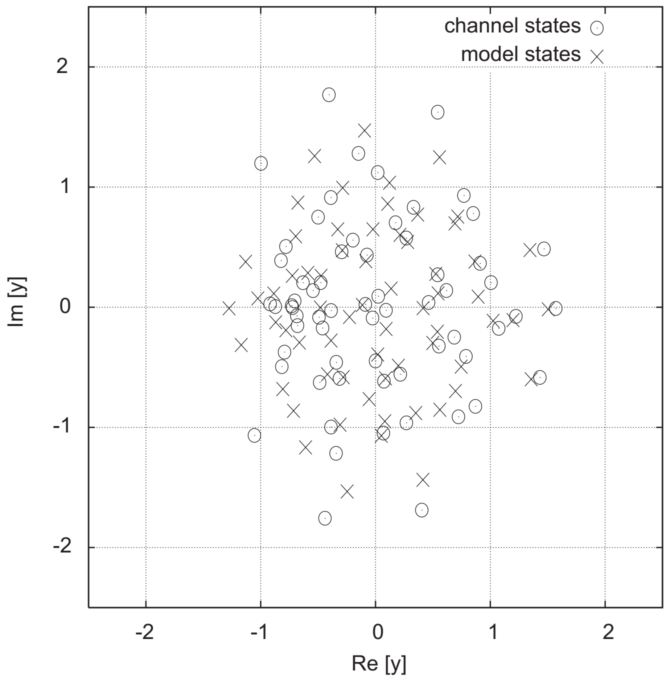


Fig. 3. State constellation for the 4-QAM nonlinear channel, where circles indicate channel states while crosses indicate thin-plate-spline RBF model states.

data sets had $N = 400$ points. The fully CVRBF networks having the node response (4) and with both the Gaussian and thin-plate-spline basis functions were applied to the training data set using the combined LROLS and *D*-optimality algorithm. For this example, it was found that the weighting $\beta$ was not critical at all and any value in $10^2$ to $10^{-6}$ gave the same excellent modelling performance. For the Gaussian RBF network, the RBF variance was set to $\rho^2 = 0.5$ via cross validation. The algorithm automatically selected 15 RBF nodes for both the Gaussian and thin-plate-spline RBF models. Table 1 summarises the modelling performance of the two selected RBF models. It can be seen from Table 1 that the two RBF network models had similarly good generalisation performance. Fig. 2 plots the model output state set $\hat{\mathscr{Y}}$ for the Gaussian RBF model, while Fig. 3 displays $\hat{\mathscr{Y}}$ of the thin-plate-spline RBF model, in comparison with the true channel state set $\bar{\mathscr{Y}}$. The state errors, defined by $\bar{y}_l - \hat{y}_l$, $1 \leqslant l \leqslant N_{\mathrm{st}}$, are plotted in Figs. 4 and 5, respectively, for the two RBF network models.

Next the 16-QAM nonlinear channel was investigated, again given the noise variance $\sigma_n^2 = 0.1$. In this case, the number of input states was increased to $N_{\mathrm{st}} = 4096$, but the number of data points used was only $N = 600$ for both the training and testing data sets. For the thin-plate-spline RBF model, an appropriate value for the *D*-optimality weighting was found to be $\beta = 10.0$ empirically, while for the Gaussian RBF network, $\beta = 10^{-6}$ was found to be appropriately. For the Gaussian RBF network, the RBF variance was chosen to be $\rho^2 = 1.5$ via cross validation. The algorithm selected 50 RBF nodes for the Gaussian RBF model and 57 RBF nodes for the thin-plate-spline RBF model. The modelling performance of these two RBF networks is listed in Table 2, which shows that the two constructed RBF networks had similar good generalisation performance.

## 4. OFS based on fisher ratio for classifier construction

The OFS algorithm based on the multi-class FRCSM is first derived for constructing sparse fully CVRBF classifiers, and this is followed by its application to nonlinear beamforming for multiple-antenna aided QPSK wireless communication systems.

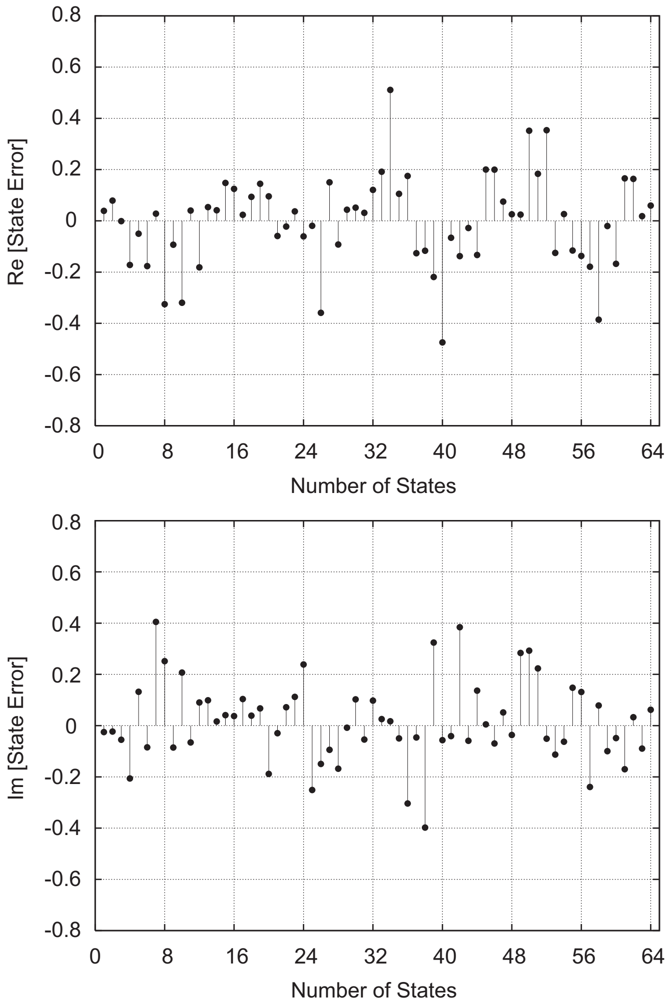


Fig. 4. State errors between the channel and Gaussian RBF model for the 4-QAM nonlinear channel.

