Support Vector Machine Multiuser Receiver for DS-CDMA Signals in Multipath Channels

S. Chen, A.K. Samingan and L. Hanzo

Abstract

The problem of constructing an adaptive multiuser detector (MUD) is considered for direct-sequence code-division multiple-access (DS-CDMA) signals transmitted through multipath channels. The emerging learning technique, called support vector machines (SVMs), is proposed as a method of obtaining a nonlinear MUD from a relatively small training data block. Computer simulation is used to study this SVM MUD, and the results show that it can closely match the performance of the optimal Bayesian one-shot detector. Comparisons with an adaptive radial basis function (RBF) MUD trained by an unsupervised clustering algorithm are discussed.

Keywords

DS-CDMA, multiuser interference, multiuser detector, linear MMSE detector, optimal one-shot detector, support vector machines, unsupervised clustering.

I. Introduction

DS-CDMA [1] constitutes an attractive multiuser scheme that allows users to transmit at the same carrier frequency in an uncoordinated manner. However, this creates multiuser interference (MUI) which, if not controlled, can seriously degrade the quality of reception. Mutually orthogonal spreading codes for different users can provide an inherent immunity to MUI in the case of synchronous systems. Unfortunately, multipath distortions are often encountered in DS-CDMA systems and will reduce this inherent immunity to MUI. A variety of receivers, known as MUDs, have been proposed for DS-CDMA systems [2]–[4]. In a DS-CDMA system, the objective of the receiver is to detect the transmitted information bits of one (at mobile-end) or many (at base station) users. In this paper, the first case is considered. This is usually referred to as the downlink scenario, the communication channel from base station to mobile. In particular, the one-shot detector, that is, the receiver detecting a user bit at a bit period, is studied. Furthermore, we will concentrate on the scenario where a few strong interfering users exist, as the use of an MUD is especially effective in such a situation.

For such applications, the linear minimum mean square error (MMSE) MUD [5]-[9] is widely used, as it is computationally very simple and can readily be implemented using standard adaptive filter techniques. A more complicated linear minimum bit error rate MUD has also been studied [10]-[14]. Linear detectors, however, can only work when the underlying noise-free signal classes are linearly

TNN Topic Area: Signal Processing and Communications
Ref.No.: TNN B085
The authors are with Department of Electronics and Computer Science, University of Southampton, Highfield, Southampton SO17 1BJ, U.K.
Contact author: S. Chen, Tel./Fax: +44 (0)23 8095 6660/4508; E-mail: sqc@ecs.soton.ac.uk.

separable. As nonlinear separable cases are common in DS-CDMA channels, neural networks have been considered as nonlinear MUDs. In the work [15], a multilayer perceptron (MLP) was applied to a CDMA system without intersymbol interference (ISI). The experience shows that the MLP MUD has better performance than the linear MUD but training times are long and unpredictable. Mitra and Poor [16] applied a RBF network to the same problem. An advantage of the RBF MUD is its intimate link with the optimal one-shot detector. Training times are better and more predictable than the MLP. A RBF receiver [17] was also considered for channels with memory, and a Volterra series detector was explored [18].

A learning technique known as the SVM has recently gaining popularity due to its many attractive features and promising empirical performance [19]-[26]. The formulation of SVM embodies the structural risk minimisation (SRM) principle, as opposed to the empirical risk minimisation (ERM) approach commonly employed in statistical learning. SRM minimises an upper bound on the generalisation error, as opposed to ERM which minimises the error on the training data. It is this difference which equips SVMs with a greater potential to generalise. For binary classification tasks, the SVM approach nonlinearly maps the input space into a high dimensional feature space via simple kernel representations. In the high dimensional feature space, a linear classifier with maximum margin is constructed. Apart from good generalisation properties, the learning aspect of SVMs is intriguing. An SVM classifier is determined only by a sparse set of support vectors (SVs), and these SVs are automatically selected from the training data during the learning process.

The SVM technique has been applied to channel equalization [27]–[30]. These studies have shown that the SVM approach is very effective in overcoming ISI and co-channel interference. As the idea of SVM is originated from finding an optimum hyperplane to separate two classes with maximum margin, it is also very relevant to multiuser detection in DS-CDMA. In this paper, the SVM technique is investigated as an adaptive nonlinear MUD. Our study shows that an SVM MUD trained using a relatively small block of noisy received signal samples can closely approximate the optimal MUD which requires a complete knowledge of the system. The remainder of the paper is organized as follows. Section II presents the DS-CDMA system model used and provides the necessary notations and definitions. Section III summarizes the linear MMSE MUD and the optimal Bayesian one-shot MUD. In Section IV, the proposed SVM MUD is presented. Section V gives some computer simulation results, comparing the performance of the SVM MUD with those of the linear and optimal MUDs. Section VI compares the adaptive SVM MUD with a RBF MUD trained by an unsupervised clustering algorithm [31]–[34]. The paper concludes in Section VII.

II. System model

Using notations from the multi-rate filtering literature [35], the discrete-time model of the synchronous DS-CDMA system with N users and M chips per bit is depicted in Figure 1, where $b_i(k) \in \{\pm 1\}$ denotes the k-th bit of user i, the signature code sequence for user i

$$\bar{\mathbf{s}}_i = [\bar{s}_{i,1} \ \cdots \ \bar{s}_{i,M}]^T \tag{1}$$

is normalized to have a unit length, and the transfer function of the channel impulse response (CIR) is

$$H(z) = \sum_{i=0}^{n_h-1} h_i z^{-i} \,. \tag{2}$$

It is assumed that $N < M$. The bit vector of N users at k is

$$\mathbf{b}(k) = [b_1(k) \ \cdots \ b_N(k)]^T \,, \tag{3}$$

and the received signal vector after the chip-matched filters is

$$\mathbf{r}(k) = [r_1(k) \ \cdots \ r_N(k)]^T \,. \tag{4}$$

The baseband model for $\mathbf{r}(k)$ can be represented by [36]:

$$\begin{aligned}\mathbf{r}(k) &= \bar{\mathbf{S}}^T \mathbf{H} \begin{bmatrix} \bar{\mathbf{S}}\mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{S}}\mathbf{A} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \bar{\mathbf{S}}\mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{b}(k) \\ \mathbf{b}(k-1) \\ \vdots \\ \mathbf{b}(k-L+1) \end{bmatrix} + \bar{\mathbf{S}}^T \mathbf{n}(k) \\ &= \mathbf{P} \begin{bmatrix} \mathbf{b}(k) \\ \mathbf{b}(k-1) \\ \vdots \\ \mathbf{b}(k-L+1) \end{bmatrix} + \tilde{\mathbf{n}}(k) \,, \end{aligned} \tag{5}$$

where the channel white Gaussian noise vector is

$$\mathbf{n}(k) = [n_1(k) \ \cdots \ n_M(k)]^T \ \text{ with } \ E[\mathbf{n}(k)\mathbf{n}^T(k)] = \sigma_n^2 \mathbf{I} \,; \tag{6}$$

the user signature sequence matrix is

$$\bar{\mathbf{S}} = [\bar{\mathbf{s}}_1 \ \cdots \ \bar{\mathbf{s}}_N] \,; \tag{7}$$

the diagonal user signal amplitude matrix is

$$\mathbf{A} = \text{diag}\{A_1 \ \cdots \ A_N\} \,, \tag{8}$$

that is, A_i^2 is the user i signal power; the $M \times LM$ CIR matrix $\mathbf{H}$ has the form

