Kernel-Based Nonlinear Beamforming Construction Using Orthogonal Forward Selection With the Fisher Ratio Class Separability Measure

S. Chen, *Senior Member, IEEE*, L. Hanzo, *Fellow, IEEE*, and A. Wolfgang

***Abstract*—This letter shows that the wireless communication system capacity is greatly enhanced by employing nonlinear beamforming and that the optimal Bayesian beamformer outperforms the standard linear beamformer significantly in terms of a reduced bit error rate, at a cost of increased complexity. A block-data adaptive implementation of the Bayesian beamformer is realized based on an orthogonal forward selection procedure with the Fisher ratio for class separability measure.**

***Index Terms*—Bayesian classification, Fisher ratio for class separability measure, nonlinear beamforming, orthogonal least squares.**

I. Introduction

SPATIAL processing with adaptive antenna arrays has shown real promise for substantial capacity enhancement in mobile communication [1]–[5]. Adaptive beamforming can separate signals transmitted on the same carrier frequency, provided that they are separated in the spatial domain. The beamforming processing is classically done by forming a linear combination of the signals received from the different elements of an antenna array. We refer to this classical beamforming as *linear* beamforming. Recent work [6] has investigated a linear beamforming technique based directly on minimizing the system bit error rate (BER) and developed an adaptive algorithm for realizing the linear minimum BER (LMBER) beamforming. The results in [6] have demonstrated that the LMBER beamforming provides considerable performance gains in terms of a reduced BER over the usual linear minimum mean-square error (LMMSE) beamforming.

The spatial separation in angles of arrival between the desired signal and the closest interfering signal determines the system performance and hence capacity. When this separation is below a certain threshold, linear beamforming ultimately fails because the system becomes linearly inseparable, a situation that is similar to the single-user channel equalization [7], [8]. For the sake of notational simplicity and for highlighting the basic concepts, we assume that the modulation scheme is binary phase shift keying (BPSK), the channel is nondispersive with additive white Gaussian noise, and narrowband beamforming is considered. We derive the optimal solution for nonlinear beamforming, which we refer to as the Bayesian beamforming solution. A block-data kernel-based adaptive beamformer is proposed to realize the optimal Bayesian beamformer solution using an orthogonal forward selection (OFS) procedure with the Fisher ratio for class separability measure [9]. The proposed nonlinear beamformer construction algorithm is compared with the state-of-art sparse kernel modeling based on the relevance vector machine (RVM) for classification [8], [10].

Manuscript received May 1, 2003; revised September 9, 2003. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jonathon A. Chambers.

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

Digital Object Identifier 10.1109/LSP.2004.826509

II. System Model

It is assumed that the system consists of M users (sources), and each user transmits a BPSK signal on the same carrier frequency $\omega = 2\pi f$. The baseband signal of user i is given by

$$m_i(k) = A_i b_i(k), \quad b_i(k) \in \{\pm 1\}, \qquad 1 \le i \le M \tag{1}$$

where the complex-valued A_i is the channel coefficient for user i multiplying by the transmitted signal amplitude of user i (therefore $|A_i|^2$ denotes user i received signal power) and $b_i(k)$ is the kth bit of user i. Without the loss of generality, source 1 is assumed to be the desired user, and the rest of the sources are interfering users. The linear antenna array is considered, which consists of L uniformly spaced elements, and signals at the L-element antenna array are given by

$$\begin{aligned} x_l(k) &= \sum_{i=1}^{M} m_i(k) \exp(j\omega t_l(\theta_i)) + n_l(k) \\ &= \bar{x}_l(k) + n_l(k) \end{aligned} \tag{2}$$

for $1 \le l \le L$, where $t_l(\theta_i)$ is the relative time delay at element l for source i, θ_i is the direction of arrival for source i, and $n_l(k)$ is a complex-valued white Gaussian noise with zero mean and $E[|n_l(k)|^2] = 2\sigma_n^2$. The desired SNR is defined as SNR $= |A_1|^2/2\sigma_n^2$, and the desired signal-to-interferer i ratio is given by $\text{SIR}_i = |A_1|^2/|A_i|^2$ for $i = 2, \ldots, M$. In vector form, the array input can be written as

$$\begin{aligned} \mathbf{x}(k) &= [x_1(k), \ldots, x_L(k)]^T = \bar{\mathbf{x}}(k) + \mathbf{n}(k) \\ &= \mathbf{P}\mathbf{b}(k) + \mathbf{n}(k) \end{aligned} \tag{3}$$

