Block-Adaptive Kernel-Based CDMA Multiuser Detection

S. Chen and L. Hanzo
Department of Electronics and Computer Science
University of Southampton, Southampton SO17 1BJ, U.K.

***Abstract*— The paper investigates the application of a recently introduced learning technique, referred to as the relevance vector machine (RVM) to construct a block-adaptive kernel-based nonlinear multiuser detector (MUD) for direct-sequence code-division multiple-access (DS-CDMA) signals transmitted through multipath channels. It is demonstrated that the RVM MUD is capable of closely matching the performance of the optimal Bayesian one-shot detector, with the aid of a significantly more sparse kernel representation than that required by the state-of-the-art support vector machine (SVM) technique.**

I. INTRODUCTION

Although the linear minimum mean square error (MMSE) MUD [1]–[5] is widely used for DS-CDMA downlink systems due to its simplicity, its limitation has long been recognized – a linear detector results in a residual bit error ratio (BER), unless the underlying noise-free signal classes are linearly separable. However, since linearly non-separable cases are common in DS-CDMA channels, often a better performance can be obtained by using a nonlinear MUD. Hence neural networks [6] have been considered as nonlinear MUDs [7]–[10]. However, the training period required by these nonlinear MUDs may become excessive and/or unpredictable. Furthermore, the structure of these neural network aided MUDs is often *ad hoc*.

In our previous work [11],[12] the SVM technique [13]–[15] has been applied for constructing kernel-based MUDs. Our study has shown that an SVM aided MUD trained using a relatively small block of noisy received signal samples may closely approximate the performance of the optimal MUD, although the latter requires a complete knowledge of the system, namely that of the system matrix $\mathbf{P}$ and the noise variance. Another advantage of the SVM approach over the existing nonlinear MUDs is the direct definition of the detector's structure, which is specified by a sparse set of support vectors (SVs) selected automatically from the data during the learning process. However, as the results reported in [11],[12] show, when applied to the MUD problem, the SVM technique does not produce a sufficiently sparse model in the sense that a typical SVM aided MUD will have 2 to 8 times more kernels, than the number of the noise-free signal states that is required by the optimum Bayesian detector.

Recently, Tipping [16] introduced a RVM method, which is based on a Bayesian framework [17],[18] and has an identical functional form to that of the SVM. The results given in [16] have demonstrated that the RVM has a comparable generalization performance to that of the SVM, while requiring significantly less kernel functions than the SVM. This paper investigates the application of the RVM technique to the construction of a block-adaptive kernel-based MUD. The computer simulation results confirm that the RVM assisted MUD is capable of closely matching the optimal Bayesian performance, and it exhibits a significantly sparser kernel representation than the SVM aided MUD. More specifically, an RVM assisted MUD typically has fewer kernels functions, than the number of noise-free signal states. That is, it can be typically described with the aid of a sparser representation, than the optimal Bayesian detector. The main drawback of the RVM method is, however that it involves a highly nonlinear optimization process. Hence the RVM assisted MUD's performance has to be compared to that of the SVM technique, which is required to solve a significantly simpler quadratic problem.

II. SYSTEM MODEL

The discrete-time model of the synchronous downlink DS-CDMA system supporting N users and transmitting M ($> N$) chips per bit is depicted in Fig. 1, where $b_i(k) \in \{\pm 1\}$ denotes the k-th bit of user i. Furthermore, the unit-length signature code sequence of user i is $\bar{\mathbf{s}}_i = [\bar{s}_{i,1} \cdots \bar{s}_{i,M}]^T$, and the z-domain transfer function associated with the channel impulse response (CIR) is given by:

$$H(z) = \sum_{i=0}^{n_h-1} h_i z^{-i}. \tag{1}$$

The bit vector of N users at instant k is $\mathbf{b}(k) = [b_1(k) \cdots b_N(k)]^T$, and the received signal vector after the chip-matched filters is $\mathbf{r}(k) = [r_1(k) \cdots r_N(k)]^T$. It can be shown that the baseband model for $\mathbf{r}(k)$ is:

