Symmetric Kernel Detector for Multiple-Antenna Aided Beamforming Systems

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

Abstract— We propose a powerful symmetric kernel classifier for nonlinear detection in challenging rank-deficient multiple-antenna aided communication systems. By exploiting the inherent odd symmetry of the optimal Bayesian detector, the proposed symmetric kernel classifier is capable of approaching the optimal classification performance using noisy training data. The classifier construction process is robust to the choice of the kernel width and is computationally efficient. The proposed solution is capable of providing a signal-to-noise ratio gain in excess of 8 dB against the powerfull linear minimum bit error rate benchmarker, when supporting five users with the aid of three receive antennas.

I. Introduction

Kernel-modelling techniques have found wide-ranging applications in regression and classification [1]-[16]. The standard kernel modelling method constitutes a black-box approach that seeks a (usually sparse) model representation extracted from the training data. Adopting black-box modelling is appropriate, if no *a priori* information exists regarding the underlying data generating mechanism. However, a fundamental principle in practical data modelling is that if there exists *a priori* information concerning the system to be modelled it should be incorporated in the modelling process. Many real-life phenomena exhibit inherent symmetry, but these properties are hard to infer from data with the aid of black-box-type kernel or radial basis function (RBF) models. However, by imposing symmetry on the model's structure, exploiting the symmetry properties becomes easier and this often leads to substantial improvements in the achievable modelling 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) [17],[18].

We consider nonlinear detection in multiple-antenna assisted beamforming systems. Detection in communication receivers in general 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 [19]-[28]. A kernel classifier or detector attempts to realise or approximate the underlying optimal Bayesian solution. Previous studies [19]-[28] have shown that a block-box kernel detector typically requires more kernels than the number of the channel states to approximate the Baysian detector, and moreover there often exists a performance gap between the kernel detector and the Baysian solution. This performance degradation can be explained as follows. The Bayesian

The authors are with School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK, E-mails: {sqc,aw03r,cjh,lh}@ecs.soton.ac.uk

nonlinear detection solution has an inherent odd symmetry, because the signal states corresponding to the different signal classes are distributed symmetrically with respect to the optimal decision boundary [29]. A black-box kernel classifier or detector, however, has difficulty realising this symmetry.

The novelty of this contribution is that we propose a symmetric kernel classifier for multiple-antenna aided communication systems, which renders realisation of the symmetric Baysian detection solution easier. The orthogonal forward selection (OFS) procedure [13],[16],[27],[28] can readily be applied to construct a sparse representation for this symmetric kernel classifier based on various criteria, such as the Fisher ratio of class separability measure (FRCSM) [13],[27] and the leave-one-out misclassification rate (LOO-MR) [16]. The OFS procedures based on the FRCSM and the LOO-MR are computationally very efficient, in comparion to other existing kernel construction methods. We adopt the FRCSM, since it is computationally even simpler to implement than the LOO-MR. It is interesting to see that, even though we do not directly minimise the misclassification rate, the sparse symmetric kernel classifier constructed by incrementally maximising the FRCSM is capable of approaching the minimum misclassification rate.

The advantage of the proposed symmetric kernel 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 [30],[31]. Although we apply the proposed symmetric kernel classifier in the context of multiple-antenna aided beamforming systems, it is equally applicable to other classification problems with similar symmetric properties. To the best of our knowledge, this is the first time that the symmetry is explicitly exploited in kernel classifier construction.

