Another possible definition of the dual L-class can be provided by carrying out two-dimensional convolution of the DLWD and transformed kernel

$$DC_L(t,\omega) = \int_k \int_l C_L(t-k,\omega-l)DWD_L(k,l)dk\,dl \quad (6)$$

The properties of the dual L-class of TFDs can be derived by generalising the properties of the Cohen class. This has been performed already in [3], where some properties of the Cohen class were generalised for the L-class of TFDs.

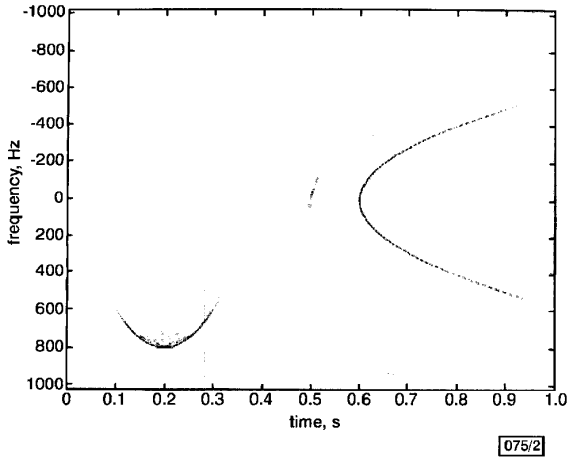

Fig. 2 *MLWD of analysed signal x(t) computed by means of recursive formula (L = 2)*

Some of the most important TFDs of the dual L-class can be obtained by taking an appropriate kernel $c_L(\theta, \tau)$. For example, by setting $c_L(\theta, \tau)$ to be equal to one, the DLWD is obtained; bysetting $c_L(\theta, \tau)$ to be equal to the window function $W(\omega)$ the pseudo-version of the DLWD is obtained. All other TFDs obtained from the DLWD by means of eqn. 5 or eqn. 6 can be understood as the smoothed versions of the DLWD. This also applies to the WD and other TFDs from the Cohen class or to the LWD and other TFDs from the L-class.

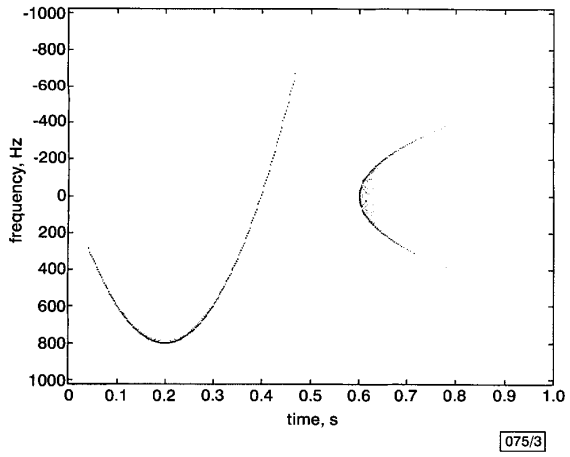

Fig. 3 *Dual MLWD of analysed signal x(t) computed by means of recursive formula (L = 2)*

Examples: In this Section, a performance example of the most important TFDs from the above-mentioned TFD classes is presented. These include the WD from the Cohen class, the LWD from the L-class and the DLWD from the dual L-class. Since the LWD and DLWD obtained by direct computation contain an increased number of cross-terms, recursive formulae for the computation of the LWD and the DLWD [3, 4] are applied in this example.

We consider a two-component signal having the following form

$$x(t) = \exp\left[-j\frac{128000}{3}(t-0.2)^3 + j1600\pi t\right] + IFT\left\{\exp\left[-j\frac{3}{128000}f^3 + j1.2\pi f\right]\right\} \quad (7)$$

The first component is created by a signal in which the instantaneous frequency changes parabolically as a function of time. The second component is created by a signal in which the instantaneous delay changes parabolically as a function of frequency.

The WD of the signal $x(t)$ is depicted in Fig. 1. It is possible to observe the presence of intensive cross-terms, which make the interpretation of the WD more difficult. Despite this, it can be seen that the WD features excellent time-frequency resolution.

The LWD of the signal $x(t)$ computed by means of a recursive formula is presented in Fig. 2. Note the reduced presence of cross-terms together with the good time-frequency resolution comparable with that of the WD. The similar result given in the Fig. 3 can be obtained by means of the DLWD computed by means of a recursive formula. Comparing Figs. 2 and 3, the duality between the LWD and the DLWD can easily be observed.