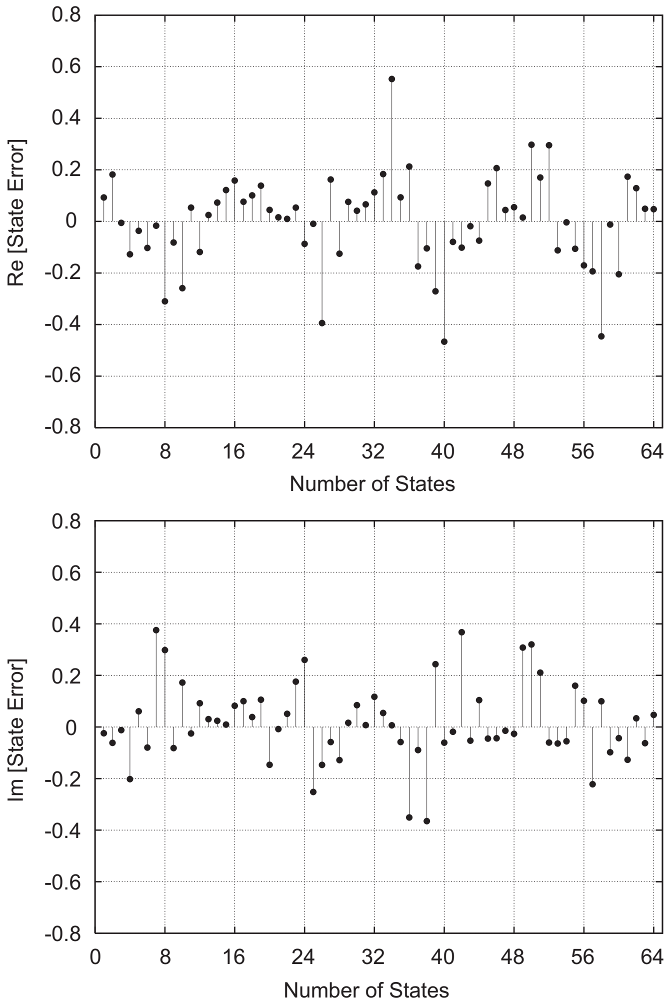


Fig. 5. State errors between the channel and thin-plate-spline RBF model for the 4-QAM nonlinear channel.

### 4.1. *Construction algorithm for fully CVRBF classifiers*

Recall from Section 2 that we are dealing with a multi-class classification problem. First divide the training feature vectors $\mathbf{X} = \{\mathbf{x}(k)\}_{k=1}^{N}$ into the $M_C$ classes ($M_C = 4$ in our case)

$$\mathbf{X}^{[i]} \triangleq \{\mathbf{x}(k) \in \mathbf{X} : y(k) = s^{[i]}\}, \quad 1 \leqslant i \leqslant M_C. \tag{37}$$

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

$$\sum_{i=1}^{M_C} N^{[i]} = N. \tag{38}$$

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

$$m_{i,l} = \frac{1}{N^{[i]}} \sum_{k=1}^{N} \delta\left(y(k) - s^{[i]}\right) w_{k,l} \tag{39}$$

and

$$\sigma_{i,l}^2 = \frac{1}{N^{[i]}} \sum_{k=1}^{N} \delta(y(k) - s^{[i]})(w_{k,l} - m_{i,l})^2, \tag{40}$$

where the indicator function

$$\delta(x) = \begin{cases} 1, & x = 0 + \mathrm{j}0, \\ 0, & x \neq 0 + \mathrm{j}0. \end{cases} \tag{41}$$

Denote the Fisher ratio of the class separation between classes $\mathbf{X}^{[i]}$ and $\mathbf{X}^{[q]}$ in the direction of basis $\mathbf{w}_l$ as $F_{i,q,l}$. Recall that Fisher ratio is defined as the ratio of the interclass difference to the intraclass spread [18], namely,

$$F_{i,q,l} = \frac{(m_{i,l} - m_{q,l})^2}{(\sigma_{i,l}^2 + \sigma_{q,l}^2)}. \tag{42}$$

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

Since we are dealing with multiple $M_C$ classes, we can define the average Fisher ratio of the class separation in the

Table 2
Modelling performance for the 16-QAM nonlinear channel

| Basis function | $D$-weighting | RBF variance | Number of RBFs | MSE for training | MSE for testing | Mean state error |
|---|---|---|---|---|---|---|
| Gaussian | $10^{-6}$ | 1.5 | 50 | 0.128931 | 0.143484 | 0.035443 |
| Thin-plate-spline | 10.0 | NA | 57 | 0.117874 | 0.146306 | 0.038081 |

direction of basis $\mathbf{w}_l$ as

$$F_l = \frac{2}{(M_C-1)M_C}\sum_{i=1}^{M_C-1}\sum_{q=i+1}^{M_C} F_{i,q,l}. \tag{43}$$

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 [27,12]. Specifically, at the $l$th stage of the OFS procedure, a regressor is chosen as the $l$th term in the selected fully CVRBF classifier if it produces the largest $F_l$ among the candidates terms, $\mathbf{w}_i$, $l \leqslant i \leqslant M$. The procedure is terminated with a sparse $n_s$-term classifier when

$$\frac{F_{n_s}}{\sum_{l=1}^{n_s} F_l} \leqslant \xi, \tag{44}$$

where the threshold $\xi$ determines the sparsity of the selected classifier. The detailed OFS procedure based on the multi-class Fisher ratio class separation measure is given in Appendix C. The LS solution for the corresponding sparse model weight vector $\boldsymbol{\theta}_{n_s}$ is readily available from $\mathbf{A}_{n_s}\boldsymbol{\theta}_{n_s} = \mathbf{g}_{n_s}$, given the LS solution of $\mathbf{g}_{n_s}$.

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, the number of users in the system is usually known and hence, the number of the subset underlying channel states, $N_{\text{sub}}$, is given (see the next subsection). Thus, we can simply set the size of the fully CVRBF classifier to $n_s = N_{\text{sub}}$. In this application, therefore, we do not need to employ costly cross validation to determine the model size and the OFS procedure is fully automatic.