II. 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 angular domain [30],[31] if the receiver is equipped with a linear antenna array consisting of $L > 1$ uniformly spaced elements. Assume that the channel is non-dispersive which does not induce intersymbol interference. Then the symbol-rate complex-valued received signal samples can be expressed as

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

for $1 \leq l \leq L$, 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 non-dispersive channel coefficient of user i, and $b_i(k)$ is the k-th symbol of user i, which assumes values from the binary phase shift keying (BPSK) symbol set, i.e. $b_i(k) \in \{\pm 1\}$. 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 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 $2 \leq i \leq S$. The received signal vector $\mathbf{x}(k) = [x_1(k) \; x_2(k) \cdots x_L(k)]^T$ can be expressed as

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

where $\mathbf{n}(k) = [n_1(k) \; n_2(k) \cdots n_L(k)]^T$, the system matrix is given by $\mathbf{P} = [A_1\mathbf{s}_1 \; A_2\mathbf{s}_2 \cdots A_S\mathbf{s}_S]$ and the steering vector for source i is $\mathbf{s}_i = [e^{j\omega t_1(\theta_i)} \; e^{j\omega t_2(\theta_i)} \cdots e^{j\omega t_L(\theta_i)}]^T$, while the transmitted BPSK symbol vector is $\mathbf{b}(k) = [b_1(k) \; b_2(k) \cdots b_S(k)]^T$.

Classically, a linear beamforming receiver is adopted to detect the desired user signal [32],[33]. The output of the linear beamformer is defined by

$$y_{\text{Lin}}(k) = \mathbf{w}^H \mathbf{x}(k) \qquad (3)$$

and the associated decision is given by

$$
\begin{aligned}
\hat{b}_1(k) &= \text{sgn}(\Re[y_{\text{Lin}}(k)]) \\
&= \begin{cases} +1, & \Re[y_{\text{Lin}}(k)] \geq 0, \\ -1, & \Re[y_{\text{Lin}}(k)] < 0, \end{cases}
\end{aligned} \qquad (4)
$$

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 $\mathbf{w}$ is set to the minimum mean square error (L-MMSE) solution [32],[33]. 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 upto $S = L$ users. If the number of users S is larger than the number of array elements L, the system is referred to as rank-deficient. The state-of-the-art design for the linear beamformer (3) in fact is the the (linear) minimum bit error rate (L-MBER) solution [34], which directly minimises the error probability or bit error rate (BER) of the linear beamformer (3). The L-MBER beamforming outperforms the L-MMSE one significantly, particularly for rank-deficient systems. The L-MBER design is optimal for the *linear* beamforming. The true optimal solution for the multiple antenna aided beamforming detector, however, is *nonlinear* [27],[28].

Let us denote the $N_b = 2^S$ legitimate combinations of $\mathbf{b}(k)$ as $\mathbf{b}_q$, $1 \leq q \leq N_b$, and denote 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 $\mathcal{X} \triangleq \{\bar{\mathbf{x}}_q = \mathbf{P}\mathbf{b}_q, \; 1 \leq q \leq N_b\}$, which can be divided into two subsets conditioned on the value of $b_1(k)$ as follows

$$\mathcal{X}^{(\pm)} \triangleq \{\bar{\mathbf{x}}_i \in \mathcal{X}, 1 \leq i \leq N_{sb} : \; b_1(k) = \pm 1\}, \qquad (5)$$

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 [35], the optimal detection strategy should be

$$\hat{b}_1(k) = \begin{cases} +1, & \text{if } \; p_+(\mathbf{x}(k)) \geq p_-(\mathbf{x}(k)), \\ -1, & \text{if } \; p_+(\mathbf{x}(k)) < p_-(\mathbf{x}(k)). \end{cases} \qquad (6)$$

By introducing the real-valued Bayesian decision variable

$$y_{\text{Bay}}(k) = f_{\text{Bay}}(\mathbf{x}(k)) \triangleq \frac{1}{2}p_+(\mathbf{x}(k)) - \frac{1}{2}p_-(\mathbf{x}(k)), \qquad (7)$$

the optimal Bayesian detection rule (6) is equivalent to $\hat{b}_1(k) = \text{sgn}(y_{\text{Bay}}(k))$. The decision variable (7) of the optimal Bayesian detector is readily expressed as [27],[28]

$$y_{\text{Bay}}(k) = \sum_{q=1}^{N_b} \text{sgn}(b_{q,1})\beta_q e^{-\frac{\|\mathbf{x}(k) - \bar{\mathbf{x}}_q\|^2}{2\sigma_n^2}} \qquad (8)$$

where β_q denotes the a priori probability of $\bar{\mathbf{x}}_q$. Since in our case, all the $\bar{\mathbf{x}}_q$ are equiprobable, we have $\beta_q = \beta > 0$.