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 & \cdots & h_{n_h-1} & & & & \\ & h_0 & h_1 & \cdots & h_{n_h-1} & & & \\ & & \ddots & \ddots & \cdots & \ddots & & \\ & & & h_0 & h_1 & \cdots & h_{n_h-1} & \end{bmatrix} ; \tag{9}$$

the $N \times LN$ system matrix $\mathbf{P}$ is defined by

$$\mathbf{P} = \bar{\mathbf{S}}^T \mathbf{H} \begin{bmatrix} \bar{\mathbf{S}}\mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{S}}\mathbf{A} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \bar{\mathbf{S}}\mathbf{A} \end{bmatrix} ; \tag{10}$$

and orthogonal code sequences are assumed, so that the noise vector at the outputs of the chip-matched filters is

$$\tilde{\mathbf{n}}(k) = [\tilde{n}_1(k) \ \cdots \ \tilde{n}_N(k)]^T \ \text{ with } \ E[\tilde{\mathbf{n}}(k)\tilde{\mathbf{n}}^T(k)] = \sigma_n^2 \mathbf{I} \, . \tag{11}$$

The ISI span L depends on the length of the CIR, n_h, related to the length of the chip sequence, M. For $n_h = 1$, $L = 1$; for $1 < n_h \le M$, $L = 2$; for $M < n_h \le 2M$, $L = 3$; and so on. The model (5) adopted in this study can readily be extended to the more general case of asynchronous DS-CDMA systems by an appropriate expansion of the system matrix $\mathbf{P}$ [6].

It is worth pointing out that we assume an MUD operating on the outputs $\mathbf{r}(k)$ of a bank of chip-matched filters, as modelled by (5). This model requires the knowledge of all the user codes. The assumption that a receiver knows the user codes is often a reasonable assumption to have. If, however, the user codes are unknown to the receiver, the MUD can directly operate on the chip-rate sampled signals $\mathbf{r}_c(k)$. For the completeness, the model for $\mathbf{r}_c(k)$ is given in Appendix, where it can be seen that operating on $\mathbf{r}_c(k)$ and $\mathbf{r}(k)$ produces equivalent results.

III. The linear and optimal detectors

The linear MUD for user i has the form:

$$\hat{b}_i(k) = \text{sgn}(y_L(k)) \tag{12}$$

with

$$y_L(k) = f_L(\mathbf{r}(k)) = \mathbf{w}^T \mathbf{r}(k) \, , \tag{13}$$

where

$$\mathbf{w} = [w_1 \ \cdots \ w_N]^T \tag{14}$$

denotes the detector weight vector for user i. The most popular solution for the detector (13) is the MMSE solution given by

$$\mathbf{w}_{MMSE} = \left(\sigma_n^2 \mathbf{I} + \mathbf{P}\mathbf{P}^T\right)^{-1} \mathbf{p}_i \,, \tag{15}$$

where $\mathbf{p}_i$ is the i-th column of $\mathbf{P}$. The linear detector (13) is computationally very simple, and standard least mean square (LMS) or recursive least squares (RLS) algorithms can be used to implement the MMSE solution adaptively.

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

$$\mathbf{b}^{(j)} = \begin{bmatrix} \mathbf{b}^{(j)}(k) \\ \mathbf{b}^{(j)}(k-1) \\ \vdots \\ \mathbf{b}^{(j)}(k-L+1) \end{bmatrix}, \ 1 \le j \le N_b \,, \tag{16}$$

and $b_i^{(j)}$ the ith element of $\mathbf{b}^{(j)}(k)$. Define the set of the N_b noise-free received signal states as

$$\mathcal{R} = \{\mathbf{r}_j = \mathbf{P}\mathbf{b}^{(j)},\ 1 \le j \le N_b\} \,. \tag{17}$$

$\mathcal{R}$ can be partitioned into two subsets:

$$\mathcal{R}_{\pm} = \{\mathbf{r}_j \in \mathcal{R} : b_i^{(j)} = \pm 1\} \,. \tag{18}$$

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 hyperplane in the N-dimensional received signal space.

As the one-shot MUD at a bit period is only concerned with one particular bit from one particular user, the optimal solution can be derived from a one-shot Bayesian classification problem, similar to the channel equalization problem [31]. Thus, the optimal one-shot detector for user i can readily be given as:

$$y_B(k) = f_B(\mathbf{r}(k)) = \sum_{\mathbf{r}_j \in \mathcal{R}} \frac{\xi_j b_i^{(j)}}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{\|\mathbf{r}(k) - \mathbf{r}_j\|^2}{2\sigma_n^2}\right) , \tag{19}$$

where ξ_j are *a priori* probabilities of $\mathbf{r}_j$ and $b_i^{(j)} \in \{\pm 1\}$ serve as class labels. Usually, all the $\mathbf{r}_j$ are equiprobable, and $\xi_j = \frac{1}{N_b}$. The optimal decision is then given by

$$\hat{b}_i(k) = \mathrm{sgn}(y_B(k)) \,. \tag{20}$$

The Bayesian decision variable (19) can also be written as

$$y_B(k) = f_B(\mathbf{r}(k)) = \sum_{j=1}^{N_b} \beta_j b_i^{(j)} \exp\left(-\frac{\|\mathbf{r}(k) - \mathbf{r}_j\|^2}{2\sigma_n^2}\right) . \tag{21}$$

An observation of (21) is that it has an equivalent form to an RBF network with Gaussian basis function or kernel.