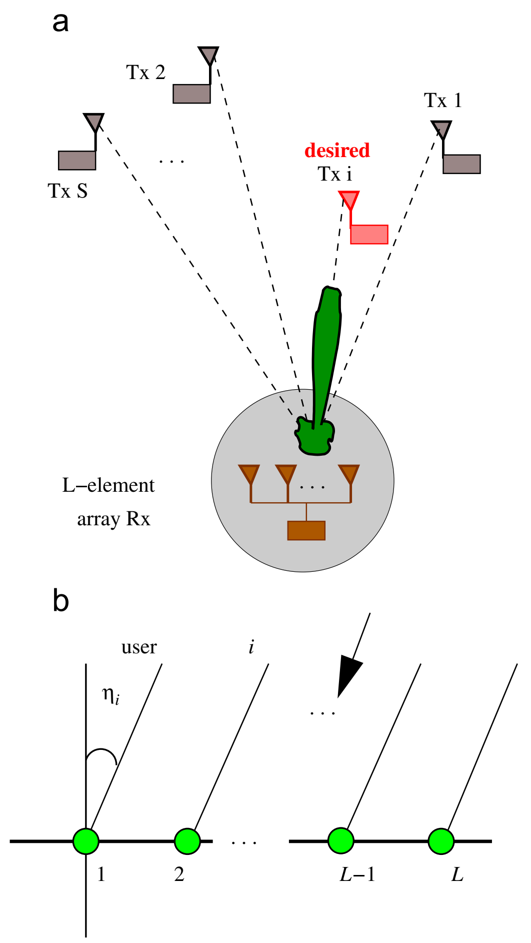


Fig. 6. (a) Beamforming based multiple-antenna receiver to support multiple users, and (b) geometric structure of uniformly spaced antenna array, where $L$ is the number of antenna-array elements, $S$ is the number of users, and $\eta_i$ is the angle of arrival for user $i$.

### 4.2. Application to nonlinear beamforming

Consider a coherent wireless communication system that supports $S$ single-transmit-antenna users of the same carrier frequency $\omega = 2\pi f$ by employing a receiver equipped with a linear antenna array consisting of $L$ uniformly spaced elements [28,30], as shown in Fig. 6. Assume that the channel is non-dispersive and it does not induce intersymbol interference. Then the received signal vector $\mathbf{x}(k) = [x_1(k)\ x_2(k)\ \cdots\ x_L(k)]^{\mathrm{T}}$ at receiver can be expressed as [25,2]

$$\mathbf{x}(k) = \mathbf{P}\mathbf{b}(k) + \mathbf{n}(k) = \bar{\mathbf{x}}(k) + \mathbf{n}(k), \tag{45}$$

where $\mathbf{P}$ is the $L \times S$ complex-valued system's channel matrix, $\mathbf{n}(k) = [n_1(k)\ n_2(k)\ \cdots\ n_L(k)]^{\mathrm{T}}$, $n_l(k)$ is the complex-valued Gaussian white noise associated with the $l$th channel having $E[|n_l(k)|^2] = 2\sigma_n^2$, $\mathbf{b}(k) = [b_1(k)\ b_2(k)\ \cdots\ b_S(k)]^{\mathrm{T}}$, $b_i(k)$ denotes the $k$th transmitted symbol of user $i$, and $b_i(k)$ takes the values from the QPSK symbol set of (1). The system's channel matrix is defined by

$$\mathbf{P} = [A_1\mathbf{s}_1\ \ A_2\mathbf{s}_2\ \ \cdots\ \ A_S\mathbf{s}_S], \tag{46}$$

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

$$\mathbf{s}_i = [\mathrm{e}^{\mathrm{j}\omega t_1(\eta_i)}\ \ \mathrm{e}^{\mathrm{j}\omega t_2(\eta_i)}\ \ \cdots\ \ \mathrm{e}^{\mathrm{j}\omega t_L(\eta_i)}]^{\mathrm{T}} \tag{47}$$

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.

Traditionally, a linear beamformer is adopted to detect the desired user's signal [25,2]. The linear beamformer for user $i$ is defined by

$$\hat{y}_{\mathrm{Lin},i}(k) = \boldsymbol{\theta}_i^{\mathrm{H}} \mathbf{x}(k), \tag{48}$$

where $\boldsymbol{\theta}_i = [\theta_{1,i} \ \theta_{2,i} \ \cdots \ \theta_{L,i}]^{\mathrm{T}}$ is the complex-valued $i$th linear beamformer's weight vector. The decision regarding the transmitted symbol $b_i(k)$ is given by $\hat{b}_i(k) = \mathrm{sgn}(\hat{y}_{\mathrm{Lin},i}(k))$, where $\hat{b}_i(k)$ denotes the estimate of $b_i(k)$ by the linear beamformer (48). The optimal weight vector designed for the linear beamformer is known to be the minimum bit error rate (L-MBER) solution [7,10]. 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 [11].

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

$$\mathscr{X} \triangleq \{\bar{\mathbf{x}}_q = \mathbf{P}\mathbf{b}_q, 1 \leqslant q \leqslant N_b\}, \tag{49}$$

and $\mathscr{X}$ can be divided into the four subsets conditioned on the values of $b_i(k) = s^{[t]}$, $1 \leqslant t \leqslant 4$, as follows:

$$\mathscr{X}^{[t,i]} \triangleq \{\bar{\mathbf{x}}_q^{[t,i]} \in \mathscr{X}, 1 \leqslant q \leqslant N_{\mathrm{sub}} : b_i(k) = s^{[t]}\}, \tag{50}$$

where the size of $\mathscr{X}^{[t,i]}$ is $N_{\mathrm{sub}} = 4^{S-1}$. If the system's channel matrix $\mathbf{P}$ is known, the channel state set $\mathscr{X}$ can be calculated and the optimal nonlinear Bayesian beamforming solution for user $i$ can be expressed as [11]

$$\begin{aligned} \hat{y}_{\mathrm{Bay},i}(k) = \sum_{q=1}^{N_{\mathrm{sub}}} \beta_q \{ & s^{[1]} \cdot \mathrm{e}^{-\|\mathbf{x}(k) - \bar{\mathbf{x}}_q^{[1,i]}\|^2 / 2\sigma_n^2} \\ & + s^{[2]} \cdot \mathrm{e}^{-\|\mathbf{x}(k) - \mathrm{j} \cdot \bar{\mathbf{x}}_q^{[1,i]}\|^2 / 2\sigma_n^2} \\ & + s^{[3]} \cdot \mathrm{e}^{-\|\mathbf{x}(k) + \bar{\mathbf{x}}_q^{[1,i]}\|^2 / 2\sigma_n^2} + s^{[4]} \cdot \mathrm{e}^{-\|\mathbf{x}(k) + \mathrm{j} \cdot \bar{\mathbf{x}}_q^{[1,i]}\|^2 / 2\sigma_n^2} \}, \end{aligned} \tag{51}$$

