Symmetric RBF Classifier for Nonlinear Detection in Multiple-Antenna-Aided Systems

Sheng Chen, *Senior Member, IEEE*, Andreas Wolfgang, *Member, IEEE*, Chris J. Harris, and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—In this paper, we propose a powerful symmetric radial basis function (RBF) classifier for nonlinear detection in the so-called "overloaded" multiple-antenna-aided communication systems. By exploiting the inherent symmetry property of the optimal Bayesian detector, the proposed symmetric RBF classifier is capable of approaching the optimal classification performance using noisy training data. The classifier construction process is robust to the choice of the RBF width and is computationally efficient. The proposed solution is capable of providing a signal-to-noise ratio (SNR) gain in excess of 8 dB against the powerful linear minimum bit error rate (BER) benchmark, when supporting four users with the aid of two receive antennas or seven users with four receive antenna elements.**

***Index Terms*—Classification, multiple-antenna system, orthogonal forward selection, radial basis function (RBF), symmetry.**

I. Introduction

RADIAL BASIS FUNCTION (RBF) or kernel modeling techniques have found wide-ranging applications in regression and classification [1]–[22]. The kernel modeling method constitutes a black-box approach that seeks a (usually sparse) model representation extracted from the training data. Adopting black-box modeling is appropriate, if no *a priori* information exists regarding the underlying data generating mechanism. However, a fundamental principle in practical data modeling is that if there exists *a priori* information concerning the system to be modeled it should be incorporated in the modeling process. The use of prior knowledge in data modeling often leads to an improved performance. In regression-type applications, the symmetric properties of the underlying system have been exploited by imposing symmetry in both RBF networks and least squares support vector machines (SVMs) [23], [24]. An important message from these two studies is worth revisiting. Many real-life phenomena exhibit inherent symmetry, but these properties are hard to infer from noisy data with the aid of black-box-type 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 regression modeling performance.

Manuscript received September 15, 2006; revised April 23, 2007; accepted September 24, 2007. This work was supported by the European Union under the auspices of the Phoenix and Newcom projects.

The authors are with the Communication Research Group, School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: sqc@ecs.soton.ac.uk; aw03r@ecs.soton.ac.uk; cjh@ecs.soton.ac.uk; lh@ecs.soton.ac.uk).

Digital Object Identifier 10.1109/TNN.2007.911745

In this paper, we consider nonlinear detection in multiple-antenna-assisted beamforming systems. Detection in communication receivers, in general, and in multiple-antenna-aided beamforming systems, in particular, can be viewed as a classification problem, and both RBF as well as other kernel models have been applied to solve this nonlinear detection problem [25]–[34]. A kernel-based classifier or detector attempts to realize or approximate the underlying optimal Bayesian solution. Previous studies [25]–[34] have shown that a black-box kernel detector typically requires more kernels than the number of the channel states to approximate the Bayesian detector, and moreover there often exists a performance gap between the kernel detector and the Bayesian solution. This performance degradation can be explained as follows. The Bayesian nonlinear detection solution has an inherent symmetry, because the signal states corresponding to the different signal classes are distributed symmetrically with respect to the Bayesian decision boundary [35]. A black-box RBF, or kernel classifier, however, has difficulty realizing this symmetry accurately. In a single-antenna–single-user system, the channel impulse response (CIR) can be identified during training and hence the channel states required to form the optimal Bayesian detector can be explicitly computed. However, for downlink situation, the receiver in a multiple-antenna-aided beamforming system only has access to the desired user's transmitted symbols during training and does not have access to all the other interfering users' transmitted data, and identifying all the users' CIRs is hard to achieve if it is not impossible. Hence, it is more realistic to implement an RBF or kernel detector directly using the channel-contaminated data for training.

The novelty of this paper is that we propose a symmetric RBF classifier for multiple-antenna-aided communication systems, which renders realization of the symmetric Bayesian detection solution easier. The orthogonal forward selection (OFS) procedure [18], [21], [33], [34] can readily be applied to construct a sparse representation for this symmetric RBF classifier based on various criteria, such as the Fisher ratio of class separability measure (FRCSM) [33] and the leave-one-out misclassification rate (LOO-MR) [21]. The OFS procedures based on the FRCSM and the LOO-MR are computationally very efficient, in comparison to other existing kernel construction methods. We adopt the FRCSM, partly because it is computationally even simpler to implement than the LOO-MR. Even though we do not directly minimize the misclassification rate, we will demonstrate that the sparse symmetric RBF classifier constructed by incrementally maximizing the FRCSM is capable of approaching the minimum misclassification rate of the optimal Bayesian detector. It is also worth pointing out that constructing classifiers by mini-

mizing classification error rate directly does not always produce good performance, particularly for problems with small training sample sets, as was demonstrated in [36]. This issue is even more crucial as in our application the error rate is typically very small (e.g. 10^{-4}). Thus, any training sample set one may have becomes relatively speaking very small, with respect to the error rate. The FRCSM on the other hand does not suffer from this kind of problem. The advantage of the proposed symmetric RBF classifier is demonstrated in challenging detection scenarios, when the number of users supported is almost twice the number of antenna elements, while conventional techniques cannot support more users than the number of antenna elements [37], [38]. Although we apply the proposed symmetric RBF classifier in the context of multiple-antenna aided beamforming systems, it is applicable in symmetric classification problems. To the best of our knowledge, this is the first time that the symmetry is explicitly exploited in RBF or kernel classifier construction.

The outline of this paper is as follows. In Section II, we present the beamforming model considered and discuss the inherent symmetric structure of the associated classification task. Based on the system model of Section II, the novel symmetric RBF classifier is derived in Section III, while its achievable performance is discussed in Section IV. In Section V, we offer our conclusions.

II. Multiple-Antenna-Aided Beamforming Receiver

Consider a coherent communication system that supports S users, where each user transmits using the same carrier frequency of $\omega = 2\pi f$. For such a system, user separation can be achieved in the spatial or angular domain [37], [38] if the receiver is equipped with a linear antenna array consisting of $L > 1$ uniformly spaced elements. Assume that the channel is nondispersive which does not induce intersymbol interference. Then, the symbol-rate complex-valued received signal samples can be expressed as [39], [40]

$$x_l(k) = \sum_{i=1}^{S} A_i b_i(k) e^{j\omega t_l(\theta_i)} + n_l(k) = \bar{x}_l(k) + n_l(k) \quad (1)$$

for $1 \le l \le L$, where $t_l(\theta_i)$ is the relative time delay at array element l for source i, with θ_i being the direction (angle) of arrival for source i, $n_l(k)$ is the complex-valued Gaussian white noise with $E[|n_l(k)|^2] = 2\sigma_n^2$, A_i is the complex-valued nondispersive channel coefficient of user i, and $b_i(k)$ is the kth symbol of user i, which assumes values from the binary phase-shift keying (BPSK) symbol set, i.e., $b_i(k) \in \{\pm 1\}$. Without loss of generality, source 1 is assumed to be the desired user and the rest of the sources are interfering users. The desired user's signal-to-noise ratio (SNR) is given by $\text{SNR} = |A_1|^2\sigma_b^2/2\sigma_n^2$, where $\sigma_b^2 = 1$ is the BPSK symbol energy, and the desired signal-to-interferer i ratio (SIR) is defined by $\text{SIR}_i = |A_1|^2/|A_i|^2$, for $2 \le i \le S$. The received signal vector $\mathbf{x}(k) = [x_1(k)\ x_2(k)\ \cdots\ x_L(k)]^T$ can be expressed as

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

where we have $\mathbf{n}(k) = [n_1(k)\ n_2(k)\ \cdots\ n_L(k)]^T$, the system matrix is given by

$$\mathbf{P} = [A_1\mathbf{s}_1\ A_2\mathbf{s}_2 \cdots A_S\mathbf{s}_S] \quad (3)$$

and the steering vector for source i is

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

while the transmitted BPSK symbol vector is $\mathbf{b}(k) = [b_1(k)\ b_2(k)\ \cdots\ b_S(k)]^T$.