Conclusions: A new class of TFDs, which we denote the dual L-class of TFDs, has been presented. The dual L-class of TFDs extends the notion of the DLWD, which is the most important TFD from this class. Although the DLWD of a multi-component signal features a greater number of cross-terms than that of the WD, a recursive formula for the DLWD computation, preserving its very good time-frequency resolution together with the reasonably small amount of disrupting cross-terms, can be found.

3 June 2000
Electronics Letters Online No: 20001218
DOI: 10.1049/el:20001218

R. Zetík and D. Kocur (*Technical University of Košice, Faculty of Electrical Engineering and Informatics, Department of Electronics and Multimedial Telecommunications, Letná 9, 04120 Košice, Slovakia*)

References

1 COHEN, L.: 'Time-frequency distributions - a review', *Proc. IEEE*, 1989, **77**, (7), pp. 941–981
2 BOASHASH, B.: 'Time-frequency signal analysis, methods and applications' (Australia, 1992)
3 STANKOVIC, L., STANKOVIC, S., and USKOKOVIC, Z.: 'Time-frequency signal analysis' (Epsilon and Montenegropublic, Podgorica, Montenegro, 1994)
4 ZETÍK, R.: 'Dual version of L-Wigner distribution'. 4th Int. Conf. DSP'99, Slovakia, 1999, pp. 66–69

Optimal decision feedback equaliser for *M*-PAM signals using support vector machine solution

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

The authors propose a method of designing the linear-combiner decision feedback equaliser for *M*-PAM signals using a support vector machine (SVM) technique. This SVM design achieves asymptotically the minimum symbol error rate solution and can be computed efficiently.

Introduction: The linear-combiner decision feedback equaliser (DFE) is widely used in practice, owing to good trade-off between performance and complexity. The most popular design strategy is the minimum mean square error (MMSE) design, which has the advantage of simple adaptive implementation. A better design in terms of symbol error rate (SER) is to choose the equaliser coefficients to minimise SER directly [1]. However, unlike the quadratic mean square error surface, the SER surface can be highly irregular and generally a gradient algorithm cannot guarantee to converge to a global minimum. The idea of using the support vector machine (SVM) technique [2] originates from finding an optimal hyperplane to separate two classes with maximum margin, which is very relevant to equalisation. The method of SVM has been applied to design the DFE for binary signals [3]. This SVM solution is easy to compute and is asymptotically the MSER solution.

We extend the results of [3] to the general case of M-PAM symbols.

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

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

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 M-PAM symbol $s(k)$ takes values from the set $\{s_i = 2i - M - 1,\ 1 \le i \le M\}$. The linear-combiner DFE produces an estimate $\hat{s}(k-d)$ of $s(k-d)$ by quantising the filter output of

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

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, $d = n_a - 1$, $m = n_a$ and $n = n_a - 1$ are chosen, as this choice is sufficient to guarantee the linear separability [1].

Define $\mathbf{s}_f(k) = [s(k) \cdots s(k-d)]^T$, $\mathbf{s}_b(k) = [s(k-d-1) \cdots s(k-d-n)]^T$, and

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

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

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 $\mathbf{r}(k)$ [1]:

$$\mathbf{r}(k) \triangleq \mathbf{y}(k) - F_2\hat{\mathbf{s}}_b(k) \tag{5}$$

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

$$f(\mathbf{r}(k)) = \mathbf{w}^T\mathbf{r}(k) \tag{6}$$

Let the $N_f = M^{d+1}$ 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

$$R \triangleq \{\mathbf{r}_j = F_1\mathbf{s}_{fj} \quad 1 \le j \le N_f\} \tag{7}$$

can be partitioned into the M subsets:

$$R^{(i)} \triangleq \{\mathbf{r}_j \in R : s(k-d) = s_i\} \quad 1 \le i \le M \tag{8}$$

Lemma (1): $R^{(i+1)}$ is a translation of $R^{(i)}$, and $R^{(l)}$, $1 \le l \le M$, are linearly separable.