IV. The support vector machine detector

The optimal detector (21) 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 $\{\mathbf{r}(k), b_i(k)\}_{k=1}^{K}$. For notational convenience, let $\mathbf{x}_k = \mathbf{r}(k)$ and $c_k = b_i(k)$, and denote the training set of K noisy received signal vectors as

$$\mathcal{X} = \{\mathbf{x}_j,\ 1 \le j \le K\} \tag{22}$$

and the set of corresponding class labels as

$$\mathcal{C} = \{c_j,\ 1 \le j \le K\}\,. \tag{23}$$

Applying the standard SVM method [19], an SVM detector can be constructed for user i:

$$y_{SVM}(k) = f_{SVM}(\mathbf{r}(k)) = \sum_{j=1}^{K} \bar{g}_j c_j F(\mathbf{r}(k), \mathbf{x}_j) + \bar{\eta}\,, \tag{24}$$

where the set of Lagrangian multipliers $\{\bar{g}_j\}$, denoted in the vector form

$$\bar{\mathbf{g}} = [\bar{g}_1 \ \cdots \bar{g}_K]^T\,, \tag{25}$$

is the solution of the quadratic programming (QP)

$$\bar{\mathbf{g}} = \arg\min_{\mathbf{g}} \left\{ \frac{1}{2} \sum_{j=1}^{K} \sum_{l=1}^{K} g_j g_l c_j c_l F(\mathbf{x}_j, \mathbf{x}_l) - \sum_{j=1}^{K} g_j \right\} \tag{26}$$

with the constraints

$$0 \le g_j \le C, \quad 1 \le j \le K\,, \tag{27}$$

and

$$\sum_{j=1}^{K} g_j c_j = 0\,. \tag{28}$$

In this particular application, it is obviously advantageous to choose the Gaussian kernel function:

$$F(\mathbf{x}_j, \mathbf{x}_l) = \exp\left(- \frac{\|\mathbf{x}_j - \mathbf{x}_l\|^2}{2\rho^2} \right)\,, \tag{29}$$

where the width parameter ρ is related to the root mean square σ_n of the channel noise, an estimate of which can be obtained. The offset constant $\bar{\eta}$ is usually determined from the "margin" SVs, i.e.

those $\mathbf{x}_j$s with the corresponding Lagrangian multipliers $0 < \bar{g}_j < C$. Because the optimal decision boundary, defined by $\{\mathbf{r} : f_B(\mathbf{r}) = 0\}$, passes through the origin of the received signal space and possesses certain symmetric properties due to the symmetric structure of $\mathcal{R}_-$ and $\mathcal{R}_+$, $\bar{\eta} = 0$ can be used. With this choice of the offset constant, the equality constraint (28) is no longer needed, and this leads to a simpler optimization task. The user-defined parameter C controls the trade-off between model complexity and training error. In our application, we will choose C empirically.

The set of SVs, denoted by $\mathcal{X}_{SVM}$, is given by those $\mathbf{x}_j$s with non-zero Lagrangian multipliers $0 < \bar{g}_j \leq C$. $\mathcal{X}_{SVM}$ is usually a small subset of the training data set $\mathcal{X}$. These SVs are determined during the optimization process. Thus the SVM MUD consists of computing the decision variable

$$y_{SVM}(k) = f_{SVM}(\mathbf{r}(k)) = \sum_{\mathbf{x}_j \in \mathcal{X}_{SVM}} \bar{g}_j c_j \exp\left(-\frac{\|\mathbf{r}(k) - \mathbf{x}_j\|^2}{2\rho^2}\right) \tag{30}$$

and making the decision with

$$\hat{b}_i(k) = \mathrm{sgn}(y_{SVM}(k))\,. \tag{31}$$

Before presenting some simulation results, we emphasize that the main purpose of this study is to investigate feasibility of the SVM method as a nonlinear MUD, and we use the standard gradient method to solve for the QP problem (26). Such an optimization technique associated with the SVM method requires a considerable amount of computation, especially for a large K. The issue of computational efficiency is more general than the MUD problem. Efficient SVM implementation is an active area of research, see for example, [23],[37],[38].

V. Simulation results

Two simulation examples were used to investigate the proposed SVM MUD and compare its performance with those of the linear MMSE and optimal MUDs. It is worth pointing out again that the linear MMSE MUD and the optimal MUD are designed based on the complete knowledge of the system (the system matrix $\mathbf{P}$ and the noise variance) while the SVM MUD is 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 of the CIR was

$$H(z) = 0.3 + 0.7z^{-1} + 0.3z^{-2}\,. \tag{32}$$

The two users had equal signal power, that is, the user 1 signal to noise ratio (SNR_1) was equal to that of user 2 (SNR_2). Figure 2 depicts the two subsets of noise-free signal states for user 1 together with the decision boundaries of the linear MMSE and optimal Bayesian MUDs, given $\mathrm{SNR}_1 = \mathrm{SNR}_2 = 16$ dB. It is clear that for user 1 it is not worth to implement a complex nonlinear detector, as a simple linear detector can perform adequately. However, 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 illustrated in Figure 3.

To construct an SVM MUD for user 2, 160 training data points were generated for each given noise variance. The influence of the parameter C on the model complexity and the bit error rate (BER) of the resulting MUD was first investigated. Tables I and II summarize the results for two different noise conditions, respectively. It can be seen that C has some influence on the construction of SVM model, but the solution is not overly sensitive to it. The decision boundary of the constructed SVM MUD for user 2 is compared with those of the linear MMSE and optimal MUDs in Figure 3, given $\mathrm{SNR}_1 = \mathrm{SNR}_2 = 20$ dB. Figure 4 plots the the BERs of the three MUDs for user 2 under different noise conditions. The results clearly show that the SVM MUD can closely match the performance of the optimal one-shot detector. For the given training set of 160 points, the number of SVs were found typically to be around 40.

Example 2. This was a 3-user system with 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 transfer function of the CIR was

$$H(z) = 0.4 + 0.9z^{-1} + 0.4z^{-2}\,. \tag{33}$$

The three users had equal signal power. For user 1, the performance of the linear MMSE detector is very close to the optimal detector, as shown in Figure 5. For users 2 and 3, however, nonlinear MUDs are certainly needed, as can be seen in Figures 6 and 7.

The number of training data for constructing SVM models was 640 for each given noise condition. The influence of C on the model complexity and performance was again found to be not overly sensitive, as illustrated in Tables III and IV for the case of user 2 with two different noise conditions. From the training data set of 640 points, typically around 200 SVs were selected. The BERs of the resulting SVM MUDs for users 2 and 3 are given in Figures 6 and 7, respectively, in comparison with the corresponding linear MMSE and optimal one-shot MUDs. The results again demonstrate that the SVM MUD can closely approximate the performance of the optimal detector.