It can readily be shown that the two subsets $\mathcal{X}^{(+)}$ and $\mathcal{X}^{(-)}$ are symmetric with respect to each other [29]. That is, 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 $\bar{\mathbf{x}}_i^{(-)} = -\bar{\mathbf{x}}_i^{(+)}$. Given this symmetry, the optimal Bayesian detector (8) can be rewritten as

$$y_{\text{Bay}}(k) = \sum_{q=1}^{N_{sb}} \beta_q \left(e^{-\frac{\|\mathbf{x}(k) - \bar{\mathbf{x}}_q^{(+)}\|^2}{2\sigma_n^2}} - e^{-\frac{\|\mathbf{x}(k) + \bar{\mathbf{x}}_q^{(+)}\|^2}{2\sigma_n^2}} \right), \qquad (9)$$

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

If the system matrix $\mathbf{P}$ is known, the signal state subset $\mathcal{X}^{(+)}$ can be computed and the Bayesian detection solution is specified. For the multiple-antenna aided beamformer, 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. But the receiver does not have access to the interfering users' data $\{b_i(k)\}$, $i \neq 1$. Thus, estimating the system matrix $\mathbf{P}$ is a challenging task. In our previous work [27],[28], standard kernel-based classifiers or detectors were constructed directly using the noisy training data set D_K to approximate the optimal Bayesian solution. It is clear that the inherent symmetry of the Bayesian detector in (9) is hard to learn by a black-box kernel classifier. We propose a novel symmetric kernel classifier which renders realisation of the symmetric Baysian detection solution easier.

III. THE SYMMETRIC KERNEL CLASSIFIER

The problem is to train a two-class kernel classifier $y_{\text{Ker}}(\mathbf{x}) : \mathcal{C}^L \to \{1, -1\}$ 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 kernel classifier of the form $\hat{d}(k) = \text{sgn}(y_{\text{Ker}}(k))$ with

$$y_{\text{Ker}}(k) = f_{\text{Ker}}(\mathbf{x}(k)) \triangleq \sum_{i=1}^{M} \theta_i \phi_i(\mathbf{x}(k)), \qquad (10)$$

where $\hat{d}(k)$ is the estimated class label for $\mathbf{x}(k)$, $\phi_i(\bullet)$ denotes the classifier's kernels, θ_i are the classifier's coefficients and M is the number of kernels. We propose to adopt the following symmetric kernel

$$\phi_i(\mathbf{x}) \triangleq \varphi(\mathbf{x}; \mathbf{c}_i, \rho^2) - \varphi(\mathbf{x}; -\mathbf{c}_i, \rho^2), \qquad (11)$$

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

$$\varphi(\mathbf{x}; \mathbf{c}_i, \rho^2) = e^{-\frac{\|\mathbf{x} - \mathbf{c}_i\|^2}{2\rho^2}}. \qquad (12)$$

Other kernel functions can also be used here. It is worth emphasising that, although we derive the symmetric kernel formulation directly through the observation of the underlying symmetric Bayesian detection solution, the proposed symmetric kernel detector can also be derived analytically by imposing the odd symmetry constraint on the standard kernel formulation, just as in the regression case [18].

Because the symmetric kernel formulation (10) has the same form to the standard kernel formulation, most of the existing sparse kernel techniques can be applied. Our previous experience with standard sparse kernel modelling suggests that the OFS procedure based on the FRCSM [27],[28] compares favourably 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 a kernel detector as sparse as possible, because the detection complexity scales with the size of the kernel classifier. We apply the OFS procedure based on the FRCSM to construct a sparse symmetric kernel 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 [35].