Proof of lemma (1): From the definitions of $R^{(i)}$ and F_1,

$$R^{(i+1)} = R^{(i)} + (s_{i+1} - s_i)\mathbf{a}_{rev} = R^{(i)} + 2\mathbf{a}_{rev} \tag{9}$$

where $\mathbf{a}_{rev} = [a_{n_a-1} \cdots a_1\ a_0]^T$. To prove the linear separability, consider the hyperplane

$$H_i(\mathbf{r}) \triangleq \hat{\mathbf{w}}^T\mathbf{r} + 2\left(\frac{M}{2} - i\right)\hat{\mathbf{w}}^T\mathbf{a}_{rev} = 0 \tag{10}$$

with

$$\hat{\mathbf{w}} = \left[0\ 0 \cdots 0\ \frac{1}{a_0}\right]^T \tag{11}$$

For any $\mathbf{r}_j \in R^{(i)}$ and any $\mathbf{r}_l \in R^{(i+1)}$, we have $H_i(\mathbf{r}_j) = -1 < 0$ and $H_i(\mathbf{r}_l) = 1 > 0$.

The MMSE solution $\mathbf{w}_{MMSE}$ is another example of separating hyperplane. The non-optimal nature of the MMSE solution is clearly illustrated in the asymptotic case of large signal-to-noise ratio (SNR) [1]: $\mathbf{w}M_{MSE} \to$ as $SNR \to \infty$.

SVM solution: We will use R_- and R_+ to represent $R^{(M/2)}$ and $R^{(M/2+1)}$. The minimum distance from the nearest point in $R_\pm$ to a separating hyperplane, $\mathbf{w}^T\mathbf{r} = 0$, is given by

$$\rho(\mathbf{w}) = \min_{\mathbf{r}_i \in R_+} \frac{|\mathbf{w}^T\mathbf{r}_i|}{\|\mathbf{w}\|} + \min_{\mathbf{r}_j \in R_-} \frac{|\mathbf{w}^T\mathbf{r}_j|}{\|\mathbf{w}\|} \tag{12}$$

which is referred to as the margin. The SVM design finds the hyperplane that maximises 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 [2] where $\mathbf{w}$ is constrained by

$$\min_{\mathbf{r}_i \in R_\pm} |\mathbf{w}^T\mathbf{r}_i| = 1 \tag{13}$$

Define the integer set $I_{R\pm} \triangleq \{i : \mathbf{r}_i \in R_\pm\}$ and the class indicator $y_i = \pm 1$, $\forall\ \mathbf{r}_i \in R_\pm$. The maximisation of the margin (eqn. 12) with the constraint (eqn. 13) using the classical Lagrangian theory [4] gives rise to the optimal separating hyperplane:

$$\mathbf{w}_{SVM} = \sum_{i \in I_{R_\pm}} \bar{g}_i y_i \mathbf{r}_i \tag{14}$$

where

$$\bar{\mathbf{g}} = \arg\min_{\mathbf{g}} \frac{1}{2} \sum_{i \in I_{R_\pm}} \sum_{j \in I_{R_\pm}} g_i g_j y_i y_j \mathbf{r}_i^T\mathbf{r}_j - \sum_{i \in I_{R_\pm}} g_i \tag{15}$$

$$g_i \ge 0 \qquad \forall i \in I_{R_\pm} \tag{16}$$

The optimisation problem (eqns. 15 and 16) is a convex quadratic programming, of which the solution $\bar{\mathbf{g}}$ can be computed efficiently. As g_i are the Lagrange multipliers of the primal problem, from the Kuhn-Tucker conditions [4]

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

only those $\mathbf{r}_i$ which satisfy $y_i\mathbf{w}_{SVM}^T\mathbf{r}_i = 1$ will have non-zero Lagrange multipliers. These points are the support vectors (SVs) [2]. All the SVs lie on the margin and the number of SVs can be very small. The hyperplane $\mathbf{w}_{SVM}$ is uniquely determined by the set of SVs.

From lemma (1), it can be seen that the optimal hyperplane for separating $R^{(i)}$ and $R^{(i+1)}$ is

$$\bar{H}_i(\mathbf{r}) = \mathbf{w}_{SVM}^T\mathbf{r} + c_i = 0 \tag{18}$$

with the bias

$$c_i = 2\left(\frac{M}{2} - i\right)\mathbf{w}_{SVM}^T\mathbf{a}_{rev} \tag{19}$$

We comment that $\mathbf{w}_{SVM}$ is independent of the noise variance and, in the asymptotic case of large SNR, it becomes the MSER solution $\mathbf{w}_{MSER}$. In general, $\mathbf{w}_{SVM}$ will not be identical to $\mathbf{w}_{MSER}$, but the difference is practically negligible for useful SNR conditions.