We would like to point out that, even though orthogonal spreading sequences were used in the simulation, the orthogonality was destroyed by the ISI channel. Interfering signals in the simulation were also kept to be relatively strong. The desired user signal to interference ratio was 0 dB for

Example 1 and approximately -5 dB for Example 2.

VI. Comparison with the clustering RBF detector

Because of its intimate link with the optimal one-shot detector (21), the RBF model of the form

$$y_{RBF}(k) = f_{RBF}(\mathbf{r}(k)) = \sum_{j=1}^{N_c} \alpha_j \exp\left(-\frac{\|\mathbf{r}(k) - \mathbf{c}_j\|^2}{2\rho^2}\right) \tag{34}$$

is a good candidate for nonlinear MUD. Assuming the number of the users is known, the number of the centers should be $N_c = N_b$, and the optimal RBF centers are the set of noise-free signal states $\mathcal{R}$. The width ρ is an estimate of the noise standard deviation, and the weights α_j can simply be set according to the class label of the corresponding center, or trained using the LMS or RLS algorithm. Such an RBF MUD achieves the exact optimal MUD.

The most efficient way of adaptively placing the RBF centers to the desired noise-free signal states is the supervised cluster algorithm [31]. In the multiuser detection context, the algorithm will require to know all the users' bits from $k-L+1$ to k, and this is impractical for downlink. Thus, unsupervised clustering has to be used. The enhanced κ-means clustering algorithm [32]–[34] is ideal for the RBF MUD. By using a cluster variation-weighted measure, this algorithm always converges to an optimal or near optimal cluster configuration, independent of the initial center locations. Furthermore, the variance of every cluster is equal after convergence. This property is particular relevant to our application since all the cluster variances in this case should be equal.

The enhanced κ-means clustering method [33] adjusts the RBF centers according to:

$$\mathbf{c}_j(k+1) = \mathbf{c}_j(k) + \mathcal{M}_j(\mathbf{r}(k))\mu(\mathbf{r}(k) - \mathbf{c}_j(k))\,, \tag{35}$$

where the membership function

$$\mathcal{M}_j(\mathbf{r}) = \begin{cases} 1, & \text{if } v_j\|\mathbf{r} - \mathbf{c}_j\|^2 \le v_l\|\mathbf{r} - \mathbf{c}_l\|^2 \ \forall l \ne j \\ 0, & \text{otherwise} \end{cases} \tag{36}$$

and v_j is the "variance" of the j-th cluster. To estimate v_j, the following rule is used:

$$v_j(k+1) = \gamma v_j(k) + (1-\gamma)\mathcal{M}_j(\mathbf{r}(k))\|\mathbf{r}(k) - \mathbf{c}_j(k)\|^2\,, \tag{37}$$

where γ is a constant slightly less than 1.0. The initial $v_j(0)$, $1 \le j \le N_c$, can be set to the same small value. The learning rate for centers, μ, is self-adjusting based on an "entropy" formula [33].

To illustrate the optimal properties of the enhanced κ-means clustering method, consider the user 2 of Example 1 given in the previous section with SNR$_1$ =SNR$_2$ =20 dB. Started from $\mathbf{c}_j(0) = \mathbf{0}$,

$1 \le j \le 16$, the trajectories of cluster centers in 10000-samples adaptation are shown in Figure 8. Table V summarizes the final cluster center positions and variations, in comparison with the true noise-free signal states. Similar clustering performance can be obtained with only 4000 data samples when the first 16 data points are used as the initial centers, as illustrated in Figure 9. The BER performance of such a clustering RBF MUD is indistinguishable with the optimal one-shot MUD.

The following comparisons for the SVM and clustering RBF MUDs can be made. The adaptation of the SVM model is based on a block of data while the clustering RBF is implemented sample-by-sample. The sample-by-sample adjustment is more desired in real-time applications. The clustering RBF MUD can achieve the optimal MUD, provided that a sufficient number of training data are available. The SVM MUD has an important advantage as it requires a relatively short training data set. For our simple Example 1, The SVM MUD requires less than 200 training data. In contrast, the clustering RBF MUD needs a few thousands of training data, due to the nature of the unsupervised clustering process. Such a long training data length is difficult to achieve in practice. As the SVM approach places the kernel centers directly on some noisy data points called SVs, it will require more centers than the clustering RBF model. For Example 1, the SVM MUD needs typically 40 kernels to closely match the optimal performance, while the clustering RBF MUD has only 16 kernel functions.

VII. Conclusions

The SVM technique has been applied to adaptive nonlinear multiuser detection for DS-CDMA systems. It has been shown that the SVM MUD can closely match the performance of the optimal Bayesian one-shot detector with an important advantage of requiring a relatively small training data set. Comparisons with an adaptive clustering RBF MUD have been made. A disadvantage of the SVM approach is its block-based adaptation nature. Although the SVM approach has an excellent model reduction ability, the resulting SVM MUD still has a larger model size, in comparison with the number of noise-free signal states. Future research will investigate how to reduce the number of support vectors further without sacrificing the BER performance too much and how to incorporate the sample-by-sample adaptive methodology with the SVM approach.

Appendix

During one bit period, the chip rate sampler produces M samples:

$$\mathbf{r}_c(k) = [r_{c,1}(k) \ \cdots \ r_{c,M}(k)]^T . \tag{38}$$

The baseband model for $\mathbf{r}_c(k)$ is represented by [36]:

$$\mathbf{r}_c(k) = \mathbf{H}\begin{bmatrix} \bar{\mathbf{S}}\mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{S}}\mathbf{A} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \bar{\mathbf{S}}\mathbf{A} \end{bmatrix}\begin{bmatrix} \mathbf{b}(k) \\ \mathbf{b}(k-1) \\ \vdots \\ \mathbf{b}(k-L+1) \end{bmatrix} + \mathbf{n}(k) = \mathbf{P}_c \begin{bmatrix} \mathbf{b}(k) \\ \mathbf{b}(k-1) \\ \vdots \\ \mathbf{b}(k-L+1) \end{bmatrix} + \mathbf{n}(k)\,, \tag{39}$$

where the $M \times LN$ system matrix $\mathbf{P}_c$ is defined as

$$\mathbf{P}_c = \mathbf{H}\begin{bmatrix} \bar{\mathbf{S}}\mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{S}}\mathbf{A} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \bar{\mathbf{S}}\mathbf{A} \end{bmatrix}. \tag{40}$$

An MUD can directly operate on $\mathbf{r}_c(k)$, instead of the chip-matched filtered version $\mathbf{r}(k)$. For example, an alternative linear MUD takes the form:

$$y_L(k) = \mathbf{w}_c^T \mathbf{r}_c(k) \tag{41}$$

with the M-dimensional weight vector $\mathbf{w}_c = [w_{c,1} \cdots w_{c,M}]^T$. It is well-known that this linear MUD is equivalent to the linear MUD of (13). In the M-dimensional space $\mathbf{r}_c(k)$, the set of the N_b noise-free received signal states is defined by

$$\mathcal{R}_c = \{\mathbf{r}_{c,j} = \mathbf{P}_c \mathbf{b}^{(j)},\ 1 \le j \le N_b\}\,. \tag{42}$$

An optimal Bayesian one-shot detector, similar to (21), can alternatively be defined.