Consider every training data point $\mathbf{x}(i)$ as a candidate kernel centre. Hence we have $M = K$ in the kernel model of (10) and $\mathbf{c}_i = \mathbf{x}(i)$ for $1 \leq i \leq K$, and the kernel variance is set to ρ^2. Let us now define $\varepsilon(i) = d(i) - y_{\text{Ker}}(i)$ as the modelling residual sequence. Then the model (10) defined over the training data set D_K can be written in matrix form as

$$\mathbf{d} = \mathbf{\Phi}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \qquad (13)$$

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

$$\mathbf{\Phi} = [\phi_1 \; \phi_2 \cdots \phi_M] \in \mathcal{R}^{K \times M} \qquad (14)$$

is the regression matrix with the column vectors $\phi_i = [\phi_i(\mathbf{x}(1)) \; \phi_i(\mathbf{x}(2)) \cdots \phi_i(\mathbf{x}(K))]^T$, $1 \leq i \leq M$. Let an

orthogonal decomposition of $\mathbf{\Phi}$ be $\mathbf{\Phi} = \mathbf{\Omega}\mathbf{A}$, where we have

$$\mathbf{A} = \begin{bmatrix} 1 & \alpha_{1,2} & \cdots & \alpha_{1,M} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \alpha_{M-1,M} \\ 0 & \cdots & 0 & 1 \end{bmatrix} \qquad (15)$$

and

$$\begin{aligned} \mathbf{\Omega} &= [\boldsymbol{\omega}_1 \; \boldsymbol{\omega}_2 \cdots \boldsymbol{\omega}_M] \\ &= \begin{bmatrix} \omega_{1,1} & \omega_{1,2} & \cdots & \omega_{1,M} \\ \omega_{2,1} & \omega_{2,2} & \cdots & \omega_{2,M} \\ \vdots & \vdots & \vdots & \vdots \\ \omega_{K,1} & \omega_{K,2} & \cdots & \omega_{K,M} \end{bmatrix} \end{aligned} \qquad (16)$$

with orthogonal columns that satisfy $\boldsymbol{\omega}_i^T \boldsymbol{\omega}_l = 0$, if $i \neq l$. The model (13) can alternatively be expressed as

$$\mathbf{d} = \mathbf{\Omega}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \qquad (17)$$

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 maximising the FRCSM using the OFS procedure, as is in [13],[27],[28]. Define the two class sets $\mathbf{X}_\pm = \{\mathbf{x}(k) : d(k) = \pm 1\}$, 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

$$m_{+,l} = \frac{1}{K_+} \sum_{k=1}^{K} \delta(d(k) - 1)\omega_{k,l}, \qquad (18)$$

$$\sigma_{+,l}^2 = \frac{1}{K_+} \sum_{k=1}^{K} \delta(d(k) - 1)(\omega_{k,l} - m_{+,l})^2, \qquad (19)$$

and

$$m_{-,l} = \frac{1}{K_-} \sum_{k=1}^{K} \delta(d(k) + 1)\omega_{k,l}, \qquad (20)$$

$$\sigma_{-,l}^2 = \frac{1}{K_-} \sum_{k=1}^{K} \delta(d(k) + 1)(\omega_{k,l} - m_{-,l})^2, \qquad (21)$$

respectively, where $\delta(x) = 1$ if $x = 0$ and $\delta(x) = 0$ if $x \neq 0$. 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 [35]

$$F_l = \frac{(m_{+,l} - m_{-,l})^2}{\sigma_{+,l}^2 + \sigma_{-,l}^2}. \qquad (22)$$

Based on this FRCSM, significant kernel terms can be selected with the aid of an OFS procedure. At the l-th stage, a candidate term is chosen as the l-th kernel 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

$$\frac{F_{M_{\text{spa}}}}{\sum_{l=1}^{M_{\text{spa}}} F_l} < \xi, \qquad (23)$$

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 orthogonalisation [9] is summarised below.