Although we assume a uniformly spaced linear antenna array, the results can be extended to other antenna array structures. In fact, our discussions are applicable to the generic multiple-input–multiple-output (MIMO) communication system [37], [38], where the i,jth element of the system matrix $\mathbf{P}$ represents the channel coefficient connecting the jth transmit antenna to the ith receive antenna. An implicit assumption for the signal model (1) is that the desired user and interfering signals are symbol synchronised. For the downlink scenario, synchronous transmission of the users is guaranteed. By contrast, in an uplink scenario, the differently delayed asynchronous signals of the users are no longer automatically synchronized. However, the quasi-synchronous operation of the system may be achieved with the aid of adaptive timing advance control as in the global system of mobile (GSM) [41]. The GSM system has a timing-advance control accuracy of 0.25-bit duration. Because synchronous systems perform better than their asynchronous counterparts, the third-generation partnership research consortium (3GPP) is also considering the employment of timing-advance control in next-generation systems.

Classically, a linear beamforming receiver is adopted to detect the desired user signal [39], [40]. The output of the linear beamformer is defined by

$$y_{\text{Lin}}(k) = \mathbf{w}^H\mathbf{x}(k) \quad (5)$$

and the associated decision is given by

$$\hat{b}_1(k) = \text{sgn}\left(\Re\left[y_{\text{Lin}}(k)\right]\right) = \begin{cases} +1, & \Re\left[y_{\text{Lin}}(k)\right] \ge 0 \\ -1, & \Re\left[y_{\text{Lin}}(k)\right] < 0 \end{cases} \quad (6)$$

where $\mathbf{w} = [w_1\ w_2\ \cdots\ w_L]^T$ denotes the linear beamformer's weight vector and $\Re[\bullet]$ the real part. Traditionally, the weight vector of the linear beamformer (5) is set to the (linear) minimum mean square error (L-MMSE) solution [39], [40]. The L-MMSE solution is based on the following consideration. An antenna array of L elements can place $L-1$ nulls. Thus, the system can support up to $S = L$ users. If the number of users S is larger than the number of array elements L, the system is called overloaded or rank-deficient. In our previous work [42], we have shown that the optimal weight vector designed for the linear beamformer is the (linear) minimum bit error rate (L-MBER) solution, which directly minimizes the error probability or bit error rate (BER) of the linear beamformer (5). The L-MBER beamforming outperforms the L-MMSE one significantly, particularly for overloaded systems. For this reason, we will use the L-MBER beamforming solution as a benchmark, rather than the L-MMSE one, in the evaluation of the proposed symmetric RBF detector. The L-MBER design is optimal for the *linear* beamforming. The optimal channel matched solution for the multiple-antenna-aided beamforming detector, however, is *nonlinear* [33], [34].

Let us denote the $N_b = 2^S$ legitimate combinations of $\mathbf{b}(k)$ as $\mathbf{b}_q, 1 \le q \le N_b$. Denote furthermore the first element of $\mathbf{b}_q$, corresponding to the desired user, as $b_{q,1}$. The noiseless channel output $\bar{\mathbf{x}}(k)$ only takes values from the signal state set

$$\mathcal{X} \triangleq \{\bar{\mathbf{x}}_q = \mathbf{P}\mathbf{b}_q,\ 1 \le q \le N_b\}. \tag{7}$$

The signal state set $\mathcal{X}$ can be divided into two subsets conditioned on the value of $b_1(k)$ as follows:

$$\mathcal{X}^{(\pm)} \triangleq \{\bar{\mathbf{x}}_i \in \mathcal{X}, 1 \le i \le N_{sb} : b_1(k) = \pm 1\} \tag{8}$$

where the size of the sets $\mathcal{X}^{(+)}$ and $\mathcal{X}^{(-)}$ is $N_{sb} = N_b/2 = 2^{S-1}$. Denote the conditional probabilities of receiving $\mathbf{x}(k)$ given $b_1(k) = \pm 1$ as $p_{\pm}(\mathbf{x}(k)) = p(\mathbf{x}(k)|b_1(k) = \pm 1)$. According to Bayes decision theory [43], the optimal detection strategy should be

$$\hat{b}_1(k) = \begin{cases} +1, & \text{if } p_+(\mathbf{x}(k)) \ge p_-(\mathbf{x}(k)) \\ -1, & \text{if } p_+(\mathbf{x}(k)) < p_-(\mathbf{x}(k)). \end{cases} \tag{9}$$

If we introduce the following real-valued Bayesian decision variable

$$y_{\mathrm{Bay}}(k) = f_{\mathrm{Bay}}(\mathbf{x}(k)) \triangleq \frac{1}{2}p_+(\mathbf{x}(k)) - \frac{1}{2}p_-(\mathbf{x}(k)) \tag{10}$$

the optimal Bayesian detection rule (9) is equivalent to $\hat{b}_1(k) = \mathrm{sgn}(y_{\mathrm{Bay}}(k))$.

The decision variable (10) of the optimal Bayesian detector is readily expressed as [33], [34]

$$y_{\mathrm{Bay}}(k) = \sum_{q=1}^{N_b} \mathrm{sgn}(b_{q,1})\beta_q e^{-\frac{\|\mathbf{x}(k)-\bar{\mathbf{x}}_q\|^2}{2\sigma_n^2}} \tag{11}$$

where β_q is proportional to the *a priori* probability of $\bar{\mathbf{x}}_q$. Because in our case, all the $\bar{\mathbf{x}}_q$ are equiprobable, we have $\beta_q = \beta = 1/N_b(2\pi\sigma_n^2)^L$. It can readily be shown that the two subsets $\mathcal{X}^{(+)}$ and $\mathcal{X}^{(-)}$ are symmetric with respect to each other [35]. Hence, provided that appropriate indexing is used, for any signal state $\bar{\mathbf{x}}_i^{(+)} \in \mathcal{X}^{(+)}$, there exists a signal state $\bar{\mathbf{x}}_i^{(-)} \in \mathcal{X}^{(-)}$, so that

$$\bar{\mathbf{x}}_i^{(-)} = -\bar{\mathbf{x}}_i^{(+)}. \tag{12}$$

Given this symmetry, the optimal Bayesian detector (11) can be rewritten as

$$y_{\mathrm{Bay}}(k) = \sum_{q=1}^{N_{sb}} \beta_q \left(e^{-\frac{\left\|\mathbf{x}(k)-\bar{\mathbf{x}}_q^{(+)}\right\|^2}{2\sigma_n^2}} - e^{-\frac{\left\|\mathbf{x}(k)+\bar{\mathbf{x}}_q^{(+)}\right\|^2}{2\sigma_n^2}} \right) \tag{13}$$

where $\bar{\mathbf{x}}_q^{(+)} \in \mathcal{X}^{(+)}$. The Bayesian detector has *odd* symmetry, as $f_{\mathrm{Bay}}(-\mathbf{x}(k)) = -f_{\mathrm{Bay}}(\mathbf{x}(k))$.