where $E[\mathbf{n}(k)\mathbf{n}^H(k)] = 2\sigma_n^2 \mathbf{I}_L$ with $\mathbf{I}_L$ denoting the $L \times L$ identity matrix, the system matrix is defined by

$$\mathbf{P} = [A_1\mathbf{s}_1, A_2\mathbf{s}_2, \ldots, A_M\mathbf{s}_M] \tag{4}$$

with the steering vector for source i being $\mathbf{s}_i = [\exp(j\omega t_1(\theta_i)), \exp(j\omega t_2(\theta_i)), \ldots, \exp(j\omega t_L(\theta_i))]^T$, and the bit vector $\mathbf{b}(k) = [b_1(k), b_2(k), \ldots, b_M(k)]^T$.

Traditionally, a linear beamformer is used, whose output is given by

$$y(k) = \mathbf{w}^H \mathbf{x}(k) \tag{5}$$

where $\mathbf{w}$ is the complex-valued beamformer weight vector. The decision for the transmitted bit $b_1(k)$ is made according to

$$\hat{b}_1(k) = \begin{cases} +1, & y_R(k) > 0 \\ -1, & y_R(k) \le 0 \end{cases} \tag{6}$$

where $y_R(k) = \Re[y(k)]$ denotes the real part of $y(k)$. The classical LMMSE beamforming solution is given by $\mathbf{w}_{\text{MMSE}} = (\mathbf{P}\mathbf{P}^H + 2\sigma_n^2 \mathbf{I}_L)^{-1}\mathbf{p}_1$, with $\mathbf{p}_1$ being the first column of $\mathbf{P}$. Recently, we have developed the LMBER beamforming solution [6]. For the linear beamformer to work adequately, the system must be linearly separable in the noise-free case. When the minimum spatial separation in angles of arrival between the desired user and interfering users is below a certain threshold, the system inevitably becomes linearly inseparable. In such a situation, the linear bermformer exhibits a high irreducible BER floor, and a nonlinear processing has to be adopted.

III. Bayesian Beamforming Solution

Given the observation vector $\mathbf{x}(k)$, the optimal solution to the beamforming problem is the maximum *a posteriori* probability solution, which we derive as follows. Denote the $N_b = 2^M$ possible sequences of $\mathbf{b}(k)$ as $\mathbf{b}_q, 1 \le q \le N_b$. Further, denote the first element of $\mathbf{b}_q$, corresponding to the desired user, as $b_{q,1}$. Obviously, $\bar{\mathbf{x}}(k)$ only takes values from the signal state set defined as $\mathcal{X} \triangleq \{\bar{\mathbf{x}}_q = \mathbf{P}\mathbf{b}_q, 1 \le q \le N_b\}$. The state set $\mathcal{X}$ can be divided into two subsets conditioned on $b_1(k)$

$$\mathcal{X}^{(\pm)} \triangleq \left\{ \bar{\mathbf{x}}_q^{(\pm)} \in \mathcal{X}, 1 \le q \le N_{\text{sb}} : b_1(k) = \pm 1 \right\} \tag{7}$$

where $N_{\text{sb}} = N_b/2$. The posterior probabilities or decision variables for $b_1(k) = \pm 1$ given $\mathbf{x}(k)$ are

$$\eta^{(\pm)}(k) = \sum_{q=1}^{N_{\text{sb}}} \frac{\xi_q^{(\pm)}}{(2\pi\sigma_n^2)^L} \exp\left(-\frac{\left\| \mathbf{x}(k) - \bar{\mathbf{x}}_q^{(\pm)} \right\|^2}{2\sigma_n^2} \right) \tag{8}$$

where $\xi_q^{(\pm)} = (1/N_{\text{sb}})$ are *a priori* probabilities of $\bar{\mathbf{x}}_q^{(\pm)}$ and $\|\mathbf{x}\|^2 = \mathbf{x}^H\mathbf{x}$. The optimal decision is given by

$$\hat{b}_1(k) = \begin{cases} +1, & \eta^{(+)}(k) \ge \eta^{(-)}(k) \\ -1, & \text{otherwise.} \end{cases} \tag{9}$$