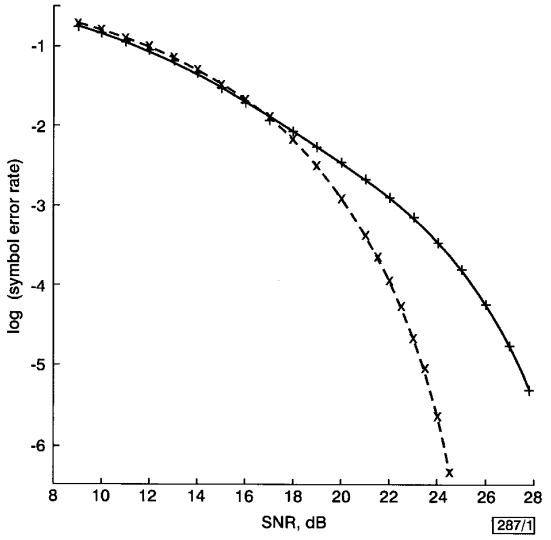

Fig. 1 *Performance comparison for CIR of* $\mathbf{a} = [-0.35\ 0.8\ 0.35]^T$ *with 4-PAM symbols*

—+— MMSE
····×···· SVM

Simulation example: The example had a 3-tap CIR of $\mathbf{a} = [-0.35\ 0.8\ 0.35]^T$ with 4-PAM symbols. The DFE structure was defined by $d = 2$, $m = 3$ and $n = 2$. The subsets $R_{\pm}$ have 32 states, and the SVM solution $\mathbf{w}_{SVM} = [0.5543\ 1.2669\ -1.0699]^T$ is determined by the eight SVs. The MMSE solution for $SNR = 23$dB is $\mathbf{w}_{MMSE} = [0.2227\ 0.5902\ -1.2405]^T$. Fig. 1 depicts the SER curves of the MMSE and SVM DFEs with detected symbols being fed back. For this example, the SERs of the MSER DFE, not included here, are practically indistinguishable from those of the SVM DFE.

Conclusions: We have extended the SVM technique to design the linear-combiner DFE for M-PAM signals. Although the SVM DFE only achieves the MSER performance asymptotically, in practice, the performance of the two DFEs are almost identical. The SVM DFE generally outperforms the MMSE DFE, but it does not have a simple adaptive implementation and can only be implemented indirectly by first estimating a CIR and then designing the equaliser weight vector. This design process however is a simple quadratic programming.

31 May 2000
Electronics Letters Online No: 20001208
DOI: 10.1049/el:20001208

S. Chen, A.K. Samingan and L. Hanzo (*Department of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, United Kingdom*)

E-mail: sqc@ecs.soton.ac.uk

References

1 CHEN, S., and MULGREW, B.: 'The minimum-SER linear-combiner decision feedback equalizer', *IEE Proc. Commun.*, 1999, **143**, (6), pp. 347–353
2 VAPNIK, V.: 'The nature of statistical learning theory' (Springer-Verlag, New York, 1995)
3 CHEN, S., GUNN, S., and HARRIS, C.J.: 'Decision feedback equalizer design using support vector machines', *IEE Proc. Vis. Image Signal Process.*, 2000, **147**, (3), pp. 213–219
4 MINOUX, M.: 'Mathematical programming: theory and algorithms' (John Wiley and Sons, Chichester, 1986)

Recurrent neural network based adaptive filtering technique for the extraction of foetal electrocardiogram

S. Selvan and R. Srinivasan

The authors propose a novel adaptive filtering technique, combining adaptive noise cancellation and adaptive signal enhancement in a single recurrent neural network employing a real time recurrent learning algorithm, which is suitable for the real-time processing of an abdominal foetal electrocardiogram and converges faster to a lower mean squared error.