References

[1] R. Prasad, *CDMA for Wireless Personal Communications*. Artech House, Inc., 1996.

[2] S. Verdú, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Information Theory*, Vol.IT-32, No.1, pp.85–96, 1986.

[3] S. Moshavi, "Multi-user detection for DS-CDMA communications," *IEEE Communications Magazine*, Vol.34, No.10, pp.124–136, 1996.

[4] S. Verdú, *Multiuser Detection*. Cambridge University Press, 1998.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

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

[10] N.B. Mandayam and B. Aazhang, "Gradient estimation for sensitivity analysis and adaptive multiuser interference rejection in code-division multi-access systems," *IEEE Trans. Communications*, Vol.45, No.7, pp.848–858, 1997.

[11] X.F. Wang, W.S. Lu and A. Antoniou, "Constrained minimum-BER multiuser detection," in *Proc. ICASSP* (Phoenix, USA), May 14-18, 1999, Vol.5, pp.2603–2606.

[12] I.N. Psaromiligkos, S.N. Batalama and D.A. Pados, "On adaptive minimum probability of error linear filter receivers for DS-CDMA channels," *IEEE Trans. Communications*, Vol.47, No.7, pp.1092–1102, 1999.

[13] C.C. Yeh, R.R. Lopes and J.R. Barry, "Approximate minimum bit-error rate multiuser detection," in *Proc. Globecom'98* (Sydney, Australia), Nov. 1998, pp.3590–3595.

[14] S. Chen, A.K. Samingan, B. Mulgrew and L. Hanzo, "Adaptive minimum-BER linear multiuser detection for DS-CDMA signals in multipath channels," submitted to *IEEE Trans. Signal Processing*, 2000.

[15] 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.

[16] 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.

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

[18] 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.

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

[20] 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.

[21] 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.

[22] V. Vapnik, "The support vector method of function estimation," in *Nonlinear Modeling: Advanced Black-Box Techniques*, J.A.K. Suykens and J. Vandewalle, eds. Boston, MA: Kluwer, 1998, pp.55–85.

[23] J. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel Methods – Support Vector Learning*, B. Schölkopf, C.J.C. Burges and A.J. Smola, eds. Cambridge, MA: MIT Press, 1999, pp.185–208.

[24] Special Issue on VC Learning Theory and Its Applications, *IEEE Trans. Neural Networks*, Vol.10, No.5, September 1999.

[25] A. Lyhyaoui, M. Martínez, I. Mora, M. Vázquez, J.L. Sancho and A.R. Figueiras-Vidal, "Sample selection via clustering to construct support vector-like classifiers," *IEEE Trans. Neural Networks*, Vol.10, No.6, pp.1474–1481, 1999.

[26] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines.* Cambridge University Press, 2000.

[27] S. Chen, S. Gunn and C.J. Harris, "Decision feedback equalizer design using support vector machines," *IEE Proc. Vision, Image and Signal Processing*, Vol.147, No.3, pp.213–219, 2000.

[28] D.J. Sebald and J.A. Bucklew, "Support vector machine techniques for nonlinear equalization," *IEEE Trans. Signal Processing*, to appear, 2000.

[29] 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, Wisconsin), Aug.23-25, 1999, pp.49–57.

[30] F. Pérez-Cruz, A. Navia-Vázquez, P.L. Alarcón-Diana and A. Artés-Rodríguez, "SVC-based equalizer for burst TDMA transmission," *Signal Processing*, accepted, 2000.

[31] 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.

[32] S. Chen, "Nonlinear time series modelling and prediction using Gaussian RBF networks with enhanced clustering and RLS learning," *Electronics Letters*, Vol.31, No.2, pp.117–118, 1995.

[33] C. Chinrungrueng and C.H. Séquin, "Optimal adaptive κ-means algorithm with dynamic adjustment of learning rate," *IEEE Trans. Neural Networks*, Vol.6, No.1, pp.157–169, 1995.

[34] S. Chen, S. McLaughlin, B. Mulgrew and P.M. Grant, "Bayesian decision feedback equaliser for overcoming co-channel interference," *IEE Proc. Communications*, Vol.143, No.4, pp.219–225, 1996.

[35] P.P. Vaidyanathan, *Multirate Systems and Filter Banks.* Englewood Cliffs, NJ: Prentice Hall, 1993.

[36] B. Mulgrew, "Nonlinear signal processing for adaptive equalisation and multi-user detection," in *Proc. EUSIPCO 98* (Rhodes, Greece), September 1998, pp.537–544.

[37] S.S. Keerthi, S.K. Shevade, C. Bhattacharyya and K.R.K. Murthy, "A fast iterative nearest point algorithm for support vector machine classifier design," *Technical Report TR-ISL-99-03*, Intelligent Systems Lab, Dept. of Computer Science & Automation, Indian Institute of Science, 1999.

[38] T. Joachims, "Making large-scale support vector machine learning practical," in *Advances in Kernel Methods – Support Vector Learning*, B. Schölkopf, C.J.C. Burges and A.J. Smola, eds. Cambridge, MA: MIT Press, 1999, pp.169–184.

Sheng Chen (SM'97) received the PhD degree in control engineering from the City University, London in 1986. From 1986 to 1999, he held various research and academic appointments at the Universities of Sheffield, Edinburgh and Portsmouth, UK. Since 1999, he has been with the Department of Electronics and Computer Science, the University of Southampton, UK, where he currently holds a post of Readership in Communications Signal Processing. Dr. Chen is a Senior Member of the Institute of Electrical and Electronics Engineers. His research interests include modelling and identification of nonlinear systems, adaptive nonlinear signal processing, artificial neural network research, finite-precision digital controller design, evolutionary computation methods and optimization. He has published over 140 research papers.