Let us redefine a single decision variable as

$$y_B(k) = \sum_{q=1}^{N_b} c_q \exp\left(-\frac{\|\mathbf{x}(k) - \bar{\mathbf{x}}_q\|^2}{2\sigma_n^2} \right) \tag{10}$$

where $c_q = \text{sgn}(b_{q,1})/(N_b(2\pi\sigma_n^2)^L)$. Then, the optimal decision (9) is equivalent to

$$\hat{b}_1(k) = \begin{cases} +1, & y_B(k) \ge 0 \\ -1, & y_B(k) < 0. \end{cases} \tag{11}$$

IV. Block-Data Kernel-Based Nonlinear Beamformer Construction

Given a block of N training data $\{\mathbf{x}(k), b_1(k)\}_{k=1}^N$, consider the nonlinear beamformer of the form

$$y(\mathbf{x}) = \sum_{l=1}^{N} \beta_l \phi_l(\mathbf{x}) \tag{12}$$

where β_l are the real-valued weights, and $\phi_l(\mathbf{x}) = \phi(\mathbf{x}, \mathbf{x}(l))$ are chosen kernel basis functions. In our application, $\phi(\cdot,\cdot)$ can be chosen as the Gaussian kernel function of the form

$$\phi(\mathbf{x}, \mathbf{x}(l)) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}(l)\|^2}{2\rho^2} \right) \tag{13}$$

where the kernel variance ρ^2 is related to the noise variance σ_n^2. The RVM method [8], [10] can be applied to construct a sparse beamformer of N_{spa} terms from (12). A drawback of the RVM method is its high computational complexity. The algorithm contains two loops, with the inner loop for updating the kernel weights and the outer loop for the associated hyperparameters. Both loops involve expensive nonlinear optimization. Furthermore, the RVM method starts with the full model set and removes those kernel terms that have large values in their associated hyperparameters. Because the Hessian matrix associated with the full model set is typically ill-conditioned and may even be noninvertible, the RVM method is inherently ill-conditioned, and its iterative procedure generally converges with slow rate and may suffer from numerical instability.

An alternative way of constructing a sparse kernel model from the full model (12) is the OFS procedure based on Fisher ratio class separability measure [9], which is computationally attractive and numerically robust. Define the modeling residual as $\epsilon(k) = d(k) - y(k) = b_1(k) - y(\mathbf{x}(k))$. Then, the kernel model (12) over the training dataset can be collected together as

$$\mathbf{d} = \mathbf{\Phi}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{14}$$

where the target vector $\mathbf{d} = [d(1), \ldots, d(N)]^T = [b_1(1), \ldots, b_1(N)]^T$, the regression matrix $\mathbf{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N]$ with

$$\begin{aligned} \boldsymbol{\phi}_i &= [\phi_i(1), \ldots, \phi_i(N)]^T \\ &= [\phi(\mathbf{x}(1), \mathbf{x}(i)), \ldots, \phi(\mathbf{x}(N), \mathbf{x}(i))]^T \end{aligned} \tag{15}$$

for $1 \le i \le N$, the kernel weight vector $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_N]^T$, and the residual vector $\boldsymbol{\epsilon} = [\epsilon(1), \ldots, \epsilon(N)]^T$. Let an orthogonal decomposition of the regression matrix $\mathbf{\Phi}$ be $\mathbf{\Phi} = \mathbf{V}\mathbf{A}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & \alpha_{1,2} & \ldots & \alpha_{1,N} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \alpha_{N-1,N} \\ 0 & \ldots & 0 & 1 \end{bmatrix} \tag{16}$$

$$\mathbf{V}=[\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_N]=\begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,N} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ v_{N,1} & v_{N,2} & \cdots & v_{N,N} \end{bmatrix} \quad (17)$$

with orthogonal columns that satisfy $\mathbf{v}_i^T\mathbf{v}_q = 0$ if $i \neq q$. The kernel model (14) can alternatively be expressed as

$$\mathbf{d} = \mathbf{V}\mathbf{g} + \boldsymbol{\epsilon} \quad (18)$$

where $\mathbf{g}=[g_1,\dots,g_N]^T$ satisfies the triangular system $\mathbf{A}\boldsymbol{\beta}=\mathbf{g}$.