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

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

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

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

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

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

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

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

2) Find

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

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

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

3) The procedure is monitored and terminated at the index value $l = M_{\mathrm{spa}}$, when for example the condition (23) 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

$$\gamma_l = \frac{\boldsymbol{\omega}_l^T \mathbf{d}}{\boldsymbol{\omega}_l^T \boldsymbol{\omega}_l}. \tag{24}$$

Instead of using the condition (23) 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
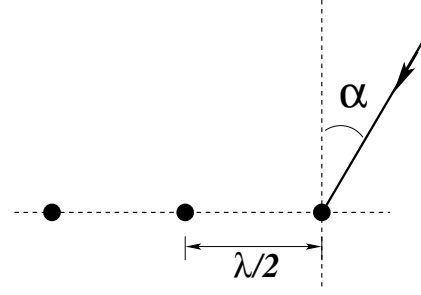

Fig. 1. Geometric structure of the three-element linear array having $\lambda/2$ spacing, where λ is the wavelength and α the angle of arrival of a user.

criteria such as the information based criteria and optimal experimental design criteria of [14] may also be used. The kernel 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 kernel classifier is not sensitive to the value of ρ^2 used, and there exists a large range of ρ^2 values which enables the sparse symmetric kernel 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. It has been shown [36] that the performance of the Bayesian detectors using $0.2\sigma_n^2$ and $5\sigma_n^2$ to substitute the noise variance σ_n^2 is indistinguishable from that of the exact Bayesian solution.

IV. SIMULATION STUDY

The example consisted of a three-element antenna array supporting five BPSK users. Fig. 1 shows the antenna array geometric structure and Table I lists the angular locations of the five users with respect to the antenna array. The simulated channel conditions were $A_i = 1 + j0$, $1 \leq i \leq 5$. The desired user and all the four interfering users had equal signal power, and therefore $\mathrm{SIR}_i = 0$ dB for $i = 2, 3, 4$. Fig. 2 portrays the BER performance of both the theoretical L-MBER beamformer and the Bayesian detector for the desired user 1. For each SNR value, $K = 600$ training samples $D_K = \{\mathbf{x}(k), b_1(k)\}_{k=1}^{K}$ were used to construct the symmetric kernel classifier employing the FRCSM-based OFS algorithm as outlined in the previous section. The kernel variance was chosen to be $\rho^2 = 3\sigma_n^2$. As the size of the Bayesian detector was $N_{sb} = 16$, we terminated the kernel classifier construction at $M_{\mathrm{spa}} = 16$. The BER performance of the 16-term symmetric kernel detector is also depicted in Fig. 2. It can be seen from Fig. 2 that the symmetric kernel detector is capable of closely approaching the optimal Bayesian performance when the number of symmetric

TABLE I
LOCATIONS OF USERS IN TERMS OF ANGLE OF ARRIVAL (AOA) FOR THE THREE-ELEMENT ANTENNA ARRAY SYSTEM SUPPORTING FIVE USERS.

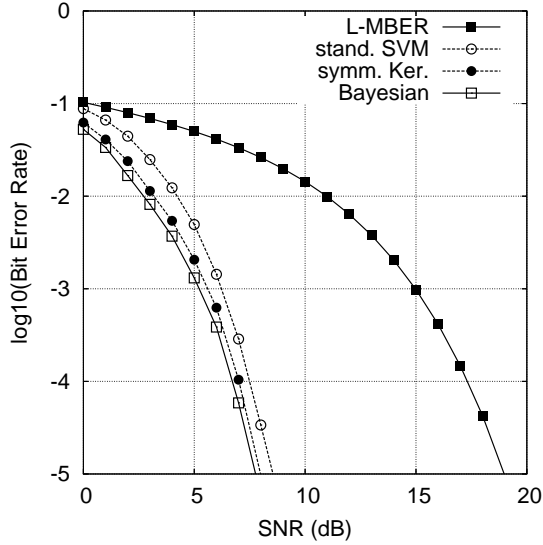| user i	1	2	3	4	5
AOA α	$0°$	$10°$	$-17°$	$15°$	$20°$

Fig. 2. The desired-user's bit error rate performance in the context of four detectors. The symmetric kernel classifier, constructed from 600 noisy training samples using the FRCSM-based OFS, has $M_{\mathrm{spa}} = 16$ symmetry kernels and a kernel variance $\rho^2 = 3\sigma_n^2$. The standard SVM classifier, constructed from the same 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 $6\sigma_n^2$.

kernels is no larger than that of the Baysian detector, and hence outperforms the black-box kernel methods of [27],[28]. 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. To demonstrate this point, the standard SVM detector was trained using the same 600-sample training data set. The size of the SVM detector constructed ranged from 40 to 60 support vectors, 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. 2.

