DESIGN OF THE OPTIMAL SEPARATING HYPERPLANE FOR THE DECISION FEEDBACK EQUALIZER USING SUPPORT VECTOR MACHINES

S. Chen and C.J. Harris

Department of Electronics and Computer Science
University of Southampton, Southampton SO17 1BJ, U.K.

ABSTRACT

The conventional decision feedback equalizer (DFE) separates the different signal classes using a single hyperplane. It is well known that the popular minimum mean square error (MMSE) design is generally not the optimal minimum bit error rate (MBER) solution. We propose a method of designing the separating hyperplane for the conventional DFE based on support vector machines (SVMs). The SVM design achieves asymptotically the MBER solution and can be computed efficiently.

1. INTRODUCTION

The conventional or linear-combiner DFE is widely used in practice, since it provides a good balance between performance and complexity. Geometrically, a DFE forms a hyperplane in the signal space. The most popular design strategy is the MMSE design with an advantage that it leads to a simple adaptive implementation. A better design in terms of performance is to choose the equalizer coefficients to minimize bit error rate (BER) directly [1],[2]. The performance of the resulting MBER DFE is superior to the MMSE DFE, but its adaptive implementation is computationally much more complex. Moreover, unlike the mean square error surface, which is quadratic, the BER surface can be highly irregular and a gradient algorithm cannot generally guarantee to converge to a global minimum.

The method of SVMs is very powerful in solving various classification and regression problems [3],[4]. The idea of SVMs originates from finding an optimal hyperplane to separate two classes with maximum margin, and this is very relevant to the DFE. We use the method of SVMs to obtain the separating hyperplane for the DFE. The SVM design leads to a well-conditioned quadratic programming defined by a few channel states called support vectors (SVs) that lie on the margin. The SVM solution is unique and can be computed efficiently. Moreover it does not depend on the noise variance and it is asymptotically the MBER solution. Simulation results indicate that the performance of the SVM DFE is virtually indistinguishable from the MBER DFE. Adaptive implementation of the SVM DFE is also discussed.

2. THE DFE STRUCTURE

We will assume that the real-valued channel generates the received signal samples of:

$$y(k) = \sum_{i=0}^{n_a-1} a_i s(k-i) + e(k)\,, \qquad (1)$$

where n_a is the channel impulse response (CIR) length, a_i denotes the channel taps, the Gaussian white noise $e(k)$ has zero mean and variance σ_e^2, and the symbol sequence $\{s(k)\}$ takes values from the set $\{\pm 1\}$. An extension to M-PAM channels will be given later. The DFE studied in this paper produces an estimate $\hat{s}(k-d)$ of $s(k-d)$ by quantizing the filter output of:

$$f(\mathbf{y}(k), \hat{\mathbf{s}}_b(k)) = \mathbf{w}^T \mathbf{y}(k) + \mathbf{b}^T \hat{\mathbf{s}}_b(k)\,, \qquad (2)$$

where $\mathbf{y}(k) = [y(k)\ y(k-1) \cdots y(k-m+1)]^T$ and $\hat{\mathbf{s}}_b(k) = [\hat{s}(k-d-1) \cdots \hat{s}(k-d-n)]^T$ are the observation and past detected symbol vectors, respectively, while $\mathbf{w} = [w_0\ w_1 \cdots w_{m-1}]^T$ and $\mathbf{b} = [b_1 \cdots b_n]^T$ are the coefficient vectors of the feedforward and feedback filters, respectively. Without the loss of generality, the decision delay $d = n_a - 1$, feedforward order $m = n_a$ and feedback order $n = n_a - 1$ are chosen, as this choice is sufficient to guarantee the linear separability.