where $\bar{\mathbf{x}}_q^{[1,i]} \in \mathscr{X}^{[1,i]}$, and $\beta_q$ are positive constants related to the *a priori* probabilities of $\bar{\mathbf{x}}_q^{[1,i]}$. The derivation of the Bayesian beamforming solution (51) is also given in Appendix D. In general, the system' channel matrix is unknown. Given a training data set $D_N = \{\mathbf{x}(k), y_i(k) = b_i(k)\}_{k=1}^N$, our aim is to construct a sparse fully CVRBF classifier or beamformer $\hat{y}_i(k)$ with $n_{\mathrm{s}} = N_{\mathrm{sub}}$ RBF nodes for detecting the user-$i$ data, using the OFS based on FRCSM. In the light of the symmetric structure of the underlying Bayesian beamforming solution (51), we choose the RBF node's response function (7) with Gaussian basis function and set all the RBF variances to a constant $\rho^2$, where appropriate value of $\rho^2$ is determined via cross validation.

In the simulation investigation, a three-element linear antenna array with half wavelength spacing was employed to support four QPSK users. The angular positions of the four users are listed in Table 3. The simulated channel conditions were $A_i = 1.0 + \mathrm{j}0.0$, $1 \leqslant i \leqslant 4$, and all the four users had a equal power. First we consider beamforming for user 1. Fig. 7 depicts the BER performance of the optimal linear beamformer, namely, the L-MBER solution, and the Bayesian beamformer. For user 1, the underlying system was linearly separable. That is, there existed linear beamformers which could separate the four subsets $\mathscr{X}^{[t,1]}$, $1 \leqslant t \leqslant 4$, correctly. Given each signal to noise ratio (SNR), a training set of $N = 600$ samples was generated to construct the fully CVRBF network using the multi-class FRCSM based OFS. For this example, $N_{\mathrm{sub}} = 64$, therefore we stopped the selection procedure after choosing $n_{\mathrm{s}} = 64$ nodes. The value of the RBF variance $\rho^2$ was determined using cross validation, and appropriate values were in the range of 0.6–2.0 depending on the SNR value and noise realisation in the training data. The BER performance of the 64-term fully CVRBF classifier is also plotted in Fig. 7. It can be seen from Fig. 7 that at low SNR values the 64-term fully CVRBF network performed slightly better than 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

Table 3
Angular positions of the four QPSK users with respect to the three-element linear array having half wavelength spacing

| User $i$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Angle of arrival $\eta_i$ (°) | 15 | −20 | 45 | −70 |

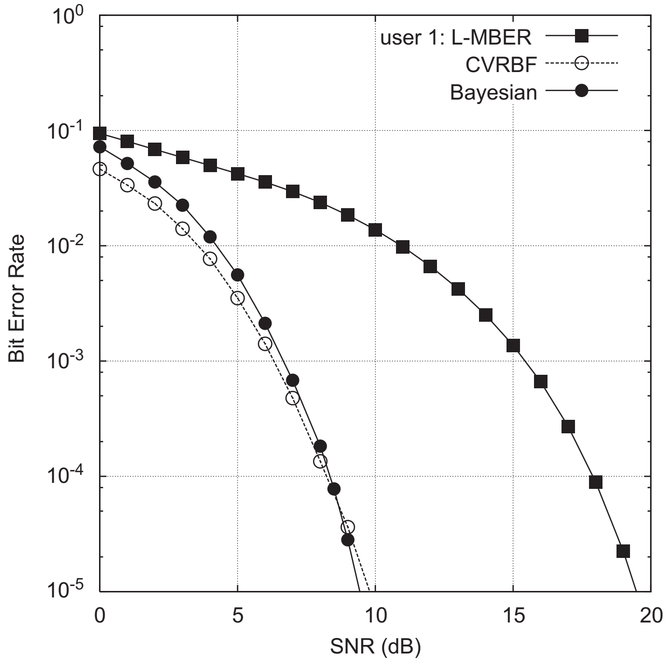


Fig. 7. Comparison of bit error rate performance of three beamformers for desired user 1. The fully CVRBF classifier constructed by the OFS based on FRCSM had 64 RBF nodes.

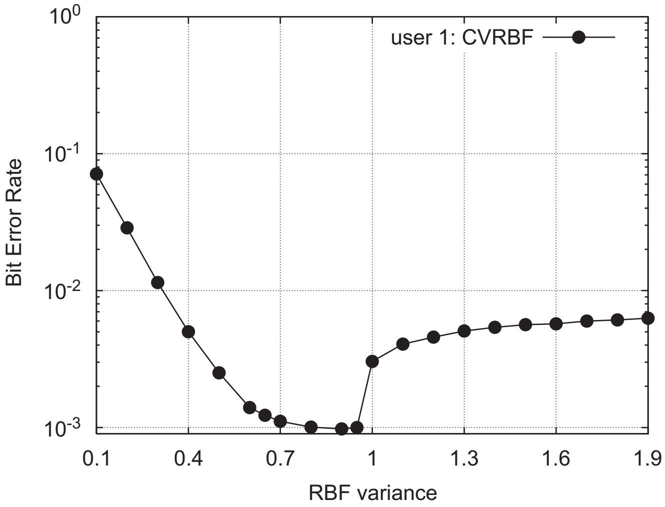


Fig. 8. Influence of the RBF variance $\rho^2$ to the performance of the fully CVRBF classifier for user 1, given $\mathrm{SNR} = 6\,\mathrm{dB}$. The fully CVRBF classifier constructed by the OFS based on FRCSM had 64 RBF nodes.