The properties of the proposed FRCSM-based OFS in-

Fig. 4. The influence of the training data length on the bit error rate performance of the symmetric kernel classifier. We used SNR= 5 dB, a kernel model size $M_{\mathrm{spa}} = 16$ and a kernel variance $\rho^2 = 3\sigma_n^2$.

voked for constructing the symmetric kernel detector were studied. First, the influence of the model size M_{spa} on the kernel classifier's performance was investigated. Given SNR= 5 dB and a training data length of $K = 600$, Fig. 3 shows the performance of the symmetric kernel classifier as a function of the model size M_{spa}. The kernel variance ρ^2 was tuned according to the model size M_{spa}, and was in range of $3\sigma_n^2$ to $5\sigma_n^2$. Next the influence of the training data length K was investigated. Given SNR= 5 dB, a kernel variance of $\rho^2 = 3\sigma_n^2$ and a kernel model size of $M_{\mathrm{spa}} = 16$, Fig. 4 plots the performance of the symmetric kernel detector as a function of the training data length K. The influence of the kernel variance on the performance of the symmetric kernel classifier was also investigated. Given SNR= 5 dB, a training data length of $K = 500$ and a kernel model size $M_{\mathrm{spa}} = 16$, Fig. 5 illustrates the performance of the symmetric kernel detector as a function of the kernel variance. The result of Fig. 5 confirms that there exists a large range of values for ρ^2, which allow the spare symmetric kernel detector to approach the optimal Bayesian performance.
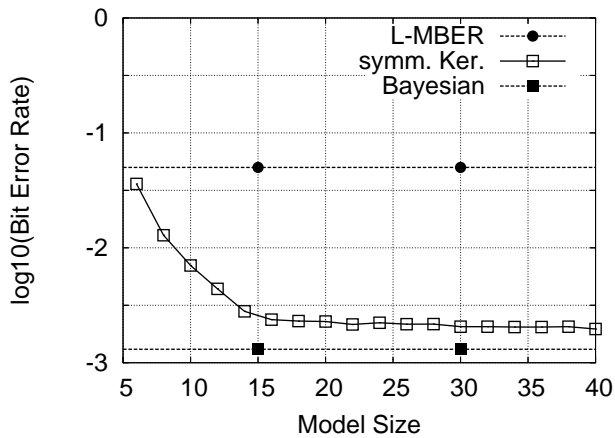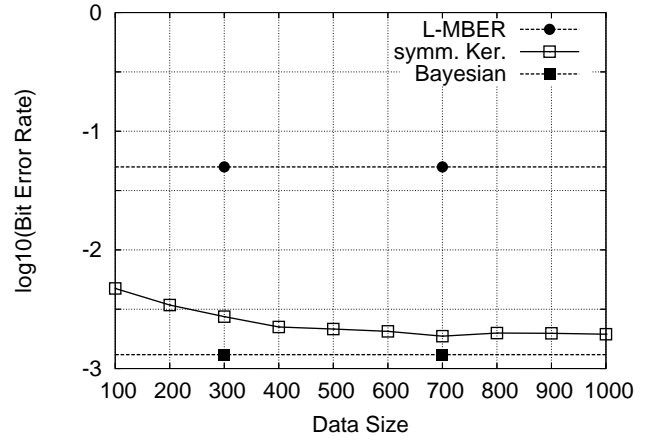