The observation vector can be expressed as: $\mathbf{y}(k) = F_1 \mathbf{s}_f(k) + F_2\ \mathbf{s}_b(k) + \mathbf{e}(k)$, where $\mathbf{s}_f(k) = [s(k) \cdots s(k-d)]^T$, $\mathbf{s}_b(k) = [s(k-d-1) \cdots s(k-d-n)]^T$, and

$$F_1 = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n_a-1} \\ 0 & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_1 \\ 0 & \cdots & 0 & a_0 \end{bmatrix} \qquad (3)$$

$$F_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ a_{n_a-1} & 0 & \ddots & \vdots \\ a_{n_a-2} & a_{n_a-1} & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ a_1 & \cdots & a_{n_a-2} & a_{n_a-1} \end{bmatrix} \quad (4)$$

are the $m \times (d+1)$ and $m \times n$ CIR matrices, respectively. Under the assumption of correct decision feedback, we have $\hat{\mathbf{s}}_b(k) = \mathbf{s}_b(k)$, and the decision feedback translates the original signal space $\mathbf{y}(k)$ into a new space $\mathbf{r}(k)$:

$$\mathbf{r}(k) \stackrel{\triangle}{=} \mathbf{y}(k) - F_2\, \hat{\mathbf{s}}_b(k)\,. \quad (5)$$

In the translated space, the DFE can be described by:

$$f(\mathbf{r}(k)) = \mathbf{w}^T \mathbf{r}(k)\,. \quad (6)$$

Let the $N_f = 2^{d+1}$ possible sequences of $\mathbf{s}_f(k)$ be $\mathbf{s}_{fj}$, $1 \le j \le N_f$. The set of the noiseless channel states in the translated space, namely $R \stackrel{\triangle}{=} \{\mathbf{r}_j = F_1\, \mathbf{s}_{fj},\, 1 \le j \le N_f\}$, can be partitioned into the two subsets:

$$R^{(\pm 1)} \stackrel{\triangle}{=} \{\mathbf{r}_j \in R : s(k-d) = \pm 1\}\,. \quad (7)$$

Lemma 1 $R^{(+1)}$ and $R^{(-1)}$ are linearly separable.

The MMSE solution $\mathbf{w}_{\mathrm{MMSE}}$ is an example of the separating hyperplane. The non-optimal nature of the MMSE solution however becomes clear in the asymptotic case of large signal to noise ratio (SNR).

Lemma 2

$$\lim_{\mathrm{SNR}\to\infty} \mathbf{w}_{\mathrm{MMSE}} = \left[0\, 0 \cdots 0\ \frac{1}{a_0}\right]^T\,. \quad (8)$$

For the proofs of these two lemmas, see [1].

The weight vector of the optimal separating hyperplane can be obtained by minimizing the BER $P_E(\mathbf{w})$ [1]

$$\mathbf{w}_{\mathrm{MBER}} = \arg\min_{\mathbf{w}} P_E(\mathbf{w})\,, \quad (9)$$

where

$$P_E(\mathbf{w}) = \frac{2}{N_s} \sum_{\mathbf{r}_j \in R^{(+1)}} Q\left(\frac{|\mathbf{w}^T \mathbf{r}_j|}{\|\mathbf{w}\|\sigma_e}\right)\,, \quad (10)$$

$N_s = 2^d$ is the number of states in $R^{(+1)}$ and

$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx\,. \quad (11)$$

3. THE SVM DFE

The minimum distance from the nearest point in R to a separating hyperplane, defined by $\mathbf{w}^T\mathbf{r} = 0$, is given by:

$$\rho(\mathbf{w}) = \min_{\mathbf{r}_i \in R^{(+1)}} \frac{|\mathbf{w}^T \mathbf{r}_i|}{\|\mathbf{w}\|} + \min_{\mathbf{r}_j \in R^{(-1)}} \frac{|\mathbf{w}^T \mathbf{r}_j|}{\|\mathbf{w}\|}\,, \quad (12)$$

which is referred to as the margin. The SVM design finds the hyperplane that maximizes this margin. Since the weight vector $\mathbf{w}$ of the hyperplane $\mathbf{w}^T\mathbf{r} = 0$ is linear dependent, it is appropriate to consider a canonical hyperplane [3] where $\mathbf{w}$ is constrained by