If the system matrix $\mathbf{P}$ of (3) is known, the signal state subset $\mathcal{X}^{(+)}$ can be computed and the Bayesian detection solution is specified. For the multiple-antenna-aided beamformer in downlink, however, the receiver only has access to the training data $D_K = \{\mathbf{x}(k), b_1(k)\}_{k=1}^K$, where K is the number of training symbols and $\{b_1(k)\}$ are the desired user's data. However, the receiver does not have access to the interfering users' data $\{b_i(k)\}, i \ne 1$. Thus, estimating the system matrix $\mathbf{P}$ is a challenging task. In our previous work [33] and [34], standard kernel-based classifiers or detectors were constructed directly using the noisy training data set D_K to approximate the optimal Bayesian solution. As discussed in Section I, the inherent symmetry of the Bayesian detector in (13) is hard to learn by a black-box RBF or kernel classifier using noisy data. In this paper, we propose a novel symmetric RBF classifier which renders realization of the symmetric Bayesian detection solution easier.

Before discussing the training algorithm for the generic two-class symmetric RBF classifier, we point out that more general symmetric structures also exist for the Bayesian detectors of communication systems that employ multiple-bits per symbol modulation schemes. For the 4-QAM system, for instance, the data symbol takes value from the 4-QAM symbol set, namely, $b_i(k) \in \{\pm 1 \pm j\}$, and the signal state set $\mathcal{X}$ can be divided into the four subsets $\mathcal{X}^{(\pm 1 \pm j)}$ depending on the value of $b_1(k)$. It is easily verified that the four state subsets satisfy the following symmetric relationship: $\mathcal{X}^{(-1+j)} = +j \cdot \mathcal{X}^{(+1+j)}$, $\mathcal{X}^{(-1-j)} = -1 \cdot \mathcal{X}^{(+1+j)}$ and $\mathcal{X}^{(+1-j)} = -j \cdot \mathcal{X}^{(+1+j)}$. This symmetric structure may be exploited to modify the four quadratic-amplitude modulation (4-QAM) RBF detector of [44]. A more general symmetric structure also exists for the generic high-order QAM system; see [45]. The derivation of the generic training algorithm for the high-order QAM RBF detector is much more complex and is beyond the scope of this work.

III. Symmetric RBF Classifier

Consider the problem of training a two-class RBF classifier $y_{\mathrm{RBF}}(\mathbf{x}) : \mathcal{C}^L \to \{1, -1\}$ based on a training data set $D_K = \{\mathbf{x}(k), d(k)\}_{k=1}^K$, where $d(k) \in \{1, -1\}$ denotes the class type for each complex-valued data sample $\mathbf{x}(k) \in \mathcal{C}^L$. We adopt the RBF classifier of the form

$$\hat{d}(k) = \mathrm{sgn}(y_{\mathrm{RBF}}(k))$$

with

$$y_{\mathrm{RBF}}(k) = f_{\mathrm{RBF}}(\mathbf{x}(k)) \triangleq \sum_{i=1}^{M} \theta_i \phi_i(\mathbf{x}(k)) \tag{14}$$

where $\hat{d}(k)$ is the estimated class label for $\mathbf{x}(k)$, $\phi_i(\bullet)$ denotes the response of the classifier's RBF bases, θ_i are the classifier's coefficients, and M is the number of RBF units. In contrast to the standard RBF classifier, here, we propose to adopt the following symmetric RBF unit:

$$\phi_i(\mathbf{x}) \triangleq \varphi(\mathbf{x}; \mathbf{c}_i, \rho^2) - \varphi(\mathbf{x}; -\mathbf{c}_i, \rho^2) \tag{15}$$

where $\mathbf{c}_i$ is the RBF center, ρ^2 is the RBF variance, and $\varphi(\bullet)$ is the classic RBF function. In this paper, we adopt the Gaussian RBF function of

$$\varphi(\mathbf{x};\mathbf{c}_i,\rho^2)=e^{-\frac{\|\mathbf{x}-\mathbf{c}_i\|^2}{2\rho^2}}. \tag{16}$$

Other RBF or kernel functions can also be used here. It is worth emphasizing that, although we derive the symmetric RBF formulation directly through the observation of the underlying symmetric Bayesian detection solution, the proposed symmetric RBF detector can also be derived analytically by imposing the odd symmetry constraint on the standard kernel formulation, just as in the regression case [24].

Because the symmetric RBF formulation (14) is "identical" in form to the standard RBF formulation, most of the existing sparse RBF or kernel techniques can be applied. Our previous experience with standard sparse kernel modeling suggests that the OFS procedure based on the FRCSM [33], [34] compares favorably with many other existing sparse kernel methods, such as the SVM techniques, in terms of efficiency of the construction process and the sparsity of the constructed model. For practical purpose, it is critical to derive an RBF or kernel detector as sparse as possible, because the detection complexity scales with the size of the RBF classifier. We apply the OFS procedure based on the FRCSM to construct a sparse symmetric RBF classifier using the training data set D_k. Note that the objective of training a classifier is to achieve maximum classification discriminative power, and Fisher ratio is a measure of discriminative power or class separability [43].

Consider every training data point $\mathbf{x}(i)$ as a candidate RBF center. Hence, we have $M=K$ in the RBF model of (14) and $\mathbf{c}_i=\mathbf{x}(i)$ for $1\le i\le K$, and the RBF variance is set to ρ^2. Let us now define $\varepsilon(i)=d(i)-y_{\mathrm{RBF}}(i)$ as the modeling residual sequence. Then, the model (14) defined over the training data set D_K can be written in a matrix form as

$$\mathbf{d}=\mathbf{\Phi}\boldsymbol{\theta}+\boldsymbol{\varepsilon} \tag{17}$$

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

$$\mathbf{\Phi}=[\boldsymbol{\phi}_1\ \boldsymbol{\phi}_2\ \cdots\ \boldsymbol{\phi}_M]\in\mathcal{R}^{K\times M} \tag{18}$$

is the regression matrix with the column vectors $\boldsymbol{\phi}_i=[\phi_i(\mathbf{x}(1))\ \phi_i(\mathbf{x}(2))\ \cdots\ \phi_i(\mathbf{x}(K))]^T$, $1\le i\le M$. Let an orthogonal decomposition of $\mathbf{\Phi}$ be $\mathbf{\Phi}=\mathbf{\Omega}\mathbf{A}$, where we have

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

and

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

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

$$\mathbf{d}=\mathbf{\Omega}\boldsymbol{\gamma}+\boldsymbol{\varepsilon} \tag{21}$$