$$\mathbf{r}(k) = \mathbf{P} \begin{bmatrix} \mathbf{b}(k) \\ \mathbf{b}(k-1) \\ \vdots \\ \mathbf{b}(k-L+1) \end{bmatrix} + \tilde{\mathbf{n}}(k), \tag{2}$$

where the $N \times LN$ system matrix is given by

$$\mathbf{P} = \bar{\mathbf{S}}^T \mathbf{H} \begin{bmatrix} \bar{\mathbf{S}}\mathbf{U} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{S}}\mathbf{U} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \bar{\mathbf{S}}\mathbf{U} \end{bmatrix}; \tag{3}$$

the user signature sequence matrix is $\bar{\mathbf{S}} = [\bar{\mathbf{s}}_1 \cdots \bar{\mathbf{s}}_N]$; the diagonal user signal amplitude matrix is $\mathbf{U} = \mathrm{diag}\{U_1 \cdots U_N\}$; the $M \times LM$ CIR matrix $\mathbf{H}$ has the form of:

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 & \cdots & h_{n_h-1} & & & \\ & h_0 & h_1 & \cdots & h_{n_h-1} & & \\ & & \ddots & \ddots & \cdots & \ddots & \\ & & & h_0 & h_1 & \cdots & h_{n_h-1} \end{bmatrix}; \tag{4}$$

0-7803-7400-2/02/$17.00 © 2002 IEEE

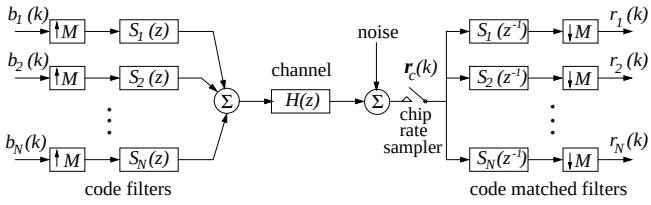

Fig. 1. Discrete-time model of the synchronous CDMA downlink.

and orthogonal code sequences are assumed, so that the noise vector $\tilde{\mathbf{n}}(k) = [\tilde{n}_1(k) \cdots \tilde{n}_N(k)]^T$ at the outputs of the chip-matched filters $\tilde{\mathbf{n}}(k) = [\tilde{n}_1(k) \cdots \tilde{n}_N(k)]^T$ has a covariance of $E[\tilde{\mathbf{n}}(k)\tilde{\mathbf{n}}^T(k)] = \sigma_n^2\mathbf{I}$. We note that the orthogonality of the codes is destroyed by the channel-induced intersymbol interference (ISI). The ISI span L depends on the length n_h of the CIR, expressed in terms of the number of chips M per spreding code. For $n_h = 1$ we have $L = 1$; for $1 < n_h \leq M$, $L = 2$; for $M < n_h \leq 2M$, $L = 3$; and so on.

III. Linear and optimal detectors

The linear MUD of user i has the form:

$$\hat{b}_i(k) = \text{sgn}(y_L(k)) \text{ with } y_L(k) = \mathbf{w}^T\mathbf{r}(k), \tag{5}$$

where $\mathbf{w} = [w_1 \cdots w_N]^T$ denotes the MUD's weight vector. The most popular solution for the MUD of (5) is the MMSE solution given by

$$\mathbf{w}_{MMSE} = \left(\sigma_n^2\mathbf{I} + \mathbf{P}\mathbf{P}^T\right)^{-1}\mathbf{p}_i, \tag{6}$$

where $\mathbf{p}_i$ denotes the i-th column of $\mathbf{P}$. The linear MUD of (5) has a low computational complexity, and the standard LMS or RLS algorithms can be used for implementing the MMSE solution adaptively.