Ahmad Kamsani Samingan received the B.Eng. (Hons.) degree in electronic engineering from the University of Southampton, UK in 1999 with an exceptional First Class honours. He was named a Zepler Prize winner in 1998 due to his excellent academic achievement. He is currently with the Communication group, the Department of Electronic and Computer Science, the University of Southampton, and is studying for his PhD degree which is due to complete in 2002. His research interest covers the field of CDMA multiuser detection, particularly in the area of adaptive multiuser detection and Support Vector Machine (SVM) based multiuser detection.

Lajos Hanzo has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the Department of Electronics and Computer Science, University of Southampton, UK, where he currently holds the chair in Telecommunications. He co-authored five books on mobile radio communications, published over 200 research papers, organised and chaired conference sessions, presented overview lectures and was awarded a number of distinctions. Currently he is managing a research team, working on a range of research projects in the field of wireless multimedia communications under the auspices of the Engineering and Physical Sciences Research Council (EPSRC) UK, the European IST Programme and the Mobile Virtual Centre of Excellence (VCE). For further information on research in progress and associated publications please refer to http://www-mobile.ecs.soton.ac.uk.

TABLE I

INFLUENCE OF C ON CONSTRUCTING THE SVM MUD FOR USER 2 OF EXAMPLE 1. $\mathrm{SNR}_1 = \mathrm{SNR}_2 = 20$ dB, AND THE NUMBER OF TRAINING DATA IS 160.

C	1.0	5.0	10.0	100.0
number of SVs	50	35	23	52
$\log_{10}$(BER)	-3.084	-2.740	-2.974	-2.840
optimal MUD $\log_{10}(\mathrm{BER}) = -3.155$				

TABLE II

INFLUENCE OF C ON CONSTRUCTING THE SVM MUD FOR USER 2 OF EXAMPLE 1. $\mathrm{SNR}_1 = \mathrm{SNR}_2 = 15$ dB, AND THE NUMBER OF TRAINING DATA IS 160.

C	1.0	5.0	10.0	100.0
number of SVs	52	39	37	36
$\log_{10}$(BER)	-1.590	-1.506	-1.440	-1.424
optimal MUD $\log_{10}(\mathrm{BER}) = -1.674$				

TABLE III

INFLUENCE OF C ON CONSTRUCTING THE SVM MUD FOR USER 2 OF EXAMPLE 2. $\mathrm{SNR}_i = 30$ dB FOR $1 \leq i \leq 3$, AND THE NUMBER OF TRAINING DATA IS 640.

C	0.5	0.6	0.8	0.9	1.0	5.0
number of SVs	220	206	192	179	173	139
$\log_{10}$(BER)	-2.840	-2.850	-2.833	-2.827	-2.797	-2.490
optimal MUD $\log_{10}(\mathrm{BER}) = -3.009$						

TABLE IV

Influence of C on constructing the SVM MUD for user 2 of Example 2. $\text{SNR}_i = 25$ dB for $1 \le i \le 3$, and the number of training data is 640.

C	0.6	0.8	1.0	5.0
number of SVs	261	240	229	181
$\log_{10}(\text{BER})$	-1.442	-1.454	-1.459	-1.429
optimal MUD $\log_{10}(\text{BER}) = -1.521$				

TABLE V

Cluster centers and variations obtained using the enhanced κ-means clustering for user 2 of Example 1. $\text{SNR}_1 = \text{SNR}_2 = 20$ dB (noise variance 0.001), initial centers are all placed at the origin, and 10000 data samples are used.

class	true center	cluster center	cluster variance
$+1$	$(-0.8552, 0.8552)$	$(-0.8615, 0.8468)$	0.0011
$+1$	$(-1.6493, -0.4276)$	$(-1.6563, -0.4339)$	0.0011
$+1$	$(-0.4276, 0.7941)$	$(-0.4238, 0.7943)$	0.0014
$+1$	$(-1.2217, -0.4887)$	$(-1.2107, -0.4959)$	0.0015
$+1$	$(-0.0611, 0.4276)$	$(-0.0664, 0.4182)$	0.0014
$+1$	$(-0.8552, -0.8552)$	$(-0.8510, -0.8662)$	0.0013
$+1$	$(0.3665, 0.3665)$	$(0.3810, 0.3704)$	0.0013
$+1$	$(-0.4276, -0.9163)$	$(-0.4306, -0.9213)$	0.0021
-1	$(0.4276, 0.9163)$	$(0.4399, 0.9148)$	0.0011
-1	$(-0.3665, -0.3665)$	$(-0.3653, -0.3769)$	0.0013
-1	$(0.8552, 0.8552)$	$(0.8550, 0.8325)$	0.0013
-1	$(0.0611, -0.4276)$	$(0.0669, -0.4298)$	0.0013
-1	$(1.2217, 0.4887)$	$(1.2286, 0.4972)$	0.0019
-1	$(0.4276, -0.7941)$	$(0.4198, -0.8084)$	0.0012
-1	$(1.6493, 0.4276)$	$(1.6459, 0.4352)$	0.0015
-1	$(0.8552, -0.8552)$	$(0.8630, -0.8501)$	0.0011

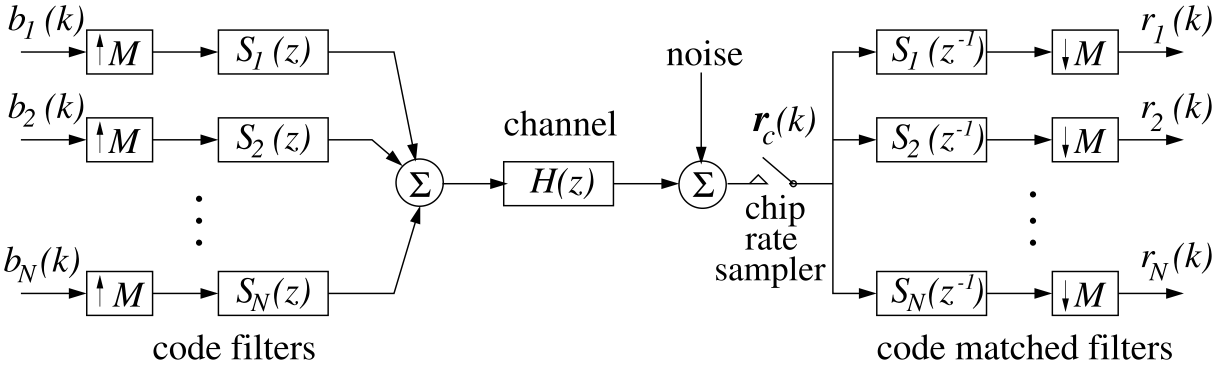

Fig. 1. Discrete-time model of synchronous CDMA downlink.