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

A sparse M_{spa}-term classifier can be selected by incrementally maximizing the FRCSM using the OFS procedure, as in [18], [33], and [34]. Define the two-class sets $\mathbf{X}_\pm=\{\mathbf{x}(k): d(k)=\pm1\}$, and let the number of points in $\mathbf{X}_\pm$ be $K_\pm$, respectively, with $K_++K_-=K$. The means and variances of the training samples belonging to class $\mathbf{X}_+$ and class $\mathbf{X}_-$ in the direction of the basis $\boldsymbol{\omega}_l$ are given by

$$m_{+,l}=\frac{1}{K_+}\sum_{k=1}^{K}\delta\left(d(k)-1\right)\omega_{k,l} \tag{22}$$

$$\sigma_{+,l}^2=\frac{1}{K_+}\sum_{k=1}^{K}\delta\left(d(k)-1\right)(\omega_{k,l}-m_{+,l})^2 \tag{23}$$

and

$$m_{-,l}=\frac{1}{K_-}\sum_{k=1}^{K}\delta\left(d(k)+1\right)\omega_{k,l} \tag{24}$$

$$\sigma_{-,l}^2=\frac{1}{K_-}\sum_{k=1}^{K}\delta\left(d(k)+1\right)(\omega_{k,l}-m_{-,l})^2 \tag{25}$$

respectively, where

$$\delta(x)=\begin{cases}1, & x=0\\ 0, & x\ne 0.\end{cases} \tag{26}$$

The Fisher ratio is defined as the ratio of the interclass difference and the intraclass spread encountered in the direction of $\boldsymbol{\omega}_l$, which is given by [43]

$$F_l=\frac{(m_{+,l}-m_{-,l})^2}{\sigma_{+,l}^2+\sigma_{-,l}^2}. \tag{27}$$

Based on this FRCSM, significant RBF terms can be selected with the aid of an OFS procedure. At the lth stage, a candidate term is chosen as the lth RBF term in the selected model, if it produces the largest F_l among the $M-l+1$ candidate terms $\boldsymbol{\omega}_i$. The procedure is terminated with a sparse M_{spa}-term model, when we have

$$\frac{F_{M_{\mathrm{spa}}}}{\sum_{l=1}^{M_{\mathrm{spa}}}F_l}<\xi \tag{28}$$

where the threshold ξ determines the sparsity level of the model selected. The appropriate value for ξ depends on the application concerned, and it must be determined empirically. The least squares solution for the corresponding sparse model weight vector $\boldsymbol{\theta}_{M_{\mathrm{spa}}}=[\theta_1\ \theta_2\ \cdots\ \theta_{M_{\mathrm{spa}}}]^T$ is readily available, given the least squares solution of $\boldsymbol{\gamma}_{M_{\mathrm{spa}}}=[\gamma_1\ \gamma_2\ \cdots\ \gamma_{M_{\mathrm{spa}}}]^T$. The detailed construction algorithm based on the Gram–Schmidt orthogonalization [14] is summarized in the following.

OFS based on the FRCSM

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

$$\alpha_{i,l}^{(q)} = \begin{cases} \dfrac{\boldsymbol{\omega}_i^T \boldsymbol{\phi}_q}{\boldsymbol{\omega}_i^T \boldsymbol{\omega}_i}, & 1 \leq i < l \\ 1, & i = l \end{cases}$$

$$\boldsymbol{\omega}_l^{(q)} = \begin{cases} \boldsymbol{\phi}_q, & l = 1 \\ \boldsymbol{\phi}_q - \sum_{i=1}^{l-1} \alpha_{i,l}^{(q)} \boldsymbol{\omega}_i, & l > 1 \end{cases}$$

$$m_{+,l}^{(q)} = \frac{1}{K_+} \sum_{k=1}^{K} \delta\left(d(k) - 1\right) \omega_{k,l}^{(q)}$$

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

$$m_{-,l}^{(q)} = \frac{1}{K_-} \sum_{k=1}^{K} \delta\left(d(k) + 1\right) \omega_{k,l}^{(q)}$$

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

$$F_l^{(q)} = \frac{\left(m_{+,l}^{(q)} - m_{-,l}^{(q)}\right)^2}{\left(\sigma_{+,l}^{(q)}\right)^2 + \left(\sigma_{-,l}^{(q)}\right)^2}.$$

2) Find

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

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

$$\boldsymbol{\omega}_l = \boldsymbol{\omega}_l^{(q_l)} = \begin{cases} \boldsymbol{\phi}_{q_l}, & l = 1 \\ \boldsymbol{\phi}_{q_l} - \sum_{i=1}^{l-1} \alpha_{i,l} \boldsymbol{\omega}_i, & l > 1. \end{cases}$$

3) The procedure is monitored and terminated at the index value $l = M_{\text{spa}}$, when, for example, the condition (28) 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 $\boldsymbol{\omega}_l^{(q)}$ has too low energy, i.e., $\left(\boldsymbol{\omega}_l^{(q)}\right)^T \boldsymbol{\omega}_l^{(q)}$ is near zero, it will not be considered. The least squares solution for the weight γ_l is simply

$$\gamma_l = \frac{\boldsymbol{\omega}_l^T \mathbf{d}}{\boldsymbol{\omega}_l^T \boldsymbol{\omega}_l}. \tag{29}$$

Instead of using the condition (28) to terminate the OFS procedure, which requires us to specify the threshold value ξ, the so-called cross-validation procedure can be used to decide when to stop the selection procedure. Automatic termination criteria such as the information-based criteria and optimal experimental design criteria of [19] may also be used.

In our particular application, the number of users is usually known, and therefore the number of the subset signal states N_{sb} is known. Thus, we may simply set $M_{\text{spa}} \leq N_{sb}$. The training set should be sufficiently rich, i.e., containing all possible sequences $\mathbf{b}_q$, $1 \leq q \leq N_b$. It is usually adequate to set the size of the training set K in the range of 10 N_b to 20 N_b. The RBF variance ρ^2 is not provided by the construction algorithm, but it may be estimated based on cross validation. Our experience suggests that the symmetric RBF classifier is not sensitive to the value of ρ^2 used, and usually there exists a range of ρ^2 values which enables the sparse symmetric RBF classifier to approach the optimal Bayesian performance. This will be further illustrated in our simulation study. This robustness to the value of ρ^2 inherently is a consequence of the Bayesian detector's robustness to the noise variance σ_n^2 used [46].

TABLE I
LOCATIONS OF USERS IN TERMS OF ANGLE OF ARRIVAL (AOA) FOR THE TWO-ELEMENT ANTENNA ARRAY SYSTEM

user i	1	2	3	4
AOA θ_i	0°	20°	-30°	-45°

IV. SIMULATION STUDY

Three simulated multiple-antenna-aided beamforming systems were used to demonstrate the efficiency of the proposed symmetric RBF classifier. The antenna array element spacing was half the wavelength. The simulated channel conditions were $A_i = 1 + j0, 1 \leq i \leq S$. The desired user and all the interfering users had equal signal power, and therefore, we had $\text{SIR}_i = 0$ dB for $2 \leq i \leq S$.