However, a linear MUD only performs adequately in certain situations. Let the $N_b = 2^{LN}$ possible combinations of $[\mathbf{b}^T(k)\ \mathbf{b}^T(k-1) \cdots \mathbf{b}^T(k-L+1)]^T$ be

$$\mathbf{b}^{(j)} = \begin{bmatrix} \mathbf{b}^{(j)}(k) \\ \mathbf{b}^{(j)}(k-1) \\ \vdots \\ \mathbf{b}^{(j)}(k-L+1) \end{bmatrix}, \ 1 \leq j \leq N_b, \tag{7}$$

and $b_i^{(j)}$ be the ith element of $\mathbf{b}^{(j)}(k)$. Let us furthermore define the set of the N_b noise-free received signal states as:

$$\mathcal{R} = \{\mathbf{r}_j = \mathbf{P}\mathbf{b}^{(j)},\ 1 \leq j \leq N_b\}. \tag{8}$$

In case of ninary transmission $\mathcal{R}$ can be partitioned into two subsets:

$$\mathcal{R}_\pm = \{\mathbf{r}_j \in \mathcal{R} : b_i^{(j)} = \pm 1\}. \tag{9}$$

If $\mathcal{R}_-$ and $\mathcal{R}_+$ are not linearly separable, a linear MUD will have an irreducible error floor even in the noise-free case, as it can only form a decision hyperplane in the N-dimensional received signal space. It was demomstrated with the aid of an example in [6] that this error floor can be potentially removed, if the decisions are cast into a higher-dimensional space.

Applying the maximum *a posteriori* probability (MAP) or Bayesian classification theory in a manner similar to the channel equalization problem [19], it can be shown that the optimal detector has the form:

$$y_B(k) = \sum_{j=1}^{N_b} \beta_j b_i^{(j)} \exp\left(-\frac{\|\mathbf{r}(k) - \mathbf{r}_j\|^2}{2\sigma_n^2}\right) \tag{10}$$

with

$$\hat{b}_i(k) = \text{sgn}(y_B(k)), \tag{11}$$

where $b_i^{(j)} \in \{\pm 1\}$ serve as class labels, and all the channel states are assumed to be equiprobable with $\beta_j = \frac{1}{N_b(2\pi\sigma_n^2)^{\frac{m}{2}}}$.

IV. The relevance vector machine detector

The optimal detector of (10) requires the knowledge of all the noise-free signal sates $\mathbf{r}_j$, which are unknown to receiver i. In general, the receiver can have access to a block of K training samples $\{\mathbf{x}_k = \mathbf{r}(k), t_k = b_i(k)\}_{k=1}^K$. Consider the kernel-based detector of user i in the form of:

$$y(\mathbf{r}(k)) = \sum_{l=1}^{K} w_l F_l(\mathbf{r}(k)), \tag{12}$$

where w_l are the "weights" and $F_l(\mathbf{r}(k)) = F(\mathbf{r}(k), \mathbf{x}_l)$. Observe that instead of the N-dimensional weight vector of the linear detector of 5, here a K-dimensional weight-vector is used. For this application, the kernel function $F(\cdot,\cdot)$ is naturally chosen to be a Gaussian function, with its variance being an estimate of the channel's noise variance. The relevance vector (RV) approach of classification [16] can readily be applied for constructing the detector (12). Denote the K-dimensional vector of previously defined training samples $\{\mathbf{x}_k = \mathbf{r}(k), t_k = b_i(k)\}_{k=1}^K$ by $\mathbf{t} = [t_1 \cdots t_K]^T$ and the weight vector by $\mathbf{w} = [w_1 \cdots w_K]^T$. The posterior probability of $\mathbf{w}$ is

$$p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}) = \frac{p(\mathbf{t}|\mathbf{w}, \boldsymbol{\alpha})p(\mathbf{w}|\boldsymbol{\alpha})}{p(\mathbf{t}|\boldsymbol{\alpha})}, \tag{13}$$

where $p(\mathbf{w}|\boldsymbol{\alpha})$ is the *a priori* probability of the weight vector $\mathbf{w}$ conditioned on $\boldsymbol{\alpha} = [\alpha_1 \cdots \alpha_K]^T$ denoting the vector of hyper-parameters, which is a term well established in statistical decision theory [17], $p(\mathbf{t}|\mathbf{w}, \boldsymbol{\alpha})$ is the so-called likelihood [17] and $p(\mathbf{t}|\boldsymbol{\alpha})$ the evidence [17]. Following the Bayesian classification framework [18], the likelihood can be expressed as