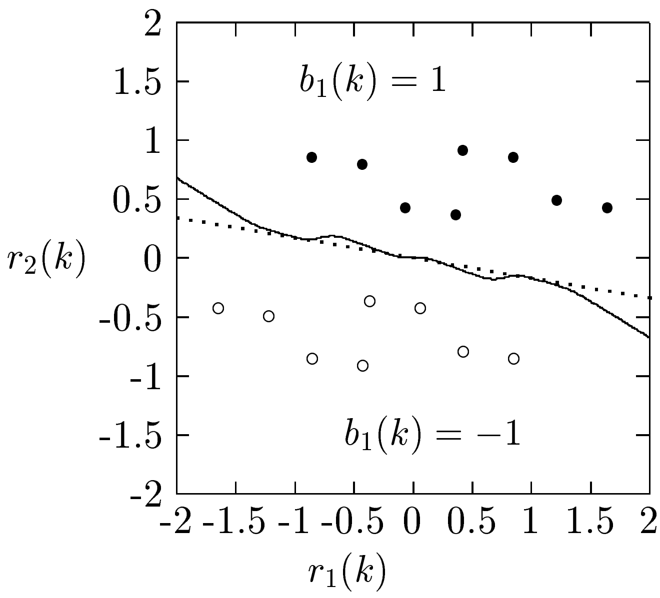

Fig. 2. The set of noise-free signal points and the two decision boundaries (dotted: linear MMSE, solid: optimal) for user 1 of Example 1. $\text{SNR}_1 = \text{SNR}_2 = 16$ dB.

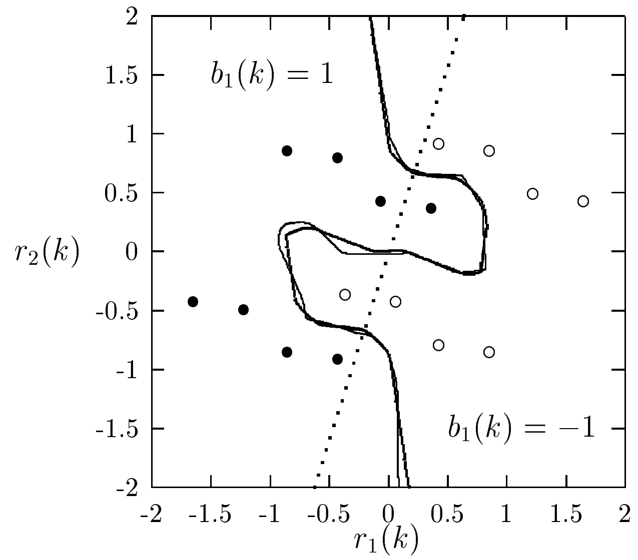

Fig. 3. The set of noise-free signal points and the three decision boundaries (dotted: linear MMSE, thick solid: optimal, thin solid: SVM) for user 2 of Example 1. $\mathrm{SNR}_1 = \mathrm{SNR}_2 = 20$ dB.

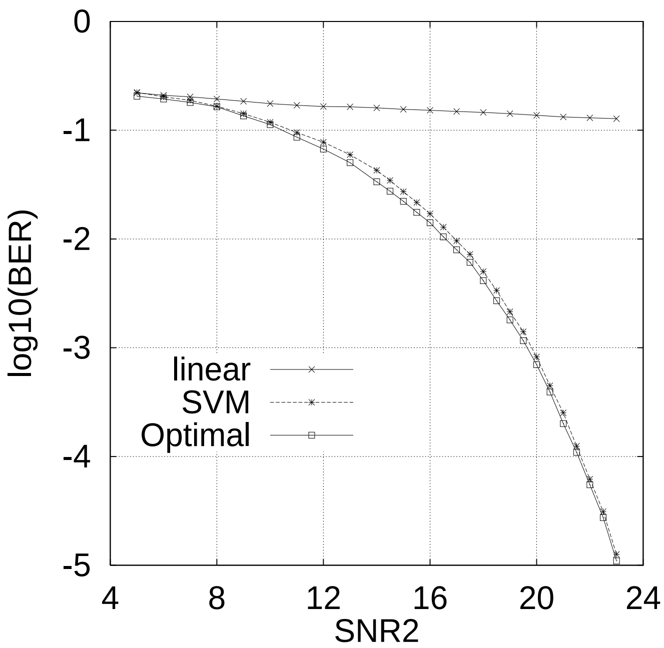

Fig. 4. Performance comparison of three MUDs for user 2 of Example 1. $\mathrm{SNR}_1 = \mathrm{SNR}_2$, and the training data set for SVM has 160 samples.

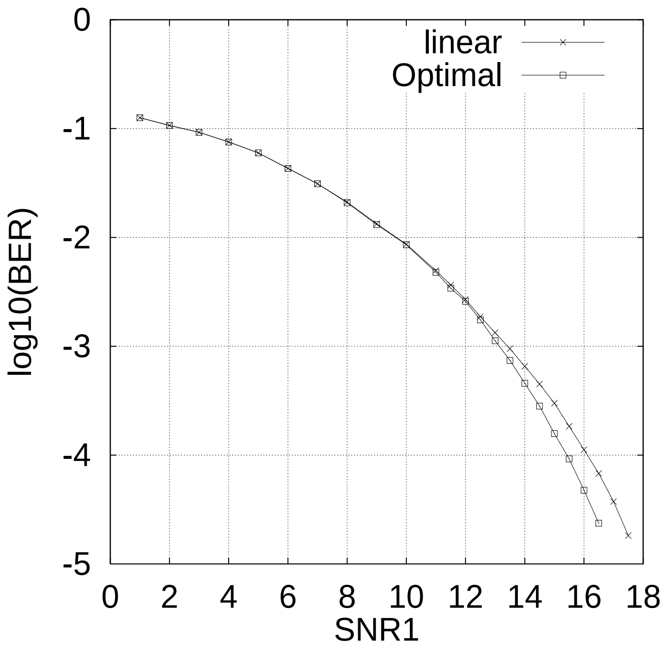

Fig. 5. Performance comparison of two MUDs for user 1 of Example 2. SNR_i, $1 \leq i \leq 3$, are identical.