Example 1: The example consisted of four BPSK signal sources and a two-element antenna array. The user angular locations are summarized in Table I. Fig. 1 portrays the BER performance of both the theoretical L-MBER beamformer and the Bayesian detector for the desired user 1. For each SNR value, $K = 500$ training data $D_K = \{\mathbf{x}(k), b_1(k)\}_{k=1}^K$ were used to construct the symmetric RBF classifier using the FRCSM-based OFS algorithm as outlined in Section III. The RBF variance was chosen to be $\rho^2 = 3\sigma_n^2$. As the size of the Bayesian detector was $N_{sb} = 8$, we terminated the RBF classifier construction at $M_{\text{spa}} = 8$. The BER of the eight-term symmetric RBF detector is also depicted in Fig. 1. It can be seen from Fig. 1 that the symmetric RBF detector is capable of closely approaching the optimal Bayesian performance even when the number of symmetric kernels is no larger than that of the Bayesian detector, and hence, outperforms the black-box kernel methods of [33], [34].

When ignoring the symmetry, a standard kernel detector would typically require more kernels than the number of Bayesian kernels N_b and yet there would be a larger performance gap between the kernel detector and the Bayesian one. We reconfirm this previous observation by constructing the standard SVM classifier [2] with the Gaussian kernel function for this example using the same 500-sample training data set. The size of the SVM detector constructed ranged from 20 to 32 support vectors (SVs) depending on the SNR value, and the typical SVM detector contained 26 SVs. The optimal kernel variance ρ^2 was determined using cross validation and its value ranged from σ_n^2 to $3\sigma_n^2$. The BER performance of the constructed SVM detector is also plotted in Fig. 1.

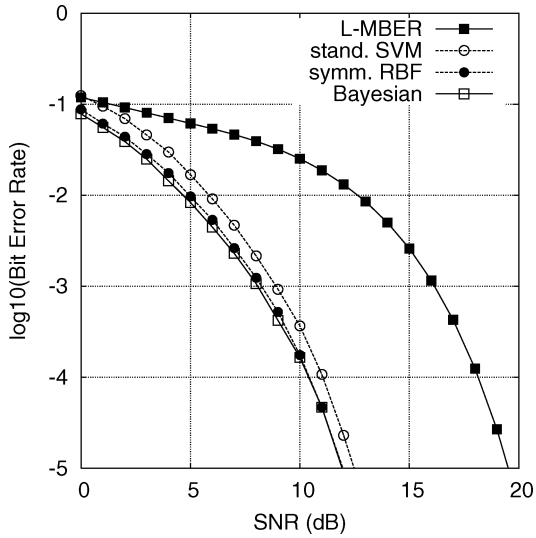

Fig. 1. Desired-user's BER performance in the context of four detectors for the two-element antenna array supporting four users. The symmetric RBF classifier, constructed from 500 noisy training samples using the FRCSM-based OFS, has $M_{\rm spa} = 8$ symmetric RBF units and the RBF variance of $\rho^2 = 3\sigma_n^2$. The standard SVM classifier, constructed from 500 noisy training samples, typically has 26 support vectors and a kernel variance in the range of σ_n^2 to $3\sigma_n^2$.

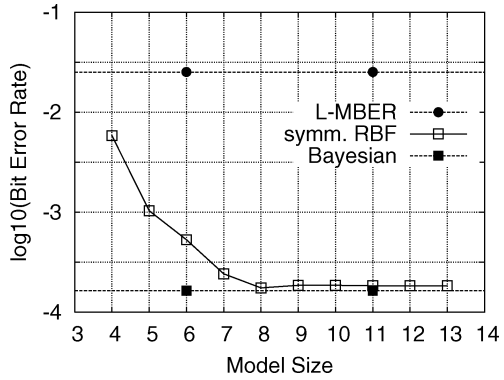

Fig. 2. Influence of the classifier's size on the BER performance of the symmetric RBF classifier for the two-element antenna array supporting four users. We used SNR = 10 dB, the training data length of $K = 500$, and the RBF variance ρ^2 was varied depending on the model size.

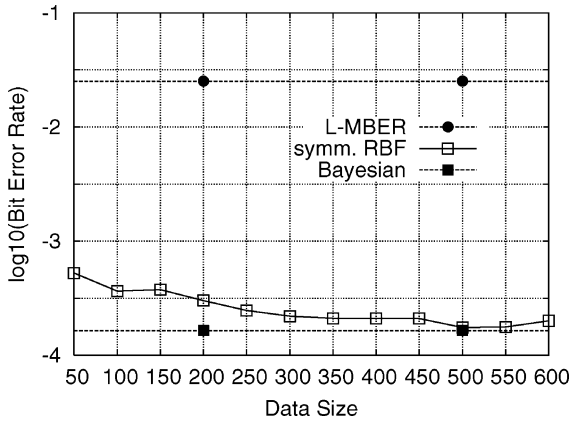

Fig. 3. Influence of the training data length on the BER performance of the symmetric RBF classifier for the two-element antenna array supporting four users. We used SNR = 10 dB, the RBF model size of $M_{\rm spa} = 8$, and the RBF variance of $\rho^2 = 3\sigma_n^2$.

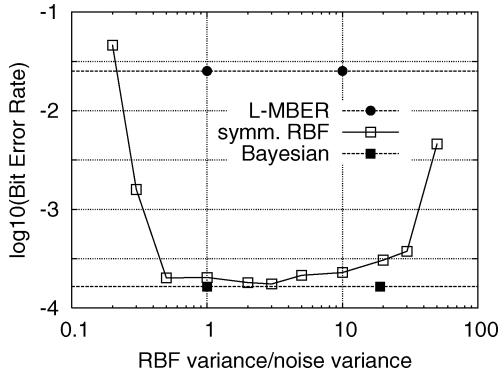

Fig. 4. Influence of the RBF variance on theBER performance of the symmetric RBF classifier for the two-element antenna array supporting four users. We used SNR = 10 dB, the training data length of $K = 500$, and the RBF model size of $M_{\rm spa} = 8$.

TABLE II
LOCATIONS OF USERS IN TERMS OF ANGLE OF ARRIVAL (AOA) FOR THE THREE-ELEMENT ANTENNA ARRAY SYSTEM

user i	1	2	3	4	5
AOA θ_i	0°	10°	-17°	15°	20°

The properties of the proposed FRCSM-based OFS invoked for constructing the symmetric RBF detector were studied. First, the influence of the model size $M_{\rm spa}$ on the RBF classifier's performance was investigated. Given SNR = 10 dB and a training data length of $K = 500$, Fig. 2 shows the performance of the symmetric RBF classifier as a function of the model size $M_{\rm spa}$. The RBF variance ρ^2 was tuned according to the model size $M_{\rm spa}$. Appropriate values were found to be $\rho^2 = 16\sigma_n^2$, $14\sigma_n^2$, $8\sigma_n^2$, and $6\sigma_n^2$ for $M_{\rm spa} = 4, 5, 6$, and 7, respectively, and $3\sigma_n^2$ for $M_{\rm spa} \geq 8$. Next, the influence of the training data length K was investigated. Given SNR = 10 dB, an RBF variance of $\rho^2 = 3\sigma_n^2$, and an RBF model size of $M_{\rm spa} = 8$, Fig. 3 plots the performance of the symmetric RBF detector as a function of the training data length K. The influence of the RBF variance on the performance of the symmetric RBF classifier was also investigated. Given SNR = 10 dB, a training data length $K = 500$, and an RBF model size $M_{\rm spa} = 8$, Fig. 4 illustrates the BER of the symmetric RBF detector as a function of the RBF variance. The result of Fig. 4 confirms that there exists a large range of values for ρ^2, which allow the spare symmetric RBF detector to approach the optimal Bayesian performance.