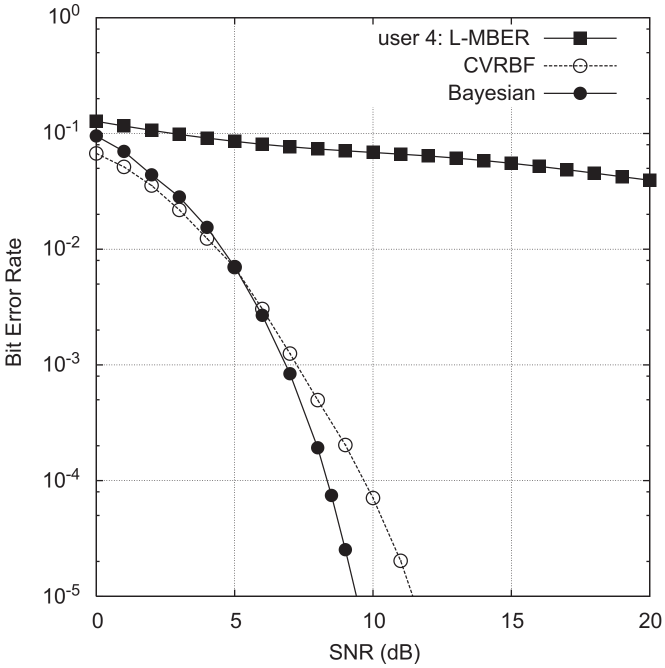


Fig. 9. Comparison of bit error rate performance of three beamformers for desired user 4. The fully CVRBF classifier constructed by the OFS based on FRCSM had 64 RBF nodes.

noise was slightly coloured. Note that the weights of the fully CVRBF network are complex-valued. Therefore, a 64-term fully CVRBF network has a larger model size than the Bayesian solution (whose weights are real-valued). This larger network size might have allowed the fully CVRBF network to exploit the noise statistics in the training data better. The influence of the RBF variance $\rho^2$ to the performance of the fully CVRBF classifier is demonstrated in Fig. 8, given $\mathrm{SNR} = 6\,\mathrm{dB}$.

The beamforming for detecting the user-4 data was also considered, and Fig. 9 shows the BER performance of the L-MBER beamformer and the Bayesian beamformer for desired user 4. In this case, the underlying system was linearly nonseparable as was demonstrated by the high BER floor of the L-MBER beamformer. For each SNR value, a training data set consisting of $N = 600$ samples was used to construct a 64-term fully CVRBF classifier using the OFS based on the multi-class FRCSM, and the BER performance of the resulting classifier is also depicted in Fig. 9. Again the value of the RBF variance was determined via cross validation. Detection of user-4 data was a more difficult task than detection of user-1 data as the former was a nonlinearly separable problem, and the degradation of the fully CVRBF network from the optimal Bayesian solution was more noticeable, as was confirmed in Fig. 9.

## 5. Conclusions

A fully CVRBF network has been proposed for regression and classification applications. For regression problems, the combined LROLS algorithm and the $D$-optimality design, originally derived for real-valued RBF networks, has been extended to select parsimonious fully CVRBF networks with excellent generalisation capability. A modelling example involving the identification of a nonlinear channel has been used to illustrate the proposed approach. It has been demonstrated that combining the local regularisation with the $D$-optimality experimental design provides a state-of-the-art procedure for constructing very sparse regression models with excellent generalisation performance. The performance of the algorithm is insensitive to the $D$-optimality cost weighting, and the model construction process is fully automated. For four-class classification problems, the multi-class FRCSM based OFS algorithm has been derived for constructing sparse fully CVRBF classifiers that generalise well. The capability of the multi-class FRCSM based OFS algorithm has been demonstrated by using it to construct sparse fully CVRBF classifiers in the application to nonlinear beamforming for multiple antenna aided QPSK wireless communication systems.

## Appendix A

The regularised LS solution for $\mathbf{g}$ is obtained by setting $\partial J_{\mathrm{R}}/\partial\mathbf{g} = \mathbf{0}$, that is,

$$\mathbf{W}^{\mathrm{H}}\mathbf{y} = (\mathbf{W}^{\mathrm{H}}\mathbf{W} + \mathbf{\Lambda})\mathbf{g}. \tag{52}$$

Now

$$\begin{aligned}\mathbf{y}^{\mathrm{H}}\mathbf{y} - 2\mathbf{g}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{g} &= (\mathbf{W}\mathbf{g} + \mathbf{e})^{\mathrm{H}}(\mathbf{W}\mathbf{g} + \mathbf{e}) - 2\mathbf{g}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{g} \\ &= \mathbf{g}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}}\mathbf{W}\mathbf{g} + \mathbf{e}^{\mathrm{H}}\mathbf{e} + \mathbf{g}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}}\mathbf{e} \\ &\quad + \mathbf{e}^{\mathrm{H}}\mathbf{W}\mathbf{g} - 2\mathbf{g}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{g}. \end{aligned} \tag{53}$$

Noting (52),

$$\begin{aligned}\mathbf{g}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}}\mathbf{e}-\mathbf{g}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{g} &= \mathbf{g}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}}(\mathbf{y}-\mathbf{W}\mathbf{g})-\mathbf{g}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{g}\\ &= \mathbf{g}^{\mathrm{H}}(\mathbf{W}^{\mathrm{H}}\mathbf{y}-\mathbf{W}^{\mathrm{H}}\mathbf{W}\mathbf{g}-\mathbf{\Lambda}\mathbf{g})\\ &= \mathbf{0}. \qquad (54)\end{aligned}$$

Similarly, $\mathbf{e}^{\mathrm{H}}\mathbf{W}\mathbf{g}-\mathbf{g}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{g}=\mathbf{0}$. Thus, $\mathbf{y}^{\mathrm{H}}\mathbf{y}-2\mathbf{g}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{g}=\mathbf{g}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}}\mathbf{W}\mathbf{g}+\mathbf{e}^{\mathrm{H}}\mathbf{e}$, or

$$\mathbf{e}^{\mathrm{H}}\mathbf{e}+\mathbf{g}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{g}=\mathbf{y}^{\mathrm{H}}\mathbf{y}-\mathbf{g}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}}\mathbf{W}\mathbf{g}-\mathbf{g}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{g}. \qquad (55)$$

## Appendix B

The complex-valued version of the modified Gram–Schmidt orthogonalisation procedure also calculates the $\mathbf{A}$ matrix row by row and orthogonalises $\mathbf{\Phi}$ as follows: at the $l$th stage make the columns $\boldsymbol{\phi}_i$, $l+1\leqslant i\leqslant M$, orthogonal to the $l$th column and repeat the operation for $1\leqslant l\leqslant M-1$. Specifically, denoting $\boldsymbol{\phi}_i^{(0)}=\boldsymbol{\phi}_i$, $1\leqslant i\leqslant M$, then for $l=1,2,\ldots,M-1$

$$\left.\begin{aligned}&\mathbf{w}_l=\boldsymbol{\phi}_l^{(l-1)},\\ &a_{l,i}=\mathbf{w}_l^{\mathrm{H}}\boldsymbol{\phi}_i^{(l-1)}/(\mathbf{w}_l^{\mathrm{H}}\mathbf{w}_l), \quad l+1\leqslant i\leqslant M,\\ &\boldsymbol{\phi}_i^{(l)}=\boldsymbol{\phi}_i^{(l-1)}-a_{l,i}\mathbf{w}_l, \quad l+1\leqslant i\leqslant M.\end{aligned}\right\} \qquad (56)$$