$$\min_{\mathbf{r}_i \in R} |\mathbf{w}^T \mathbf{r}_i| = 1\,. \quad (13)$$

Define the integer set $I_R \stackrel{\triangle}{=} \{i\ :\ \mathbf{r}_i \in R\}$ and the class indicator $y_i = \pm 1$, $\forall \mathbf{r}_i \in R^{(\pm 1)}$. The maximization of the margin (12) with the constraint (13) using the classical Lagrangian theory [5] gives rise to the optimal separating hyperplane:

$$\mathbf{w}_{\mathrm{SVM}} = \sum_{i \in I_R} \bar{g}_i\, y_i\, \mathbf{r}_i\,, \quad (14)$$

where

$$\bar{\mathbf{g}} = \arg\min_{\mathbf{g}} \frac{1}{2} \sum_{i \in I_R} \sum_{j \in I_R} g_i g_j\, y_i y_j\, \mathbf{r}_i^T \mathbf{r}_j - \sum_{i \in I_R} g_i\,, \quad (15)$$

$$g_i \ge 0,\ \forall i \in I_R\,. \quad (16)$$

The optimization problem (15) with (16) is a quadratic programming problem, whose solution $\bar{\mathbf{g}}$ can be computed efficiently. As g_i are the Lagrange multipliers of the primal problem, from the Kuhn-Tucker conditions [5]

$$\bar{g}_i \left(y_i \mathbf{w}_{\mathrm{SVM}}^T \mathbf{r}_i - 1\right) = 0\,, \quad (17)$$

only those points $\mathbf{r}_i$, which satisfy $y_i \mathbf{w}_{\mathrm{SVM}}^T \mathbf{r}_i = 1$, will have non-zero Lagrange multipliers. These points are the SVs [3]. All the SVs lie on the margin and the number of SVs can be very small. Let R_{SV} be the set of SVs. The hyperplane $\mathbf{w}_{\mathrm{SVM}}$ is uniquely determined by R_{SV}.

Thus, the identical solution is obtained by substituting R with R_{SV} in (14) to (16). However, the SVs are unknown *a priori*. We can select a smaller subset R_{Sub} of R, which contains all the SVs. We first point out that, because we restrict the DFE to the linear structure (6) in the translated space, the decision boundary is a hyperplane. If we allow a nonlinear structure, the decision boundary will become a hypersurface. The true optimal solution for the DFE without restricting it to linear filtering is the nonlinear Bayesian solution [6],[7]. Asymptotically, the Bayesian boundary is piecewise linear, consisting of a set of hyperplanes. Each of these hyperplanes is defined by a pair of *dominant* states in $R^{(\pm 1)}$ [8]. The following algorithm can be used to select these pairs of the dominant states that define the set of the asymptotic hyperplanes [9]:

```
FOR r_i^+ ∈ R^(+1)
    FOR r_j^- ∈ R^(-1)
        x = (r_i^+ + r_j^-)/2
        d_0 = ||r_i^+ - x||
        IF (||r_l^+ - x|| > d_0, ∀r_l^+ ∈ R^(+1), l ≠ i) AND
            (||r_l^- - x|| > d_0, ∀r_l^- ∈ R^(-1), l ≠ j)
                R_Sub ← (r_i^+, r_j^-)
        END IF
    NEXT r_j^-
NEXT r_i^+
```

As R_{Sub} can be a smaller subset of R, using it to substitute R in the quadratic programming described by (14) to (16) will result in considerable savings in computation. This is allowed, since $R_{\mathrm{SV}} \subset R_{\mathrm{Sub}}$.

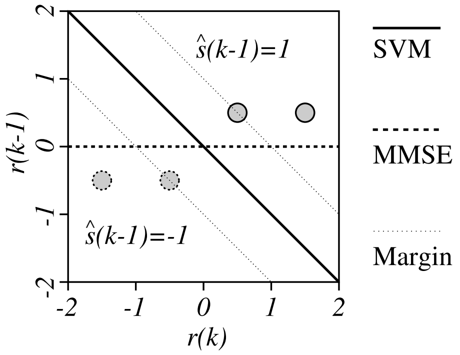

Figure 1: Two asymptotic decision hyperplanes for the CIR of $\mathbf{a} = [0.5\ \ 1.0]^T$.