Example 2: A three-element linear antenna array system was designed for supporting five BPSK users. Table II lists the user angular locations. Fig. 5 plots the BER performance of both the L-MBER beamformer and the Bayesian detector. The size of the Bayesian solution for this example is specified by $N_{sb} = 16$. Given each SNR value, $K = 600$ training samples were used to construct the symmetric RBF classifier with a model size $M_{\rm spa} = 16$ and an RBF variance of $\rho^2 = 3\sigma_n^2$, using the FRCSM-based OFS. The BER of the constructed symmetric RBF detector is depicted in Fig. 5, where it can be seen that the 16-term symmetric RBF detector closely approached the performance of the optimal Bayesian solution. As a comparison, the standard SVM detector was also trained using the same 600-sample training data set. The size of the SVM detector constructed ranged from 40 to 60 SVs, and the value of the kernel variance ρ^2, determined using cross validation, was in the range of σ_n^2 to $6\sigma_n^2$. The BER of the constructed SVM detector is also shown in Fig. 5.

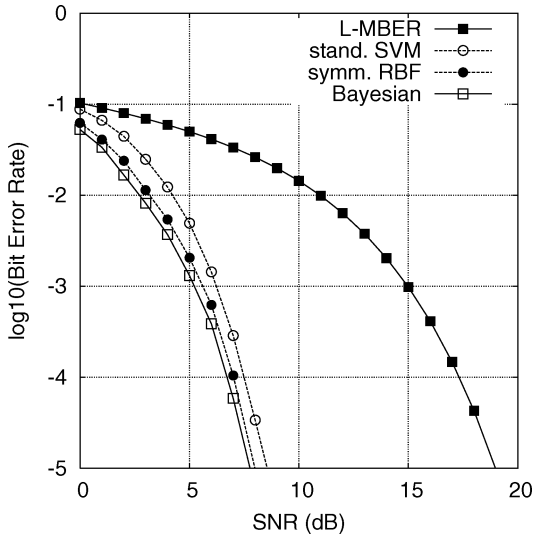

Fig. 5. Desired-user's BER performance in the context of four detectors for the three-element antenna array supporting five users. The symmetric RBF classifier, constructed from 600 noisy training samples using the FRCSM-based OFS, has $M_{\rm spa} = 16$ symmetry RBF units and the RBF variance of $\rho^2 = 3\sigma_n^2$. The standard SVM classifier, constructed from 600 noisy training samples, has the number of support vectors in the range of 40 to 60 and a kernel variance in the range of σ_n^2 to $6\sigma_n^2$.

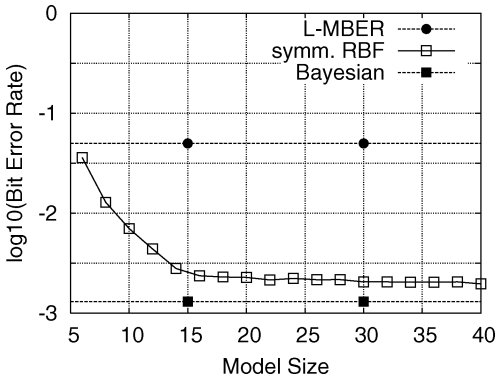

Fig. 6. Influence of the classifier's size on the BER performance of the symmetric RBF classifier for the three-element antenna array supporting five users. We used SNR = 5 dB, the training data length of $K = 600$, and the RBF variance ρ^2 was varied depending on the model size.

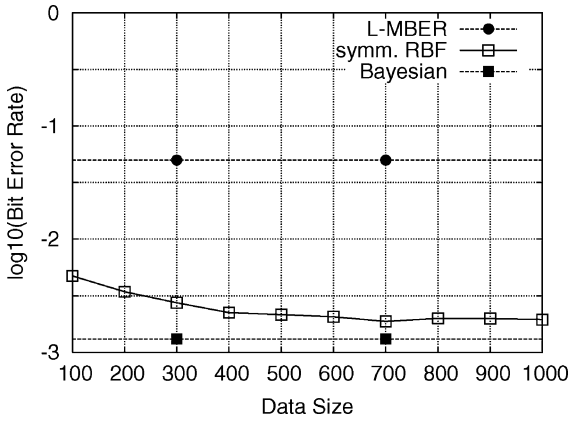

Fig. 7. Influence of the training data length on the BER performance of the symmetric RBF classifier for the three-element antenna array supporting five users. We used SNR = 5 dB, the RBF model size of $M_{\rm spa} = 16$, and the RBF variance of $\rho^2 = 3\sigma_n^2$.

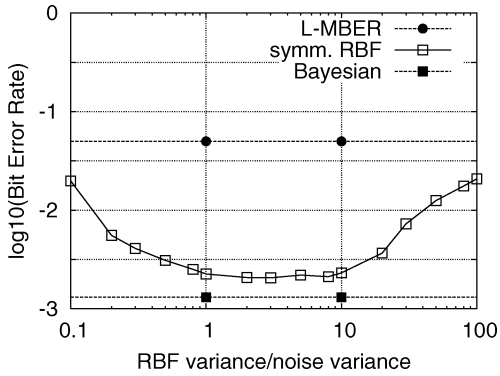

Fig. 8. Influence of the RBF variance on the BER performance of the symmetric RBF classifier for the three-element antenna array supporting five users. We used SNR = 5 dB, the training data length of $K = 600$, and the RBF model size of $M_{\rm spa} = 16$.

TABLE III
LOCATIONS OF USERS IN TERMS OF ANGLE OF ARRIVAL (AOA) FOR THE FOUR-ELEMENT ANTENNA ARRAY SYSTEM

user i	1	2	3	4	5	6	7
AOA θ_i	0°	10°	-12°	15°	20°	-18°	17°

The influence of the model size $M_{\rm spa}$ to the constructed RBF classifier's performance was studied next. Given SNR = 5 dB and the training data length $K = 600$, Fig. 6 illustrates the BER of the symmetric RBF classifier as a function of the constructed model size $M_{\rm spa}$. The RBF variance ρ^2 was tuned according to the model size $M_{\rm spa}$, and was in the range of $3\sigma_n^2$ to $5\sigma_n^2$. The influence of the training data length K was then investigated. Given SNR = 5 dB, RBF variance $\rho^2 = 3\sigma_n^2$, and RBF model size $M_{\rm spa} = 16$, Fig. 7 plots the BER of the symmetric RBF detector as a function of the training data length K. The influence of the RBF variance to the performance of the symmetric RBF classifier was also investigated. Given SNR = 5 dB, the training data length $K = 600$, and RBF model size $M_{\rm spa} = 16$, Fig. 8 depicts the BER of the symmetric RBF detector as a function of the RBF variance. The result of Fig. 8 again confirms that there exists a large range of values for ρ^2 with which the spare symmetric RBF detector can closely approximate the Bayesian performance.