The last stage of the procedure is simply $\mathbf{w}_M=\boldsymbol{\phi}_M^{(M-1)}$. The elements of $\mathbf{g}$ are computed by transforming $\mathbf{y}^{(0)}=\mathbf{y}$ in a similar way:

$$\left.\begin{aligned}&g_l=\mathbf{w}_l^{\mathrm{H}}\mathbf{y}^{(l-1)}/(\mathbf{w}_l^{\mathrm{H}}\mathbf{w}_l+\lambda_l),\\ &\mathbf{y}^{(l)}=\mathbf{y}^{(l-1)}-g_l\mathbf{w}_l,\end{aligned}\right\} \quad 1\leqslant l\leqslant M. \qquad (57)$$

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

$$\mathbf{\Phi}^{(l-1)}=[\mathbf{w}_1 \ \cdots \ \mathbf{w}_{l-1} \ \boldsymbol{\phi}_l^{(l-1)} \ \cdots \ \boldsymbol{\phi}_M^{(l-1)}]. \qquad (58)$$

If some of the columns $\boldsymbol{\phi}_l^{(l-1)},\ldots,\boldsymbol{\phi}_M^{(l-1)}$ in $\mathbf{\Phi}^{(l-1)}$ have been interchanged, this will still be referred to as $\mathbf{\Phi}^{(l-1)}$ for notational convenience. The $l$th stage of the selection procedure is given as follows:

*Step* 1: For $l\leqslant i\leqslant M$, compute

$$\left.\begin{aligned}&g_l^{(i)}=(\boldsymbol{\phi}_i^{(l-1)})^{\mathrm{H}}\mathbf{y}^{(l-1)}/((\boldsymbol{\phi}_i^{(l-1)})^{\mathrm{H}}\boldsymbol{\phi}_i^{(l-1)}+\lambda_i),\\ &[\mathrm{crerr}]_l^{(i)}=(|g_l^{(i)}|^2((\boldsymbol{\phi}_i^{(l-1)})^{\mathrm{H}}\boldsymbol{\phi}_i^{(l-1)}+\lambda_i)\\ &\qquad+\beta\log((\boldsymbol{\phi}_i^{(l-1)})^{\mathrm{H}}\boldsymbol{\phi}_i^{(l-1)}))/(\mathbf{y}^{\mathrm{H}}\mathbf{y}).\end{aligned}\right\}$$

*Step* 2: Find

$$[\mathrm{crerr}]_l=[\mathrm{crerr}]_l^{(i_l)}=\max\{[\mathrm{crerr}]_l^{(i)}, l\leqslant i\leqslant M\}.$$

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

*Step* 3: Perform the orthogonalisation as indicated in (56) to derive the $l$th row of $\mathbf{A}$ and to transform $\mathbf{\Phi}^{(l-1)}$ into $\mathbf{\Phi}^{(l)}$. Calculate $g_l$ and update $\mathbf{y}^{(l-1)}$ into $\mathbf{y}^{(l)}$ in the way shown in (57).

The selection is terminated at the $n_{\mathrm{s}}$ stage when the criterion (26) is satisfied and this produces a subset model containing $n_{\mathrm{s}}$ significant regressors. The algorithm described here is in its standard form. A fast implementation can be adopted, just as shown in the real-valued case [16], to reduce complexity.

## Appendix C

Without regularisation, the modified Gram–Schmidt orthogonalisation procedure is defined in (56) and (57), with all the regularisation parameters set to $\lambda_i=0$ in (57). Also recall from Appendix B that the definition of the regression matrix $\mathbf{\Phi}^{(l-1)}$ at the beginning of the $l$th stage is given in (58). Further introduce the notation $\boldsymbol{\phi}_q^{(l-1)}=[\phi_{1,q}^{(l-1)} \ \phi_{2,q}^{(l-1)} \ \cdots \ \phi_{N,q}^{(l-1)}]^{\mathrm{T}}$. 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 $l\leqslant q\leqslant M$:

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

Compute for $1\leqslant i\leqslant M_C$

$$m_{i,l}^{(q)}=\frac{1}{N^{[i]}}\sum_{k=1}^{N}\delta(y(k)-s^{[i]})\phi_{k,q}^{(l-1)}$$

and

$$(\sigma_{i,l}^{(q)})^2=\frac{1}{N^{[i]}}\sum_{k=1}^{N}\delta(y(k)-s^{[i]})(\phi_{i,q}^{(l-1)}-m_{i,l}^{(q)})^2.$$

Then calculate

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

and

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

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

$$\mathcal{J}_q=\{l\leqslant q\leqslant M \text{ and } q \text{ passes } \mathbf{Test}\}.$$

*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 $\mathbf{\Phi}^{(l-1)}$ is interchanged with the $l$th column of $\mathbf{\Phi}^{(l-1)}$, 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 (56) to derive the $l$th row of $\mathbf{A}$ and to transform $\mathbf{\Phi}^{(l-1)}$ into

$\mathbf{\Phi}^{(l)}$. Calculate $g_l$ and update $\mathbf{y}^{(l-1)}$ into $\mathbf{y}^{(l)}$ in the way shown in (57).

## Appendix D

Denote the conditional probabilities of receiving $\mathbf{x}(k)$ given $b_i(k) = s^{[t]}$ as $p^{[t]}(\mathbf{x}(k)) = p(\mathbf{x}(k)|b_i(k) = s^{[t]})$. According to Bayes' decision theory [18], the optimal detection strategy is

$$\hat{b}_i(k) = s^{[t^*]}, \tag{59}$$

where

$$t^* = \arg \max_{1 \leqslant t \leqslant 4} p^{[t]}(\mathbf{x}(k)). \tag{60}$$

Define the complex-valued Bayesian decision variable [15]

$$\begin{aligned} \hat{y}_{\mathrm{Bay},i}(k) \triangleq {} & s^{[1]} \cdot p^{[1]}(\mathbf{x}(k)) + s^{[2]} \cdot p^{[2]}(\mathbf{x}(k)) \\ & + s^{[3]} \cdot p^{[3]}(\mathbf{x}(k)) + s^{[4]} \cdot p^{[4]}(\mathbf{x}(k)). \end{aligned} \tag{61}$$

The optimal Bayesian detection rule (59) and (60) is equivalent to $\hat{b}_i(k) = \mathrm{sgn}(\hat{y}_{\mathrm{Bay},i}(k))$.