Introduction: The foetal electrocardiogram (FECG) that is obtained non-invasively by applying a pair of electrodes to the abdomen of a pregnant woman contains a weak foetal signal, a relatively strong maternal ECG (MECG), maternal abdominal muscle noise and power line pickup. To obtain proper information of the foetal heart, it is necessary to improve the signal-to-noise ratio of the abdominal signal. Various techniques [1 – 4] have been developed for the processing of an abdominal FECG. The methods described in [1, 2] involve off-line processing. Since the frequency spectrum of each noise source partially overlaps that of the FECG, adaptive filtering is required to achieve adequate noise reduction. Methods described in [3, 4] employ adaptive noise cancellation and adaptive signal enhancement, respectively. Artificial neural networks can offer the computational power of nonlinear techniques and signal processing techniques have been improved by the application of artificial neural networks [5]. A novel technique achieving both noise cancellation and signal enhancement in a single recurrent neural network is proposed which is applicable to real-time processing of an abdominal FECG. Its performance is compared by processing the corrupted signal using an adaptive noise canceller and a cascaded connection of an adaptive noise canceller and adaptive signal enhancer. Recurrent neural networks using the real-time recurrent learning (RTRL) algorithm [6] are employed to implement each of the above systems. The RTRL algorithm is selected since it is suitable for real-time learning. The underlying mechanism of interfering MECG and abdominal muscle noise is not well defined and hence the assumption of nonlinearity is considered to obtain a better result.

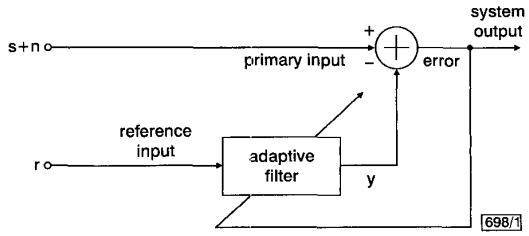

Fig. 1 *Block diagram of adaptive noise canceller*

Adaptive noise canceller and adaptive signal enhancer: The adaptive noise canceller shown in Fig. 1 uses a primary input which consists of a corrupted signal $(s + n)$ and a reference input containing the noise source r correlated in some way with the primary noise n. The signal s and noise n are uncorrelated. The output of the adaptive filter is y. By training the adaptive filter using an on-line error of $(s + n - y)$, the expected value of $E\{(n - y)^2\}$ is locally minimised [3]. In the adaptive signal enhancer shown in Fig. 2, the signals $s1$, $s2$ are correlated and $n2$ is uncorrelated with $s1$, $s2$ and $n1$. The reference input is derived from the primary input by inserting a delay such that decorrelation of the noise components is achieved. The adaptive filter is trained with an on-line error of $(s1 + n1 - y)$. Now y is the least square estimate of the signal $s1$ [4]. However, before employing the adaptive signal enhancer, the adaptive noise canceller can be used to improve the processing [4]. The adaptive filters in both the noise canceller and the signal enhancer can be implemented using neural filters.

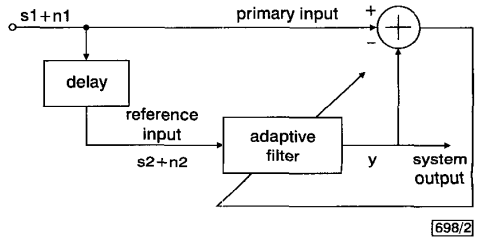

Fig. 2 *Block diagram of adaptive signal enhancer*

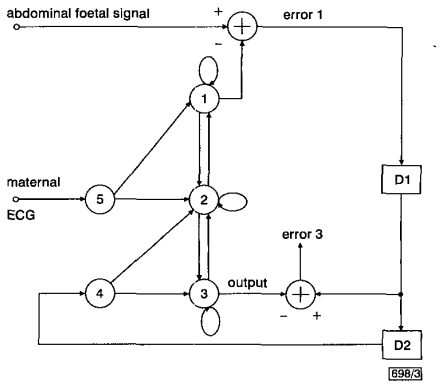

Fig. 3 *Schematic diagram of proposed adaptive filtering technique using recurrent neural network*

Proposed technique: A schematic diagram of the proposed technique is shown in Fig. 3. The network consists of five nodes where nodes 1, 2 and 3 are computational nodes and nodes 4 and 5 are input nodes. The abdominal signal obtained from the mother's abdomen is taken as the primary input. The MECG obtained from the mother is taken as the reference input, which acts as the