$$p(\mathbf{t}|\mathbf{w}, \boldsymbol{\alpha}) = \prod_{l=1}^{K} \left(f(y(\mathbf{x}_l))\right)^{t_l} \left(1 - f(y(\mathbf{x}_l))\right)^{1-t_l}, \tag{14}$$

where

$$f(x) = \frac{1}{1 + \exp(-x)} \tag{15}$$

is the logistic sigmoid function. The Gaussian *a priori* probability is chosen in the form of:

$$p(\mathbf{w}|\boldsymbol{\alpha}) = \prod_{l=1}^{K} \frac{\sqrt{\alpha_l}}{\sqrt{2\pi}} \exp\left(-\frac{\alpha_l w_l^2}{2}\right). \tag{16}$$

Since the so-called marginal likelihood $p(\mathbf{t}|\boldsymbol{\alpha})$ cannot be obtained analytically by integrating out the weights from (14), an iterative procedure is necessitated [18].

With a given fixed $\boldsymbol{\alpha}$, the MAP solution $\mathbf{w}_{\rm MAP}$ can be obtained by maximizing $\log(p(\mathbf{w}|\mathbf{t},\boldsymbol{\alpha}))$ or, equivalently, by minimizing the following cost function

$$J(\mathbf{w}|\mathbf{t},\boldsymbol{\alpha}) = \sum_{l=1}^{K}\{\frac{\alpha_l w_l^2}{2} - t_l \log(f(y(\mathbf{x}_l))) - (1-t_l)\log(1-f(y(\mathbf{x}_l)))\}\,. \quad (17)$$

The gradient of J with respect to $\mathbf{w}$ is given by:

$$\nabla J = \mathbf{A}\mathbf{w} + \boldsymbol{\Phi}^T(\mathbf{f}-\mathbf{t})\,, \quad (18)$$

where $\mathbf{A} = \text{diag}\{\alpha_1,\cdots,\alpha_K\}$, $\mathbf{f} = [f(y(\mathbf{x}_1))\cdots f(y(\mathbf{x}_K))]^T$ and the matrix $\boldsymbol{\Phi}$ has elements $\phi_{i,j} = F(\mathbf{x}_i,\mathbf{x}_j)$. The Hessian of J is

$$\mathbf{H} = \nabla^2 J = \boldsymbol{\Phi}^T\mathbf{B}\boldsymbol{\Phi} + \mathbf{A}\,, \quad (19)$$

where $\mathbf{B} = \text{diag}\{f(y(\mathbf{x}_1))(1-f(y(\mathbf{x}_1))),\cdots,f(y(\mathbf{x}_K))(1-f(y(\mathbf{x}_K)))\}$.

The *a posteriori* set of weights is approximated around $\mathbf{w}_{\rm MAP}$ with the aid of a Gaussian approximation having a covariance of:

$$\boldsymbol{\Lambda} = (\mathbf{H}|_{\mathbf{w}_{\rm MAP}})^{-1} \quad (20)$$

and the mean of:

$$\boldsymbol{\mu} = [\mu_1\cdots\mu_K]^T = \boldsymbol{\Lambda}\left(\boldsymbol{\Phi}^T\mathbf{B}\mathbf{t}|_{\mathbf{w}_{\rm MAP}}\right)\,. \quad (21)$$

The hyperparameters $\boldsymbol{\alpha}$ are updated using [17]

$$\alpha_i^{\rm new} = \frac{1-\alpha_i^{\rm old}\lambda_{i,i}}{\mu_i^2} \quad (22)$$

with $\lambda_{i,i}$ being the diagonal elements of $\boldsymbol{\Lambda}$.