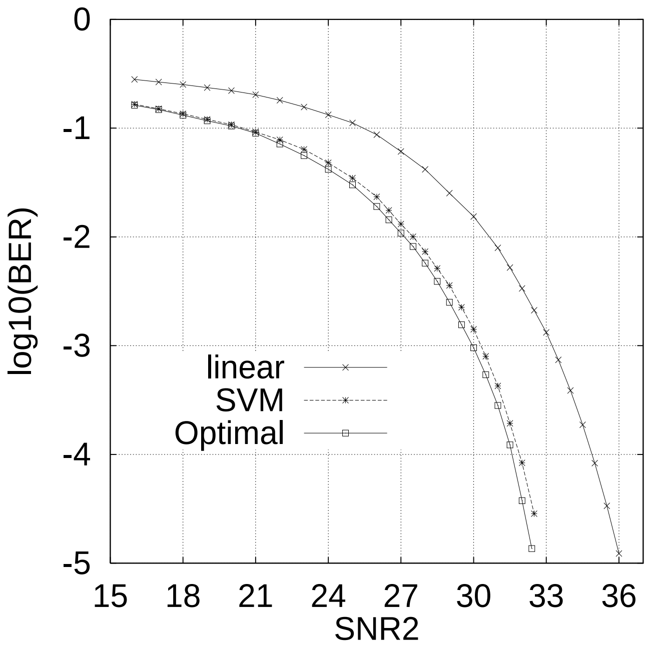

Fig. 6. Performance comparison of three MUDs for user 2 of Example 2. SNR_i, $1 \leq i \leq 3$, are identical, and the training data set for SVM has 640 samples.

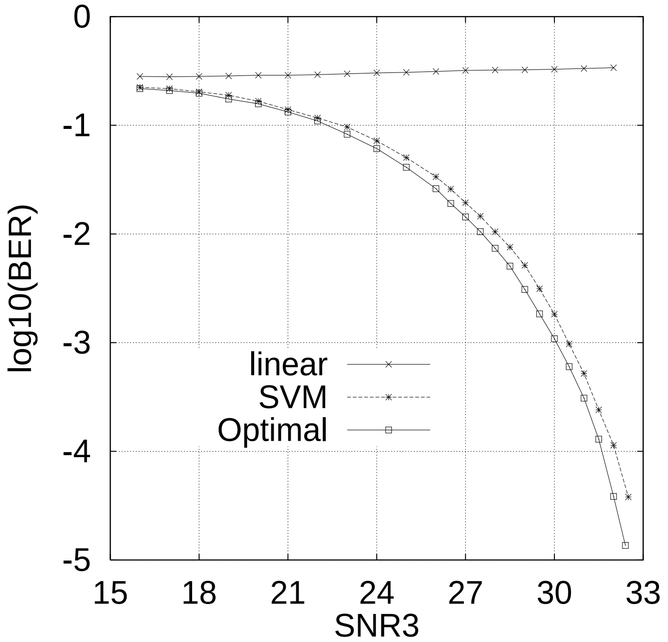

Fig. 7. Performance comparison of three MUDs for user 3 of Example 2. SNR_i, $1 \le i \le 3$, are identical, and the training data set for SVM has 640 samples.

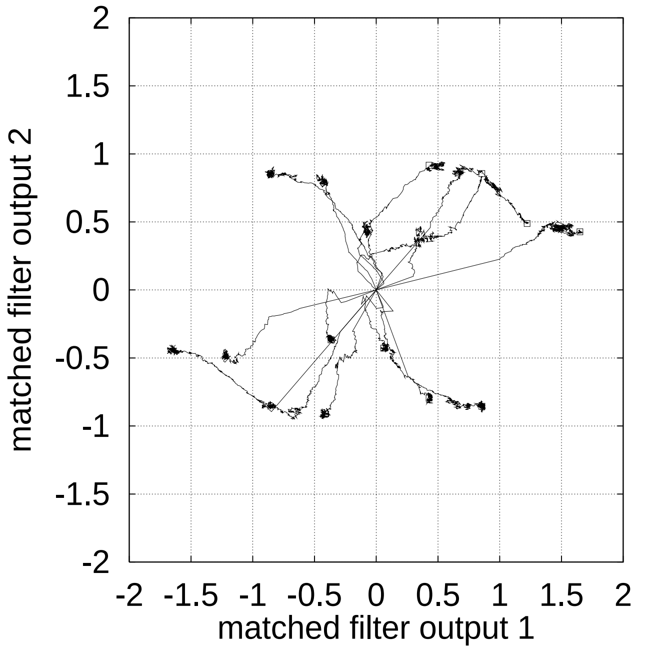

Fig. 8. Cluster center trajectories using the enhanced κ-means clustering for user 2 of Example 1. $\mathrm{SNR}_1 = \mathrm{SNR}_2 =$ 20 dB, initial centers are all placed at the origin, and 10000 data samples are used.

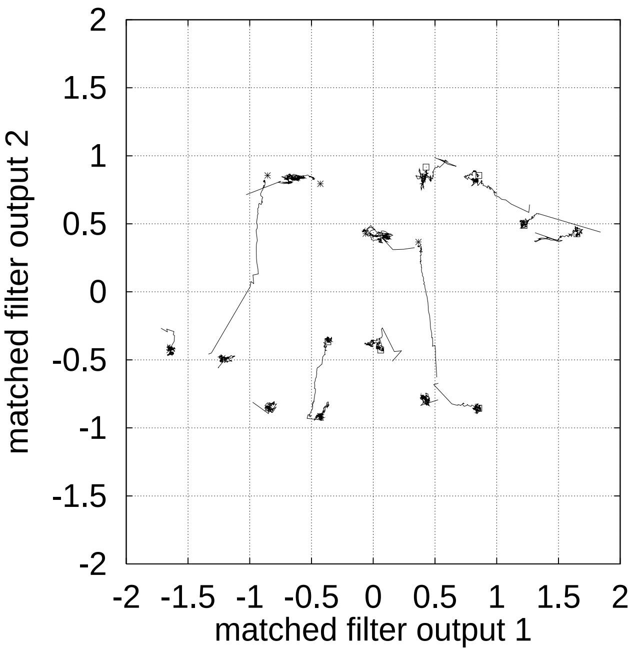

Fig. 9. Cluster center trajectories using the enhanced κ-means clustering for user 2 of Example 1. $\mathrm{SNR}_1 = \mathrm{SNR}_2 =$ 20 dB, the first 16 data points are used as initial centers, and 4000 data samples are used.