Fig. 3. The influence of the classifier's size on the bit error rate performance of the symmetric kernel classifier. We used SNR= 5 dB, a training data length $K = 600$ and the kernel variance ρ^2 was varied depending on the model size.
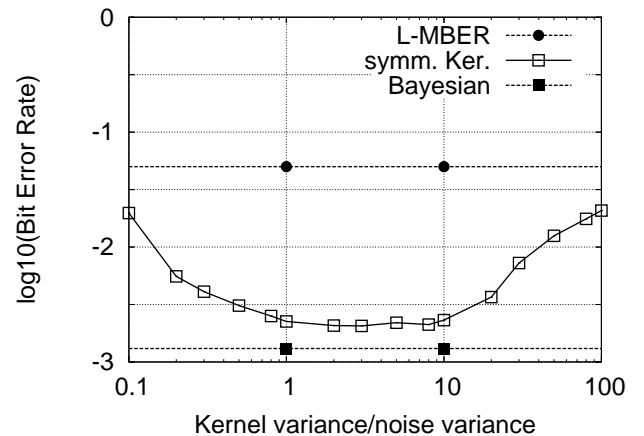

Fig. 5. The influence of the kernel variance on the bit error rate performance of the symmetric kernel classifier. We used SNR= 5 dB, a training data length $K = 600$ and a kernel model size $M_{\mathrm{spa}} = 16$.

V. Conclusions

A novel symmetric kernel 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 kernel detector from the noisy training data, which is capable of approaching the optimal Bayesian detection performance. The proposed solution provides a signal-to-noise ratio gain in excess of 8 dB against the powerful linear minimum bit error rate benchmark, when supporting five users with the aid of three receive antennas. Compared with the standard sparse kernel methods, which do not exploit the symmetry of the underlying solution, our proposed method is computationally simpler, results in a much smaller detector size, and performs better. Although we have presented this sparse symmetric kernel classifier in the context of nonlinear detection in wireless communication systems, it is generically applicable to any classification problem exhibiting a similar symmetry.

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," *Technical Report*, ISIS Research Group, Department of Electronics and Computer Science, University of Southampton, UK, May 1998.

[3] S.S. Chen, D.L. Donoho and M.A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol.43, no.1, pp.129–159, 2001.

[4] M.E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Machine Learning Research*, vol.1, pp.211–244, 2001.

[5] B. Schölkopf and A.J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press: Cambridge, MA, 2002.

[6] P. Vincent and Y. Bengio, "Kernel matching pursuit," *Machine Learning*, vol.48, no.1, pp.165–187, 2002.

[7] D. Decoste and B. Schölkopf, "Training invariant support vector machines," *Machine Learning*, 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. Machine Learning Research*, vol.5, pp.27–72, 2004.

[9] 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.

[10] S. Chen, C.F.N. Cowan and P.M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Networks*, vol.2, no.2, pp.302–309, 1991.

[11] S. Chen, Y. Wu and B.L. Luk, "Combined genetic algorithm optimisation and regularised orthogonal least squares learning for radial basis function networks," *IEEE Trans. Neural Networks*, vol.10, no.5, pp.1239–1243, 1999.

[12] 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 Networks*, vol.15, no.2, pp.263–270, 2002.

[13] K.Z. Mao, "RBF neural network center selection based on Fisher ratio class separability measure," *IEEE Trans. Neural Networks*, vol.13, no.5, pp.1211–1217, 2002.

[14] S. Chen, X. Hong and C.J. Harris, "Sparse kernel regression modelling using combined locally regularized orthogonal least squares and D-optimality experimental design," *IEEE Trans. Automatic Control*, vol.48, no.6, pp.1029–1036, 2003.

[15] S. Chen, X. Hong, C.J. Harris and P.M. Sharkey, "Sparse modelling using orthogonal forward regression with PRESS statistic and regularization," *IEEE Trans. Systems, Man and Cybernetics, Part B: Cybernetics*, vol.34, no.2, pp.898–911, 2004.

[16] X. Hong, S. Chen and C.J. Harris, "Fast kernel classifier construction using orthogonal forward selection to minimise leave-one-out misclassification rate," in *Proc. 2006 Int. Conf. Intelligent Computing* (Kunming, China), Aug.16-19, 2006, pp.106-114.