The introduction of an individual hyperparameter for every weight of the model (12) is the key feature of the RVM aided approach, and it is ultimately responsible for its attractively low number of kernels [16]. During the optimization process, many of the α_i hyperparameters are driven to large values and hence the corresponding model weights w_i are effectively pruned out. Thus the corresponding model terms $F_i(\cdot)$ can be removed from the trained model represented by (12). The simple iterative procedure that we adopt for constructing a RVM aided MUD is summarized as follows:

Initialization. The $K \times n_R$ kernel matrix $\boldsymbol{\Phi}$ is initialized with $n_R = K$, i.e. every training data point is considered as a candidate kernel. Each weight w_i is initially associated with an identical value of the hyperparameter α_i.

Step 1. Given the current value $\boldsymbol{\alpha}$, find $\mathbf{w}_{\rm MAP}$ by minimizing the cost function of (17). A simplified conjugate gradient algorithm [20] is used in the optimization. Alternatively, the iteratively-re-weighted least-square algorithm [21] can be used.

Step 2. The hyperparameters are updated using (22). If we have $\alpha_i > L_g$, where L_g is a preset large positive value, we assign $n_R := n_R - 1$, and the corresponding column in $\boldsymbol{\Phi}$ is removed. Thus the corresponding weight w_i and model term $F_i(\cdot)$ is pruned out the model.

Test. If the hyperparameters $\boldsymbol{\alpha}$ remain sufficiently unchanged in two successive iterations or a pre-set maximum number of iterations is reached, stop; otherwise go to *Step 1*.

The set of RVs $\{\mathbf{x}_l\}_{l=1}^{n_R}$ selected is typically a small subset of the training points. The RVM aided MUD thus computes the decision variable as follows:

$$y(\mathbf{r}(k)) = \sum_{l=1}^{n_R} w_l F_l(\mathbf{r}(k)) \quad (23)$$

and carries out the decision according to:

$$\hat{b}_i(k) = \text{sgn}(y(\mathbf{r}(k)))\,. \quad (24)$$

V. Simulation results

Two simulation examples were used for comparing the performance of the proposed RVM aided MUD to that of the linear MMSE, optimal Bayesian and the SVM assisted MUDs. *It is worth pointing out again that the linear MMSE and the optimal MUDs are designed based on the complete knowledge of the system (the system matrix* $\mathbf{P}$ *and the noise variance), while the SVM and RVM aided MUDs are trained using a block of the noisy received signal samples.*

Example 1. This was a two-user system with 4 chips per bit. The code sequences of the two users were $(+1,+1,-1,-1)$ and $(+1,-1,-1,+1)$, respectively, and the transfer function associated with the CIR was $H(z) = 0.3 + 0.7z^{-1} + 0.3z^{-2}$. The two users had equal signal power, that is, the signal to noise ratio SNR_1 of user 1 was equal to SNR_2 of user 2. In order to construct a kernel-based MUD for user 2, a total of 160 training data points were generated for each given noise variance. The number of SVs selected by the SVM method is influenced by the control parameter C, which provides a trade-off between the model's complexity (the number of SVs) and the training error [13]. The appropriate value for C was found in the simulations experimentally. When using $C = 8.0$, the number of SVs was found typically to be around 40. For the RVM method, there was no need to specify such a control parameter, and the numbers of RVs found ranged from 6 (for low SNRs) to 18 (for high SNRs).

Table I summarizes the results obtained using the SVM and RVM methods, respectively, in comparison to the optimal Bayesian MUD, given $\text{SNR}_1 = \text{SNR}_2 = 20$ dB. Fig. 2 depicts the typical decision boundaries of the SVM and RVM MUDs, respectively, together with that of the optimal Bayesian decision boundary. It is clear that for user 2 $\mathcal{R}_-$ and $\mathcal{R}_+$ are not linearly separable and the linear MMSE detector will have an irreducible error floor of 0.125, as can be seen in Fig. 3, where the BERs of the optimal and the RVM aided MUDs are also shown. Compared to the results given in [11], it can be seen that the RVM aided MUD has a similar performance to that of the SVM assisted MUD. However, it requires a significantly lower number of kernels, than the SVM assisted MUD for closely approximating the optimal MUD's performance.