A sparse N_{spa}-term model can be selected by incrementally maximizing a class separability measure in an OFS procedure [9]. Define the two class sets $\mathcal{C}_\pm = \{\mathbf{x}(k) : d(k) = \pm 1\}$, and let the numbers of points in $\mathcal{C}_\pm$ be $N_\pm$, respectively, with $N_+ + N_- = N$. The means and variances of training samples belonging to classes $\mathcal{C}_\pm$ in the direction of basis $\mathbf{v}_l$ are given by

$$\begin{aligned} m_{+,l} &= \frac{1}{N_+}\sum_{i=1}^{N}\delta(d(i)-1)v_{i,l} \\ \sigma_{+,l}^2 &= \frac{1}{N_+}\sum_{i=1}^{N}\delta(d(i)-1)(v_{i,l}-m_{+,l})^2 \end{aligned} \quad (19)$$

$$\begin{aligned} m_{-,l} &= \frac{1}{N_-}\sum_{i=1}^{N}\delta(d(i)+1)v_{i,l} \\ \sigma_{-,l}^2 &= \frac{1}{N_-}\sum_{i=1}^{N}\delta(d(i)+1)(v_{i,l}-m_{-,l})^2 \end{aligned} \quad (20)$$

respectively, where $\delta(x) = 1$ for $x = 0$ and $\delta(x) = 0$ for $x \neq 0$. The Fisher ratio, defined as the ratio of the interclass difference and the intraclass spread, in the direction of $\mathbf{v}_l$ is given by [11]

$$F_l = \frac{(m_{+,l}-m_{-,l})^2}{\sigma_{+,l}^2+\sigma_{-],l}^2}. \quad (21)$$

Base on this Fisher ratio, significant kernel terms can be selected in an OFS procedure. At the lth stage, a term is chosen as the lth term in the selected model if it produces the largest F_l among the candidate terms $\mathbf{v}_i, l \le i \le N$. The procedure is terminated with a sparse N_{spa}-term model when

$$\frac{F_{N_{\text{spa}}}}{\sum_{l=1}^{N_{\text{spa}}} F_l} < \xi \quad (22)$$

where the threshold ξ determines the sparsity of the selected model. We have found out empirically that the appropriate values for ξ is in the range of 0.005–0.01. The least square solution for the corresponding sparse model weight vector $\boldsymbol{\beta}_{N_{\text{spa}}}$ is readily available given the least square solution of $\mathbf{g}_{N_{\text{spa}}}$.

The modified Gram–Schmidt orthogonalization procedure [12] is first summarized. Denote $\boldsymbol{\phi}_i^{(0)} = \boldsymbol{\phi}_i, 1 \le i \le N$. For $l = 1,2,\dots,N-1$

$$\left.\begin{array}{l} \mathbf{v}_l = \boldsymbol{\phi}_l^{(l-1)}, \\ a_{l,i} = \mathbf{v}_l^T\boldsymbol{\phi}_i^{(l-1)} \big/ \left(\mathbf{v}_l^T\mathbf{v}_l\right), \quad l+1 \le i \le N \\ \boldsymbol{\phi}_i^{(l)} = \boldsymbol{\phi}_i^{(l-1)} - a_{l,i}\mathbf{v}_l, \quad l+1 \le i \le N \end{array}\right\}. \quad (23)$$

The last stage is simply $\mathbf{v}_N = \boldsymbol{\phi}_N^{(N-1)}$. The elements of $\mathbf{g}$ are computed by transforming $\mathbf{d}^{(0)} = \mathbf{d}$ in a similar way

$$\left.\begin{array}{l} g_l = \mathbf{v}_l^T\mathbf{d}^{(l-1)} \big/ \left(\mathbf{v}_l^T\mathbf{v}_l\right) \\ \mathbf{d}^{(l)} = \mathbf{d}^{(l-1)} - g_l\mathbf{v}_l \end{array}\right\} 1 \le l \le N. \quad (24)$$

Next, define $\boldsymbol{\Phi}^{(l-1)} = [\mathbf{v}_1,\dots,\mathbf{v}_{l-1},\boldsymbol{\phi}_l^{(l-1)},\dots,\boldsymbol{\phi}_N^{(l-1)}]$ and give a very small positive number T_z. With the notation $\boldsymbol{\phi}_q^{(l-1)} = [\phi_{1,q}^{(l-1)},\phi_{2,q}^{(l-1)},\dots,\phi_{N,q}^{(l-1)}]^T$, the lth stage of the selection procedure is given as follows.