Example 3: A four-element linear antenna array system was used to support seven BPSK users. The users' angular locations are summarized in Table III. Fig. 9 depicts the BER performance of both the L-MBER beamformer and the optimal Bayesian detector. The size of the Bayesian solution for this example is specified by $N_{sb} = 64$. At each SNR value, $K = 1000$ training samples were used to construct the symmetric RBF classifier with a model size $M_{\rm spa} = 60$ and an RBF variance of $\rho^2 = 3\sigma_n^2$, using the FRCSM-based OFS. The BER performance of the resulting symmetric RBF detector is also shown in Fig. 9 in comparison to that of the optimal Bayesian solution.

We found that a standard kernel classifier constructed from the same 1000-sample training data set would require a significantly larger model size than the $N_b = 128$ number of signal states for this example. In our simulation, the SVM detector required several hundreds of SVs. Moreover, the performance of

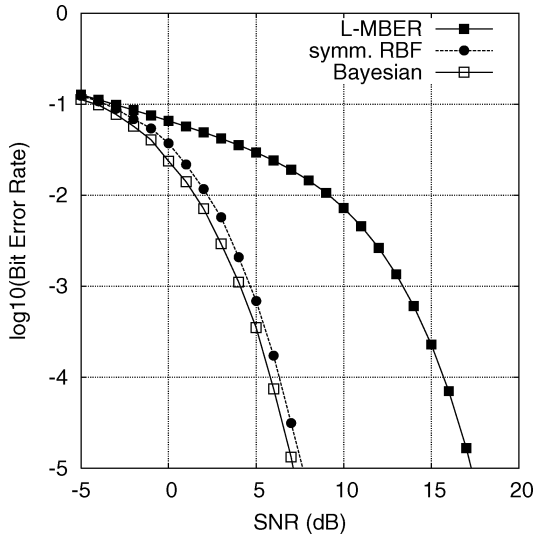

Fig. 9. Desired-user's BER performance in the context of three detectors for the four-element antenna array supporting seven users. The symmetric RBF classifier, constructed from 1000 noisy training samples using the FRCSM-based OFS, has $M_{\rm spa} = 60$ symmetric RBF units and the RBF variance of $\rho^2 = 3\sigma_n^2$.

this standard SVM detector was poorer than that of the symmetric RBF detector.[1]

V. Conclusion

A novel symmetric RBF classifier has been proposed for nonlinear detection which is capable of substantially outperforming previous solutions in the extremely challenging scenario of supporting almost twice as many users as the number of antenna elements in multiple-antenna-aided communication systems. The orthogonal forward selection procedure based on the Fisher ratio of class separability measure provides a fast and efficient means of constructing a sparse symmetric RBF detector from the noisy training data, which is capable of approaching the optimal Bayesian detection performance. The proposed solution is capable of providing an SNR gain in excess of 8 dB against the powerful linear minimum BER benchmark, when supporting four users with the aid of two receive antennas or seven users employing four receive antenna elements. Although we have presented this sparse symmetric RBF classifier in the context of nonlinear detection in wireless communication systems, it is generically applicable to any classification problem exhibiting a similar symmetry.

The orthogonal forward selection algorithm based on the Fisher ratio of class separability measure provides an efficient procedure for constructing the optimal minimum symbol–error-rate symmetric RBF detector for the stationary or slow flat fading systems. We are currently conducting the research to develop a stochastic adaptive near-minimum symbol–error-rate algorithm for sample-by-sample adaptation of the symmetric RBF detector in order to extend its application to fast fading systems.

[1]It was also computationally very expensive to compute the BER of the SVM detector for high SNR values using simulation. This is because in order to guarantee the estimation accuracy of the BER simulation we require that at least 500 errors occur in the simulation. For the BER level at 10^{-5}, this will require at least 5×10^7 data symbols. Computing the BER of the SVM detector with several hundreds of kernels at this BER level is computationally too costly.

References

[1] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
[2] S. Gunn, "Support vector machines for classification and regression," ISIS Res. Group, Dept. Electron. Comput. Sci., Univ. Southampton, Southampton, U.K., May 1998, Tech. Rep..
[3] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.
[4] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.
[5] B. Schölkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press, 2002.
[6] P. Vincent and Y. Bengio, "Kernel matching pursuit," *Mach. Learn.*, vol. 48, no. 1, pp. 165–187, 2002.
[7] D. Decoste and B. Schölkopf, "Training invariant support vector machines," *Mach. Learn.*, vol. 46, no. 1, pp. 161–190, 2002.
[8] G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan, "Learning the kernel matrix with semidefinite programming," *J. Mach. Learn. Res.*, vol. 5, pp. 27–72, 2004.
[9] Y.-J. Lee and S.-Y. Huang, "Reduced support vector machines: A statistical theory," *IEEE Trans. Neural Netw.*, vol. 18, no. 1, pp. 1–13, Jan. 2007.
[10] L. Jiao, L. Bo, and L. Wang, "Fast sparse approximation for least squares support vector machine," *IEEE Trans. Neural Netw.*, vol. 18, no. 3, pp. 685–697, May 2007.
[11] E. Romero and D. Toppo, "Comparing support vector machines and feedforward neural networks with similar hidden-layer weights," *IEEE Trans. Neural Netw.*, vol. 18, no. 3, pp. 959–963, May 2007.
[12] X. Xu, D. Hu, and X. Lu, "Kernel-based least squares policy iteration for reinforcement learning," *IEEE Trans. Neural Netw.*, vol. 18, no. 4, pp. 973–992, Jul. 2007.
[13] G. L. Wang, Y. F. Li, and D. X. Bi, "Support vector networks in adaptive friction compensation," *IEEE Trans. Neural Netw.*, vol. 18, no. 4, pp. 1209–1219, Jul. 2007.
[14] 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.
[15] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Netw.*, vol. 2, no. 2, pp. 302–309, Mar. 1991.
[16] S. Chen, Y. Wu, and B. L. Luk, "Combined genetic algorithm optimization and regularized orthogonal least squares learning for radial basis function networks," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1239–1243, Sep. 1999.
[17] S. A. Billings and K. L. Lee, "Nonlinear Fisher discriminant analysis using a minimum squared error cost function and the orthogonal least squares algorithm," *Neural Netw.*, vol. 15, no. 2, pp. 263–270, 2002.
[18] K. Z. Mao, "RBF neural network center selection based on Fisher ratio class separability measure," *IEEE Trans. Neural Netw.*, vol. 13, no. 5, pp. 1211–1217, Sep. 2002.
[19] S. Chen, X. Hong, and C. J. Harris, "Sparse kernel regression modeling using combined locally regularized orthogonal least squares and D-optimality experimental design," *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 1029–1036, Jun. 2003.
[20] S. Chen, X. Hong, C. J. Harris, and P. M. Sharkey, "Sparse modeling using orthogonal forward regression with PRESS statistic and regularization," *IEEE Trans. Syst. Man Cybern. B, Cybern.*, vol. 34, no. 2, pp. 898–911, Apr. 2004.
[21] X. Hong, S. Chen, and C. J. Harris, "Fast kernel classifier construction using orthogonal forward selection to minimize leave-one-out misclassification rate," in *Proc. Int. Conf. Intell. Comput.*, Kunming, China, Aug. 16–19, 2006, pp. 106–114.
[22] X. Hong, S. Chen, and C. J. Harris, "A kernel-based two-class classifier for imbalanced data sets," *IEEE Trans. Neural Netw.*, vol. 18, no. 1, pp. 28–41, Jan. 2007.
[23] 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," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, pp. 026701-1–026701-11, 2004.
[24] 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 Control/Eur. Control Conf.*, Dec. 12–15, 2005, pp. 5716–5721.
[25] S. Chen and B. Mulgrew, "Overcoming co-channel interference using an adaptive radial basis function equalizer," *Signal Process.*, vol. 28, no. 1, pp. 91–107, 1992.
[26] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," *IEEE Trans. Neural Netw.*, vol. 4, no. 4, pp. 570–579, Jul. 1993.