TABLE I

BER PERFORMANCE AND NUMBER OF KERNELS USED BY VARIOUS MUDS FOR USER 2 OF EXAMPLE 1, GIVEN $\text{SNR}_1 = \text{SNR}_2 = 20$ DB. THE SVM USED $C = 8.0$.

model	$\log_{10}$(BER)	kernels
SVM	-3.019	40
RVM	-3.045	18
Bayesian	-3.155	16

Example 2. This was a 3-user system employing 8 chips per bit. The code sequences for the three users were $(+1,+1,+1,+1,-1,-1,-1,-1)$, $(+1,-1,+1,-1,-1,+1,-1,+1)$ and $(+1,-1,-1,+1,-1,+1,+1,-1)$, respectively, and the z-domain transfer function associated with the CIR was $H(z) = 0.5 + 1.0z^{-1} - 0.5z^{-2}$. The three users had equal signal power. The number of training data used for constructing the kernel-based MUDs was 640 for each given SNR. For user 3, typically 200 SVs were selected from the training data set, while the number of RVs selected ranged from 14 (for low SNRs) to 38 (for high SNRs). Table II summarizes the results for three MUDs, given $\text{SNR}_i = 15$ dB, $1 \le i \le 3$. The BERs of the resultant RVM aided MUDs of user 3 experienced under different SNR conditions are given in Fig. 4, in comparison to the corresponding linear MMSE and optimal MUDs. The results again demonstrate that the RVM aided MUD is capable of closely approximate the performance of the optimal detector using a low number of kernels. This reduces the complexity of classifying the received signal vector into one of the legitimate classes, which is ultimately required for carrying out a binary decision.

VI. CONCLUSIONS

The RVM technique has been applied for adaptive nonlinear multiuser detection in DS-CDMA systems. It has been shown that the RVM MUD trained using noisy data is capable of closely approximating the performance of the optimal Bayesian one-shot detector using less kernels than the Bayesian MUD. Compared to the SVM aided technique, the RVM assisted MUD results in a less complex received vector classification in the MUD. A disadvantage of the RVM aided MUD is, however that it requires solving a more complex nonlinear optimization problem. Like the SVM method, the RVM method is a block-based technique. Future research is required for investigating how to incorporate a sample-by-sample

TABLE II

BER PERFORMANCE AND NUMBER OF KERNELS USED BY VARIOUS MUDS FOR USER 3 OF EXAMPLE 2, GIVEN $\text{SNR}_i = 15$ DB, $1 \le i \le 3$. THE SVM USED $C = 5.0$.