Step 1) For $l \le q \le N$:

Test: Conditioning number check. If $(\boldsymbol{\phi}_q^{(l-1)})^T\boldsymbol{\phi}_q^{(l-1)} < T_z$, the qth candidate is not considered.

Compute the following:

$$\begin{aligned} m_{+,l}^{(q)} &= \frac{1}{N_+}\sum_{i=1}^{N}\delta(d(i)-1)\phi_{i,q}^{(l-1)}, \\ \left(\sigma_{+,l}^{(q)}\right)^2 &= \frac{1}{N_+}\sum_{i=1}^{N}\delta(d(i)-1)\left(\phi_{i,q}^{(l-1)}-m_{+,l}^{(q)}\right)^2 \\ m_{-,l}^{(q)} &= \frac{1}{N_-}\sum_{i=1}^{N}\delta(d(i)+1)\phi_{i,q}^{(l-1)} \\ \left(\sigma_{-,l}^{(q)}\right)^2 &= \frac{1}{N_-}\sum_{i=1}^{N}\delta(d(i)+1)\left(\phi_{i,q}^{(l-1)}-m_{-,l}^{(q)}\right)^2 \\ F_l^{(q)} &= \frac{\left(m_{+,l}^{(q)}-m_{-,l}^{(q)}\right)^2}{\left(\sigma_{+,l}^{(q)}\right)^2+\left(\sigma_{-,l}^{(q)}\right)^2}. \end{aligned}$$

Let the index set $\mathcal{J}_q$ be: $\mathcal{J}_q = \{l \le q \le N$ and q passes **Test**$\}$.

Step 2) Find: $F_l = F_l^{(q_l)} = \max\{F_l^{(q)}, q \in \mathcal{J}_q\}$.

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

Step 3) Perform the orthogonalization as indicated in (23) to derive the lth row of $\mathbf{A}$ and to transform $\boldsymbol{\Phi}^{(l-1)}$ into $\boldsymbol{\Phi}^{(l)}$. Calculate g_l and update $\mathbf{d}^{(l-1)}$ into $\mathbf{d}^{(l)}$ in the way shown in (24).

V. Simulation Example

The example consisted of four signal sources and a two-element antenna array. Fig. 1 shows the locations of the desired source and three interfering sources graphically. The simulated channel conditions were $A_i = 1 + j0, 1 \le i \le 4$, and all the four users had equal signal power. The minimum spatial separation was the difference in angles of arrival between the desired user 1 and the interferer 2, which was $\theta \le 30°$. Fig. 2 compares the BERs of the LMMSE, LMBER, and Bayesian beamformers for the two cases of $\theta = 30°$ and $\theta = 10°$, respectively. It is seen from Fig. 2(a) that for $\theta = 30°$, the LMMSE beamformer could not achieve linear separability and exhibited a high BER floor, but the LMBER beamformer achieved linear separability and

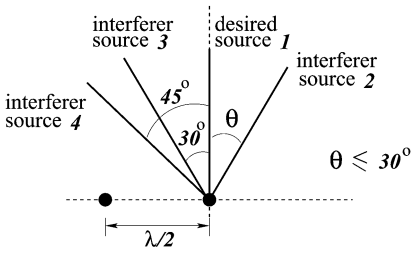

Fig. 1. Locations of the desired and interfering sources with respect to the two-element linear antenna array having $\lambda/2$ spacing, where λ is the wavelength.

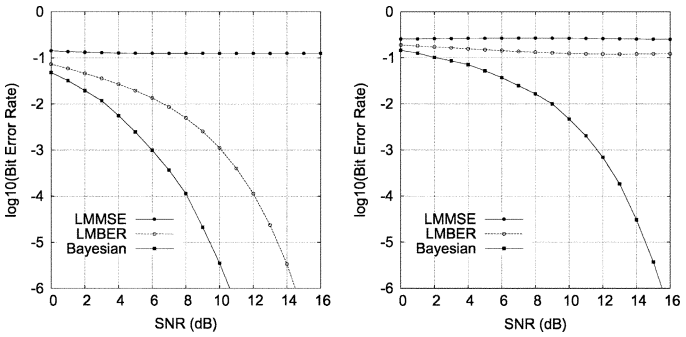

Fig. 2. Comparison of the bit error rates of three theoretical beamformers. (a) $\theta = 30°$. (b) $\theta = 10°$.