[27] F. Albu and D. Martinez, "The application of support vector machines with Gaussian kernels for overcoming co-channel interference," in *Proc. 9th IEEE Int. Workshop Neural Netw. Signal Process.*, Madison, WI, Aug. 23–25, 1999, pp. 49–57.
[28] D. J. Sebald and J. A. Bucklew, "Support vector machine techniques for nonlinear equalization," *IEEE Trans. Signal Process.*, vol. 48, no. 11, pp. 3217–3226, Nov. 2000.
[29] S. Chen, A. K. Samingan, and L. Hanzo, "Support vector machine multiuser receiver for DS-CDMA signals in multipath channels," *IEEE Trans. Neural Netw.*, vol. 12, no. 3, pp. 604–611, May 2001.
[30] S. Chen, S. R. Gunn, and C. J. Harris, "The relevance vector machine technique for channel equalization application," *IEEE Trans. Neural Netw.*, vol. 12, no. 6, pp. 1529–1532, Nov. 2001.
[31] F. Pérez-Cruz, A. Navia-Vázquez, P. L. Alarcón-Diana, and A. Artés-Rodríguez, "SVC-based equalizer for burst TDMA transmissions," *Signal Process.*, vol. 81, no. 8, pp. 1571–1787, 2001.
[32] A. Wolfgang, S. Chen, and L. Hanzo, "Radial basis function network assisted space-time equalization for dispersive fading environments," *Electron. Lett.*, vol. 40, no. 16, pp. 1006–1007, 2004.
[33] S. Chen, L. Hanzo, and A. Wolfgang, "Kernel-based nonlinear beamforming construction using orthogonal forward selection with Fisher ratio class separability measure," *IEEE Signal Process. Lett.*, vol. 11, no. 5, pp. 478–481, May 2004.
[34] S. Chen, L. Hanzo, and A. Wolfgang, "Nonlinear multi-antenna detection methods," *EURASIP J. Appl. Signal Process.*, vol. 2004, no. 9, pp. 1225–1237, 2004.
[35] S. Chen, B. Mulgrew, and L. Hanzo, "Asymptotic Bayesian decision feedback equalizer using a set of hyperplanes," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3493–3500, Dec. 2000.
[36] X. Zhou and K. Z. Mao, "The ties problem resulting from counting-based error estimators and its impact on gene selection algorithms," *Bioinformatics*, vol. 22, no. 20, pp. 2507–2515, 2006.
[37] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[38] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[39] J. Litva and T. K. Y. Lo, *Digital Beamforming in Wireless Communications*. London, U.K.: Artech House, 1996.
[40] J. S. Blogh and L. Hanzo, *Third Generation Systems and Intelligent Wireless Networking—Smart Antenna and Adaptive Modulation*. Chichester, U.K.: Wiley, 2002.
[41] R. Steele and L. Hanzo, *Mobile Radio Communications*. Piscataway, NJ: IEEE Press, 1999.
[42] S. Chen, N. N. Ahmad, and L. Hanzo, "Adaptive minimum bit error rate beamforming," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 341–348, Mar. 2005.
[43] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
[44] S. Chen, S. McLaughlin, and B. Mulgrew, "Complex-valued radial basis function network, Part II: Application to digital communications channel equalization," *Signal Process.*, vol. 36, pp. 175–188, 1994.
[45] S. Chen, H.-Q. Du, and L. Hanzo, "Adaptive minimum symbol error rate beamforming assisted receiver for quadrature amplitude modulation systems," in *Proc. Veh. Technol. Conf.—Spring*, Melbourne, Australia, May 7–10, 2006, vol. 5, pp. 2236–2240.
[46] S. Chen, S. McLaughlin, B. Mulgrew, and P. M. Grant, "Adaptive Bayesian decision feedback equalizer for dispersive mobile radio channels," *IEEE Trans. Commun.*, vol. 43, no. 5, pp. 1937–1946, May 1995.

Sheng Chen (M'90–SM'96) received the Ph.D. degree in control engineering from the City University, London, U.K., in 1986 and the Doctor of Sciences (D.Sc.) degree by the University of Southampton, Southampton, U.K., in 2005.

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

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

Andreas Wolfgang (M'03) was born in Germany, in 1977. He received the Dipl. Ing. degree in electrical engineering from Karlsruhe University of Technology, Karlsruhe, Germany, in 2003. He is currently working towards the Ph.D. degree at the University of Southampton, Southampton, U.K.

His primary research interests include wideband communication systems, iterative detection as well as nonlinear signal processing.

Chris J. Harris received the Ph.D. degree and the Doctor of Sciences (D.Sc.) degree from the University of Southampton, Southampton, U.K., in 1972 and 2001, respectively.

He previously held appointments at the University of Hull, Hull, U.K., the UMIST, Manchester, U.K., the University of Oxford, Oxford, U.K., and the University of Cranfield, Cranfield, U.K. He was also employed by the U.K. 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, later, 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 and control, intelligent control, and estimation of dynamic processes, multisensor data fusion, and systems integration. He has authored and coauthored 12 research books and over 400 research papers.

Dr. Harris is the Associate Editor of numerous international journals. He was elected to Fellow of the Royal Academy of Engineering (FREng), U.K., in 1996. He 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 (M'91–SM'92–F'04) received the M.S. degree in electronics and the Ph.D. degree from Technical University of Budapest, Budapest, in 1976 and 1983, respectively, and the Doctor of Sciences (D.Sc.) degree from the University Southampton, Southampton, U.K., in 2004.

During his 30-year career in telecommunications he has held various research and academic posts in Hungary, Germany, and United Kingdom. Since 1986 he has been with the School of Electronics and Computer Science, the University of Southampton, U.K., where he holds the chair in telecommunications. He has coauthored 12 Wiley/IEEE Press books totalling about 10 000 pages on mobile radio communications, published in excess of 600 research papers, organized 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.

Prof. Hanzo is a Fellow of the Royal Academy of Engineering (FREng), U.K. He is the IEEE Distinguished Lecturer of both the Communications Society and the Vehicular Technology Society as well as a Fellow of the Institute of Electrical Engineers. He is a nonexecutive director of the Virtual Centre of Excellence (VCE) in mobile communications, U.K., a governor of the IEEE Vehicular Technology society, and an executive board member of the Pan-European Network of Excellence known as NEWCOM.