[17] L.A. Aguirre, R.A.M. Lopes, G.F.V. Amaral and C. Letellier, "Constraining the topology of neural networks to ensure dynamics with symmetry properties," *Physical Review E*, vol.69, pp.026701-1–026701-11, 2004.

[18] M. Espinoza, J.A.K. Suykens and B. De Moor, "Imposing symmetry in least squares support vector machines regression," in *Proc. Joint 44th IEEE Conf. Decision and Control, and the European Control Conf. 2005* (Seville, Spain), Dec.12-15, 2005, pp.5716–5721.

[19] S. Chen and B. Mulgrew, "Overcoming co-channel interference using an adaptive radial basis function equaliser," *Signal Processing*, vol.28, no.1, pp.91–107, 1992.

[20] S. Chen, B. Mulgrew and P.M. Grant, "A clustering technique for digital communications channel equalisation using radial basis function networks," *IEEE Trans. Neural Networks*, vol.4, no.4, pp.570–579, 1993.

[21] 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 Networks for Signal Processing* (Madison, WI), Aug.23-25, 1999, p.49-57.

[22] D.J. Sebald and J.A. Bucklew, "Support vector machine techniques for nonlinear equalization," *IEEE Trans. Signal Processing*, vol.48, no.11, pp.3217–3226, 2000.

[23] S. Chen, A.K. Samingan and L. Hanzo, "Support vector machine multiuser receiver for DS-CDMA signals in multipath channels," *IEEE Trans. Neural Networks*, vol.12, no.3, pp.604-611, 2001.

[24] S. Chen, S.R. Gunn and C.J. Harris, "The relevance vector machine technique for channel equalization application," *IEEE Trans. Neural Networks*, vol.12, no.6, pp.1529–1532, 2001.

[25] F. Pérez-Cruz, A. Navia-Vázquez, P.L. Alarcón-Diana and A. Artés-Rodrguez, "SVC-based equalizer for burst TDMA transmissions," *Signal Processing*, vol.81, no.8, pp.1571–1787, 2001.

[26] A. Wolfgang, S. Chen and L. Hanzo, "Radial basis function network assisted space-time equalisation for dispersive fading environments," *Electronics Letters*, vol.40, no.16, pp.1006–1007, 2004.

[27] S. Chen, L. Hanzo and A. Wolfgang, "Kernel-based nonlinear beamforming construction using orthogonal forward selection with Fisher ratio class separability measure," *IEEE Signal Processing Letters*, vol.11, no.5, pp.478–481, 2004.

[28] S. Chen, L. Hanzo and A. Wolfgang, "Nonlinear multi-antenna detection methods," *EURASIP J. Applied Signal Processing*, vol.2004, no.9, pp.1225–1237, 2004.

[29] S. Chen, B. Mulgrew and L. Hanzo, "Asymptotic Bayesian decision feedback equalizer using a set of hyperplanes," *IEEE Trans. Signal Processing*, vol.48, no.12, pp.3493–3500, 2000.

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

[31] D. Tse, and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press: Cambridge, UK, 2005.

[32] J. Litva and T. K.Y. Lo, *Digital Beamforming in Wireless Communications*. London: Artech House, 1996.

[33] J.S. Blogh and L. Hanzo, *Third Generation Systems and Intelligent Wireless Networking – Smart Antenna and Adaptive Modulation*. Chichester: John Wiley, 2002.

[34] S. Chen, N.N. Ahmad and L. Hanzo, "Adaptive minimum bit error rate beamforming," *IEEE Trans. Wireless Communications*, vol.4, no.2, pp.341–348, 2005.

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

[36] S. Chen, S. McLaughlin, B. Mulgrew and P.M. Grant, "Adaptive Bayesian decision feedback equaliser for dispersive mobile radio channels," *IEEE Trans. Communications*, vol.43, no.5, pp.1937–1946, 1995.