We comment that $\mathbf{w}_{\mathrm{SVM}}$ is the asymptotic MBER solution, that is, $\lim_{\mathrm{SNR}\to\infty} \mathbf{w}_{\mathrm{MBER}} = \mathbf{w}_{\mathrm{SVM}}$. In general, $\mathbf{w}_{\mathrm{SVM}}$ will not be identical to $\mathbf{w}_{\mathrm{MBER}}$, but the difference is practically negligible for useful SNR conditions. Consider the CIR of $\mathbf{a} = [0.5\ \ 1.0]^T$. The hyperplane of the SVM solution is depicted in Fig. 1. This is identical to the asymptotic MBER solution. When the SNR is reduced to 15 dB, $\mathbf{w}_{\mathrm{SVM}}$ remains unchanged with a slope of -1, but $\mathbf{w}_{\mathrm{MBER}}$ is changed from a slope of -1 to -1.02. Such a small difference will hardly cause any difference in BER performance.

A numerical example with a four-tap CIR of

$$\mathbf{a} = [0.35\ \ 0.80\ \ 1.00\ \ 0.80]^T \tag{18}$$

was used in our simulations. The DFE was accordingly chosen to be $d = 3$, $m = 4$ and $n = 3$. The full set of states R has sixteen points. The subset selection produced a subset R_{Sub} of eight states, four of them being the SVs. The BERs of the MMSE and SVM DFEs with detected symbols being fed back are plotted in Fig.2. The BER of the MBER DFE was practically indistinguishable from that of the SVM DFE. Therefore, the BER curve of the MBER DFE is not included. From Fig.2, it can be seen that the performance of the SVM DFE is significantly better than that of the MMSE DFE. At the BER of 10^{-4}, the SVM DFE has a SNR gain of about 2 dB over the MMSE solution.

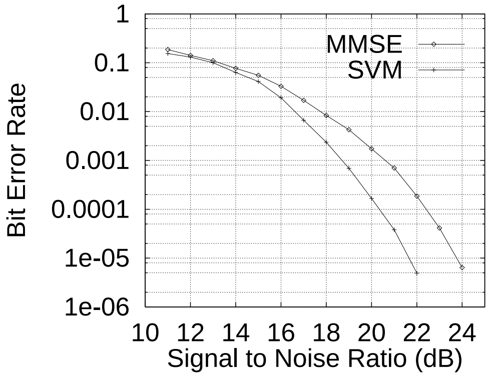

Figure 2: Performance comparison for the CIR of $\mathbf{a} = [0.35\ \ 0.80\ \ 1.00\ \ 0.80]^T$.

4. EXTENSION TO M-PAM

For the M-PAM case, $s(k)$ takes values from the set

$$s_i = 2i - M - 1,\ 1 \le i \le M\,. \tag{19}$$

Let the $N_f = M^{d+1}$ sequences of $\mathbf{s}_f(k)$ be $\mathbf{s}_{fj}$, $1 \le j \le N_f$. The set of the noiseless channel states in the translated space, $R \stackrel{\triangle}{=} \{\mathbf{r}_j = F_1 \mathbf{s}_{fj},\ 1 \le j \le N_f\}$, can be partitioned into M subsets

$$R^{(i)} \stackrel{\triangle}{=} \{\mathbf{r}_j \in R:\ s(k-d) = s_i\},\ 1 \le i \le M\,. \tag{20}$$

Lemma 3 $R^{(i)}$, $1 \le i \le M$, are linearly separable.

The decision boundary of the DFE consists of $M - 1$ parallel hyperplanes defined by: $\{\mathbf{r} : \mathbf{w}^T\mathbf{r} = 2i - M\}$, $1 \le i \le M - 1$. One of the hyperplanes is $\{\mathbf{r} : \mathbf{w}^T\mathbf{r} = 0\}$. The minimum symbol error rate (MSER) solution $\mathbf{w}_{\mathrm{MSER}}$ can be obtained by minimizing the symbol error rate [10].