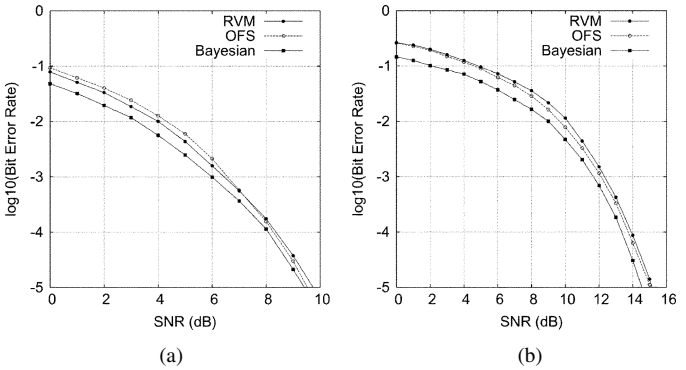

Fig. 3. Performance comparison of the Bayesian beamformer with the RBF beamformers constructed by the RVM algorithm and the OFS with the Fisher ratio, respectively. (a) $\theta = 30°$. (b) $\theta = 10°$.

had a much better BER performance than the LMMSE beamformer. The Bayesian beamformer provided the optimal performance and had a 4-dB improvement in SNR at the BER level of 10^{-3}, compared with the LMBER beamformer. When the spatial separation was reduced to $\theta = 10°$, the system became inherently linearly inseparable, and while the linear beamformer failed in this situation, the Bayesian beamformer still performed adequately. This is demonstrated clearly in Fig. 2(b).

The OFS algorithm with the Fisher ratio and the RVM algorithm were used to construct a RBF beamformer. The number of training data for each given SNR was $N = 160$. The Gaussian kernel variance ρ^2 was determined empirically, and the appropriate values for ρ^2 were found to be in the range of $2\sigma_n^2$ to $10\sigma_n^2$, depending on the SNR. The numbers of RBF centers or kernel terms identified by the two algorithms over the given SNR values were similar, ranging from $N_{\rm spa} = 14$ to 20 with the typical value of $N_{\rm spa} = 18$. The BERs of the RVM and OFS beamformers are compared in Fig. 3. It can be seen that both kernel-based beamformers have similarly good performance with similar model sparsity. The OFS algorithm based on the Fisher ratio, however, has considerable computational and numerical advantages during the construction process.

VI. Conclusion

The optimal nonlinear beamforming assisted receiver has been derived, and it has been shown that this optimal Bayesian beamformer outperforms the linear beamformer significantly in terms of a reduced bit error rate. This demonstrates the potential of system capacity enhancement by employing nonlinear beamforming. Block-data kernel-based adaptive implementation of the optimal Bayesian beamformer is investigated using the OFS algorithm based on the Fisher ratio for class separability measure. Empirical results have demonstrated that this construction algorithm has excellent performance similar to that of the RVM algorithm, but it is computationally much simpler and numerically much more robust.

References

[1] J. H. Winters, J. Salz, and R. D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Trans. Commun.*, vol. 42, pp. 1740–1751, Feb./Mar./Apr. 1994.

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

[3] L. C. Godara, "Applications of antenna arrays to mobile communications, part I: Performance improvement, feasibility, and system considerations," *Proc. IEEE*, vol. 85, pp. 1031–1060, July 1997.

[4] J. H. Winters, "Smart antennas for wireless systems," *IEEE Personal Commun.*, vol. 5, no. 1, pp. 23–27, 1998.

[5] J. S. Blogh and L. Hanzo, *Third Generation Systems and Intelligent Wireless Networking—Smart Antenna and Adaptive Modulation*. New York: Wiley, 2002.

[6] S. Chen, L. Hanzo, and N. N. Ahmad, "Adaptive minimum bit error rate beamforming assisted receiver for wireless communications," in *Proc. ICASSP*, Hong Kong, China, Apr. 6–10, 2003, pp. 640–643.

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

[8] S. Chen, S. R. Gunn, and C. J. Harris, "The relevance vector machine technique for channel equalization application," *IEEE Trans. Neural Networks*, vol. 12, pp. 1529–1532, Nov. 2001.

[9] K. Z. Mao, "RBF neural network center selection based on Fisher ratio class separability measure," *IEEE Trans. Neural Networks*, vol. 13, pp. 1211–1217, Sept. 2002.

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

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

[12] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to nonlinear system identification," *Int. J. Control*, vol. 50, no. 5, pp. 1873–1896, 1989.