The conditional probability $p^{[t]}(\mathbf{x}(k))$ can be expressed as

$$p^{[t]}(\mathbf{x}(k)) = \sum_{q=1}^{N_{\mathrm{sub}}} \beta_q \mathrm{e}^{-\|\mathbf{x}(k) - \bar{\mathbf{x}}_q^{[t,i]}\|^2 / 2\sigma_n^2}, \tag{62}$$

where $\bar{\mathbf{x}}_q^{[t,i]} \in \mathscr{X}^{[t,i]}$, and $\beta_q$ is proportional to the *a priori* probability of $\bar{\mathbf{x}}_q^{[t,i]}$. Since all the $\bar{\mathbf{x}}_q^{[t,i]}$ are equiprobable, $\beta_q = \beta = 1/N_{\mathrm{sub}}(2\pi\sigma_n^2)^L$. It can be seen that the optimal Bayesian decision variable (61) takes the structure of a CVRBF network [14] with a Gaussian RBF function. The state subsets $\mathscr{X}^{[t,i]}$, $1 \leqslant t \leqslant 4$, are distributed sym-metrically with respect to each other as summarised in the following lemma.

**Lemma.** *The four subsets* $\mathscr{X}^{[t,i]}$, $1 \leqslant t \leqslant 4$, *satisfy*

$$\begin{cases} \mathscr{X}^{[2,i]} = +\mathrm{j} \cdot \mathscr{X}^{[1,i]}, \\ \mathscr{X}^{[3,i]} = -1 \cdot \mathscr{X}^{[1,i]}, \\ \mathscr{X}^{[4,i]} = -\mathrm{j} \cdot \mathscr{X}^{[1,i]}. \end{cases} \tag{63}$$

**Proof.** Consider any $\bar{\mathbf{x}}_q^{[1,i]} = \mathbf{P}\mathbf{b}_q^{[1,i]} \in \mathscr{X}^{[1,i]}$, where the $i$th element of $\mathbf{b}_q^{[1,i]}$ is $s^{[1]} = +1 + \mathrm{j}$. Noting $\mathrm{j} \cdot s^{[1]} = -1 + \mathrm{j} = s^{[2]}$, $\mathrm{j} \cdot \bar{\mathbf{x}}_q^{[1,i]} = \mathbf{P}(\mathrm{j} \cdot \mathbf{b}_q^{[1,i]}) \in \mathscr{X}^{[2,i]}$. This proves the first relationship. The proofs of the other two relationships are similar. □

Substituting the symmetric property (63) into the optimal Bayesian solution (61) leads to the expression (51), where $\bar{\mathbf{x}}_q^{[1,i]} \in \mathscr{X}^{[1,i]}$.

## References

[1] A.C. Atkinson, A.N. Donev, Optimum Experimental Designs, Clarendon Press, Oxford, 1992.

[2] J.S. Blogh, L. Hanzo, Third Generation Systems and Intelligent Wireless Networking—Smart Antennas and Adaptive Modulation, Wiley, Chichester, UK, 2002.

[3] C. Botoca, G. Budura, Symbol decision equalizer using a radial basis functions neural network, in: Proceedings of the 7th WSEAS International Conference on Neural Networks, Cavta, Croatia, June 12–14, 2006, pp. 79–84.

[4] I. Cha, S.A. Kassam, Channel equalization using adaptive complex radial basis function networks, IEEE J. Sel. Areas Commun. 131 (1) (1995) 122–131.

[5] S. Chen, Locally regularised orthogonal least squares algorithm for the construction of sparse kernel regression models, in: Proceedings of the 6th International Conference on Signal Processing, vol. 2, Beijing, China, August 26–30, 2002, pp. 1229–1232.

[6] S. Chen, Local regularization assisted orthogonal least squares regression, Neurocomputing 69 (4–6) (2006) 559–585.

[7] S. Chen, N.N. Ahmad, L. Hanzo, Adaptive minimum bit error rate beamforming, IEEE Trans. Wireless Commun. 4 (2) (2005) 341–348.

[8] S. Chen, S.A. Billings, W. Luo, Orthogonal least squares methods and their application to non-linear system identification, Int. J. Control 50 (5) (1989) 1873–1896.

[9] S. Chen, C.F.N. Cowan, P.M. Grant, Orthogonal least squares learning algorithm for radial basis function networks, IEEE Trans. Neural Networks 2 (2) (1991) 302–309.

[10] S. Chen, L. Hanzo, N.N. Ahmad, A. Wolfgang, Adaptive minimum bit error rate beamforming assisted receiver for QPSK wireless communication, Digital Signal Process. 15 (6) (2005) 545–567.

[11] S. Chen, L. Hanzo, S. Tan, Nonlinear beamforming for multiple-antenna assisted QPSK wireless systems, to be presented at ICC 2008, Beijing, China, May 19–23, 2008.

[12] S. Chen, L. Hanzo, A. Wolfgang, Kernel-based nonlinear beamforming construction using orthogonal forward selection with Fisher ratio class separability measure, IEEE Signal Process. Lett. 11 (5) (2004) 478–481.

[13] S. Chen, X. Hong, C.J. Harris, Sparse kernel regression modeling using combined locally regularized orthogonal least squares and *D*-optimality experimental design, IEEE Trans. Autom. Control 48 (6) (2003) 1029–1036.

[14] S. Chen, S. McLaughlin, B. Mulgrew, Complex-valued radial basis function network, part I: network architecture and learning algorithms, Signal Process. 35 (1994) 19–31.

[15] S. Chen, S. McLaughlin, B. Mulgrew, Complex-valued radial basis function network, part II: application to digital communications channel equalisation, Signal Process. 36 (1994) 175–188.

[16] S. Chen, J. Wigger, Fast orthogonal least squares algorithm for efficient subset model selection, IEEE Trans. Signal Process. 43 (7) (1995) 1713–1715.

[17] J. Deng, N. Sundararajan, P. Saratchandran, Communication channel equalization using complex-valued minimal radial basis function neural networks, IEEE Trans. Neural Networks 13 (3) (2002) 687–696.