model	$\log_{10}$(BER)	kernels
SVM	-2.452	199
RVM	-2.489	30
Bayesian	-2.780	64

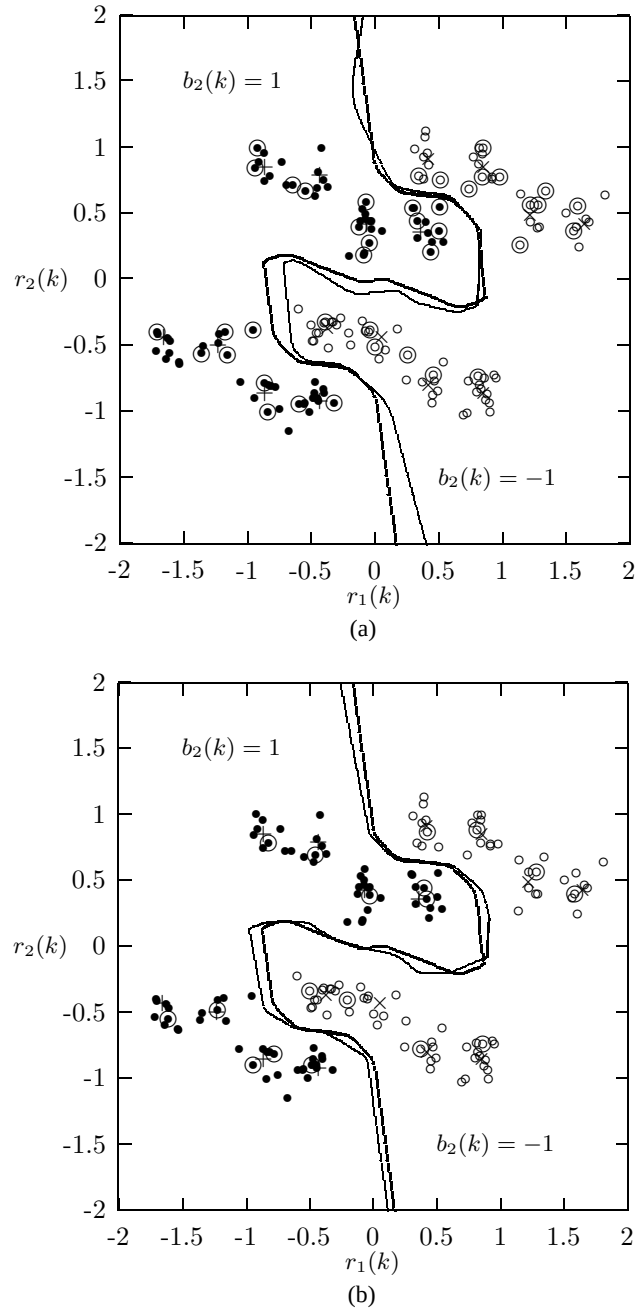

Fig. 2. Comparison of the optimal Bayesian decision boundary (thick solid) with those (thin solid) of the SVM MUD (a) and the RVM MUD (b) for user 2 of Example 1, given $\text{SNR}_1 = \text{SNR}_2 = 20$ dB. The $\times$ and $+$ are the two classes of the noise-free signal states, small circles and dishes are the two classes of the training data, respectively, and big circles denote SVs (a) and RVs (b).

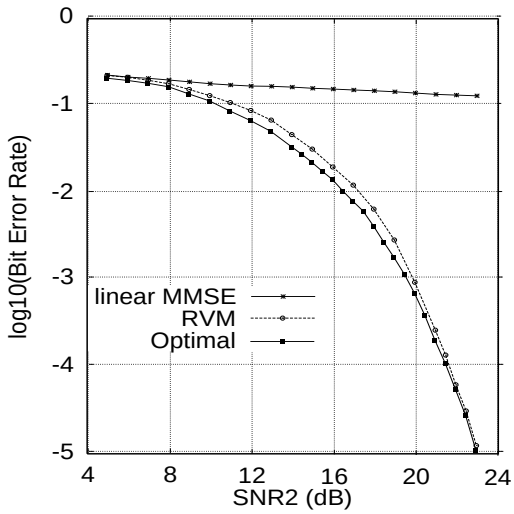

Fig. 3. Performance comparison of three MUDs, linear MMSE, adaptive RVM and optimal detectors, for user 2 of Example 1. $\text{SNR}_1 = \text{SNR}_2$, and the training data set for RVM had 160 samples. The numbers of RVs found ranged from 6 (for low SNRs) to 18 (for high SNRs).

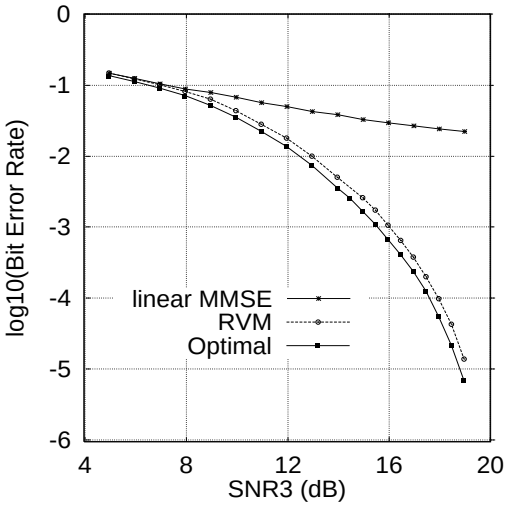

Fig. 4. Performance comparison of three MUDs, linear MMSE, adaptive RVM and optimal detectors, for user 3 of Example 2. SNR_i, $1 \leq i \leq 3$, were identical, and the training data set for RVM had 640 samples. The numbers of RVs found ranged from 14 (for low SNRs) to 38 (for high SNRs).

adaptive methodology into the RVM approach.