It can readily be seen from (3) that the subset $R^{(i+1)}$ is a translation of $R^{(i)}$:

$$R^{(i+1)} = R^{(i)} + (s_{i+1} - s_i)\mathbf{a}_{\text{rev}} = R^{(i)} + 2\mathbf{a}_{\text{rev}} \,, \quad (21)$$

where $\mathbf{a}_{\text{rev}} = [a_{n_a-1} \cdots a_1 a_0]^T$. Thus, the weight vector $\mathbf{w}$ is linear dependent and has to obey the constraint:

$$\mathbf{w}^T \mathbf{a}_{\text{rev}} = 1 \,. \quad (22)$$

Similar to the binary case, the SVM solution $\mathbf{w}_{\text{SVM}}$ is determined by the two subsets $R^{(l)}$ and $R^{(l+1)}$, where $l = M/2$, via a simple quadratic programming with the extra constraint (22). This SVM solution is the asymptotic MSER solution.

5. ADAPTIVE IMPLEMENTATION

An indirect adaptation scheme is suitable for adaptive implementation of the SVM DFE. The scheme first estimates a channel model $\hat{\mathbf{a}}$ using the LMS algorithm and then computes the weight vector $\mathbf{w}_{\text{SVM}}$ of the SVM DFE based on the channel estimate $\hat{\mathbf{a}}$. This indirect approach, as opposed to a direct adaptation of the equalizer weight vector using the LMS algorithm, has the advantage of shorter training period. This is because the correlation matrix of the LMS channel estimator has an eigenvalue spread of one, while the correlation matrix of the LMS algorithm for updating the equalizer weight vector can have a large eigenvalue spread.

Implementing the SVM DFE in data storage systems is particularly simple, as in many commercial disk drives, the equalizers are trained at the factory floor and then "frozen" before shipping. Thus training can be done off-line in one go. For time-varying communication links, it is possible to implement the SVM DFE based on a block-by-block adaptation. For example, in some TDMA communication systems, transmission is organized in frames. Each frame contains a training sequence, which can be used in channel estimation. The estimated channel is then used to design the SVM DFE to detect data in the frame. The adaptive SVM DFE is computationally more complex than the adaptive MMSE DFE. The increase in computation, however, can partly be justified by an improved performance.

6. CONCLUSIONS

We have proposed a new DFE design based on the method of SVMs. The SVM DFE, being the asymptotic MBER solution, is superior in terms of performance to the MMSE DFE. The SVM approach results in an efficient quadratic programming solution. Adaptive implementation of the SVM DFE is also discussed, and it is possible to realize the SVM DFE in data storage systems and over slow time-varying communication links.

7. REFERENCES

[1] S. Chen, B. Mulgrew, E.S. Chng and G. Gibson, "Space translation properties and the minimum-BER linear-combiner DFE," *IEE Proc. Communications*, Vol.145, No.5, pp.316–322, 1998.

[2] E. Shamash and K. Yao, "On the structure and performance of a linear decision feedback equalizer based on the minimum error probability criterion," in *Proc. ICC'74*, 1974, pp.25F1–25F5.

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

[4] 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.

[5] M. Minoux, *Mathematical Programming: Theory and Algorithms*. Chichester: John Wiley and Sons, 1986.

[6] S. Chen, B. Mulgrew and S. McLaughlin, "Adaptive Bayesian equaliser with decision feedback," *IEEE Trans. Signal Processing*, Vol.41, No.9, pp.2918–2927, 1993.

[7] 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.

[8] R.A. Iltis, "A randomized bias technique for the importance sampling simulation of Bayesian equalizers," *IEEE Trans. Communications*, Vol.43, No.2/3/4, pp.1107–1115, 1995.

[9] E.S. Chng, B. Mulgrew, S. Chen and G. Gibson, "Optimum lag and subset selection for radial basis function equaliser," in *Proc. 5th IEEE Workshop Neural Networks for Signal Processing* (Cambridge, USA), Aug.31-Sept.2, 1995, pp.593–602.

[10] S. Chen and B. Mulgrew, "The minimum-SER linear-combiner decision feedback equalizer," *IEE Proc. Communications*, accepted, 1999.