[18] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.

[19] Q. Gan, P. Saratchandran, N. Sundararajan, K.R. Subramanian, A complex valued radial basis function network for equalization of fast time varying channels, IEEE Trans. Neural Networks 10 (4) (1999) 958–960.

[20] A. Hirose, Complex-Valued Neural Networks, Springer, Berlin, 2006.

[21] X. Hong, C.J. Harris, Nonlinear model structure design and construction using orthogonal least squares and D-optimality design, IEEE Trans. Neural Networks 13 (5) (2002) 1245–1250.

[22] G.-B. Huang, L. Chen, Enhanced random search based incremental extreme learning machine, Neurocomputing (2008) to appear.

[23] M.-B. Li, G.-B. Huang, P. Saratchandran, N. Sundararajan, Fully complex extreme learning machine, Neurocomputing 68 (2005) 306–314.

[24] T. Lim, T. Adali, Approximation by fully complex multilayer perceptrons, Neural Comput. 15 (7) (2003) 1641–1666.

[25] J. Litva, T.K.Y. Lo, Digital Beamforming in Wireless Communications, Artech House, London, 1996.

[26] D.J.C. MacKay, Bayesian interpolation, Neural Comput. 4 (3) (1992) 415–447.

[27] K.Z. Mao, RBF neural network center selection based on Fisher ratio class separability measure, IEEE Trans. Neural Networks 13 (5) (2002) 1211–1217.

[28] A. Paulraj, R. Nabar, D. Gore, Introduction to Space-Time Wireless Communications, Cambridge University Press, Cambridge, UK, 2003.

[29] S. Pupolin, L.J. Greenstein, Performance analysis of digital radio links with nonlinear transmit amplifier, IEEE J. Sel. Areas Commun. SAC-5 (3) (1987) 534–546.

[30] D. Tse, P. Viswanath, Fundamentals of Wireless Communication, Cambridge University Press, Cambridge, UK, 2005.

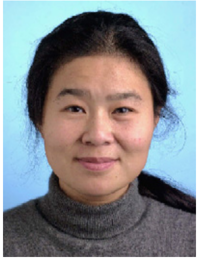
[31] A. Uncini, L. Vecci, P. Campolucci, F. Piazza, Complex-valued neural networks with adaptive spline activation function for digital ratio links nonlinear equalization, IEEE Trans. Signal Process. 47 (2) (1999) 505–514.

[32] C.-C. Yang, N.K. Bose, Landmine detection and classification with complex-valued hybrid neural network using scattering parameters dataset, IEEE Trans. Neural Networks 16 (3) (2005) 743–753.

**Sheng Chen** received his BEng degree in control engineering from Chinese Petroleum University, China, in 1982 and his PhD in control engineering from the City University, London, UK, in 1986. In 2005, he was awarded the Doctor of Sciences (DSc) degree by the University of Southampton, Southampton, UK.

He joined the School of Electronics and Computer Science, University of Southampton, Southampton, UK, in September 1999. He previously held research and academic appointments at the University of Sheffield, Sheffield, the University of Edinburgh, Edinburgh, and the University of Portsmouth, Portsmouth, all in UK. Professor Chen's research works include wireless communications, machine learning and neural networks, finite-precision digital controller design, and evolutionary computation methods. He has published over 300 research papers.

In the database of the world's most highly cited researchers, compiled by Institute for Scientific Information (ISI) of the USA, Dr. Chen is on the list of the highly cited researchers in the engineering category.



**Xia Hong** received her university education at National University of Defence Technology, PR China (BSc, 1984, MSc, 1987), and University of Sheffield, UK (PhD, 1998), all in automatic control.

She worked as a research assistant in Beijing Institute of Systems Engineering, Beijing, China from 1987 to 1993. She worked as a research fellow in the Department of Electronics and Computer Science at University of Southampton from 1997 to 2001. She is currently a lecturer at School of Systems Engineering, University of Reading. She is actively engaged in research into nonlinear systems identification, data modelling, estimation and intelligent control, neural networks, pattern recognition, learning theory and their applications. She has published over 80 research papers, and coauthored a research book.

She was awarded a Donald Julius Groen Prize by IMechE in 1999.

**Chris J. Harris** received his PhD from the University of Southampton, Southampton, UK. He was awarded the Doctor of Sciences (DSc) degree by the University of Southampton.

He previously held appointments at the University of Hull, Hull, the UMIST, Manchester, the University of Oxford, Oxford, and the University of Cranfield, Cranfield, all in UK, as well as being employed by UK. Ministry of Defence. He returned to the University of Southampton as the Lucas Professor of Aerospace Systems Engineering in 1987 to establish the Advanced Systems Research Group and, more recently, Image, Speech and Intelligent Systems Group. His research interests lie in the general area of intelligent and adaptive systems theory and its application to intelligent autonomous systems such as autonomous vehicles, management infrastructures such as command & control, intelligent control, and estimation of dynamic processes, multi-sensor data fusion, and systems integration. He has authored and co-authored 12 research books and over 400 research papers, and he is the associate editor of numerous international journals.

Dr. Harris was elected to the Royal Academy of Engineering in 1996, was awarded the IEE Senior Achievement medal in 1998 for his work in autonomous systems, and the highest international award in IEE, the IEE Faraday medal, in 2001 for his work in intelligent control and neurofuzzy systems.

**Lajos Hanzo** received his Master degree in electronics in 1976 and his doctorate in 1983. In 2004 he was awarded the Doctor of Sciences (DSc) degree by the University Southampton, Southampton, UK.

During his 30-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the School of Electronics and Computer Science, the University of Southampton, UK, where he holds the chair in telecommunications. He has coauthored 14 John Wiley/IEEE Press books totalling about 10 000 pages on mobile radio communications, published in excess of 600 research papers, organised and chaired conference sessions, presented overview lectures and has been awarded a number of distinctions. He is an enthusiastic supporter of industrial-academic liaison. He also offers a range of industrial research overview courses.

Professor Hanzo is a Fellow of the Royal Academy of Engineering (FREng), UK. He is an IEEE Distinguished Lecturer of both the Communications Society and the Vehicular Technology Society as well as a Fellow of both the IEEE and IEE. He is a non-executive director of the Virtual Centre of Excellence (VCE) in mobile communications, UK, and a governor of the IEEE VT society.