REFERENCES

[1] Z. Xie, R.T. Short and C.K. Rushforth, "A family of suboptimum detectors for coherent multiuser communications," *IEEE J. Selected Areas in Communications*, Vol.8, No.4, pp.683–690, 1990.

[2] U. Madhow and M.L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Communications*, Vol.42, No.12, pp.3178–3188, 1994.

[3] S.L. Miller, "An adaptive direct-sequence code-division multiple-access receiver for multiuser interference rejection," *IEEE Trans. Communications*, Vol.43, No.2/3/4, pp. 1746–1755, 1995.

[4] H.V. Poor and S. Verdú, "Probability of error in MMSE multiuser detection," *IEEE Trans. Information Theory*, Vol.43, No.3, pp.858–871, 1997.

[5] G. Woodward and B.S. Vucetic, "Adaptive detection for DS-CDMA," *Proc. IEEE*, Vol.86, No.7, pp.1413–1434, 1998.

[6] L. Hanzo, C.H. Wong, M.S. Yee: Adaptive wireless transceivers: Turbo-Coded, Turbo-Equalised and Space-Time Coded TDMA, CDMA and OFDM systems, John Wiley, in press

[7] B. Aazhang, B.P. Paris and G.C. Orsak, "Neural networks for multiuser detection in code-division multiple-access communications," *IEEE Trans. Communications*, Vol.40, No.7, pp.1212–1222, 1992.

[8] U. Mitra and H.V. Poor, "Neural network techniques for adaptive multiuser demodulation," *IEEE J. Selected Areas in Communications*, Vol.12, No.9, pp.1460-1470, 1994.

[9] D.G.M. Cruickshank, "Radial basis function receivers for DS-CDMA," *Electronics Letters*, Vol.32, No.3, pp.188-190, 1996.

[10] R. Tanner and D.G.M. Cruickshank, "Volterra based receivers for DS-CDMA," in *Proc. 8th IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications*, September 1997, Vol.3, pp.1166-1170.

[11] S. Chen, A.K. Samingan and L. Hanzo, "Adaptive multiuser receiver using support vector machine technique," Proceedings of *VTC 2001 Spring Conf.* (Rhodes, Greece), May 6-9 2001, pp 604-608

[12] 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, May 2001, pp 604-611

[13] V. Vapnik, *The Nature of Statistical Learning Theory.* New York: Springer-Verlag, 1995.

[14] B. Schölkopf, K.K. Sung, C.J.C. Burges, F. Girosi, P. Niyogi, T. Poggio and V. Vapnik, "Comparing support vector machines with Gaussian kernels to radial basis function classifiers," *IEEE Trans. Signal Processing*, Vol.45, No.11, pp.2758–2765, 1997.

[15] C.J.C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, Vol.2, No.2, pp.121–167, 1998.

[16] M.E. Tipping, "The relevance vector machine," in Sara A. Solla, Todd K. Leen and Klaus-Robert Müller, eds., *Advances in Neural Information Processing Systems 12*, Cambridge, MA: MIT Press, 2000.

[17] D.J.C. MacKay, "Bayesian interpolation," *Neural Computation*, Vol.4, No.3, pp.415–447, 1992.

[18] D.J.C. MacKay, "The evidence framework applied to classification networks," *Neural Computation*, Vol.4, pp.720–736, 1992.

[19] 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.

[20] M.S. Bazaraa, H.D. Sherali and C.M. Shetty, *Nonlinear Programming: Theory and Algorithms.* New York: John Wiley, 1993.

[21] I.T. Nabney, "Efficient training of RBF networks for classification," in *Proc. 9th ICANN*, 1999, Vol.1, pp.210–215.