MULTI-CRITERIA QUADRATIC PROGRAMMING BASED LOW COMPLEXITY NONLINEAR CHANNEL EQUALISATION

Mohamed Musbah and Xu Zhu

Department of Electrical Engineering and Electronics, University of Liverpool
Brownlow Hill, L69 3GJ, Liverpool, UK
email: {m.musbah,xuzhu}@liverpool.ac.uk

ABSTRACT

We propose a low complexity multi-criteria quadratic programming (MCQP) based approach for nonlinear channel equalisation in wireless communications viewed as a classification problem. Compared to the standard SVM method, the proposed MCQP approach not only provides a better performance by introducing additional criteria to the objective function, but also requires a much lower complexity by solving a set of linear equations. Besides, the MCQP based equaliser achieves nearly the same performance as the optimal Bayesian detector. Furthermore, MCQP demonstrates its effectiveness in tracking the time variation of communication channels.

1. INTRODUCTION

Nonlinear channel equalisation is a challenging analytical problem for many digital communication systems [11]. This is because of the nonlinear inter-symbol interference (ISI), additive white Gaussian noise (AWGN) [2][6] and effects of time-varying channels that can severely degrade the system performance [5]. Further, algorithms performing nonlinear equalisation are often too computationally intensive to be implemented in real time. The process of equalisation is, thus, aimed to reconstruct transmitted symbols based on observations of the corrupted channel. It is desirable that a receiver requires a small set of training data to characterise the transmission channel, hence making better use of bandwidth. Equalisation is treated as a natural inverse filter [10], and the equalizer forms an approximation of the inverse of the distorting channel.

The recent advances in digital signal processing area allow many researchers to consider the problem of designing new techniques for efficient nonlinear equalisers. Among these techniques; the multilayer perceptron (MLP) [8] and wavelet neural networks (WNN) [13]. Despite the impressive performance obtained from these techniques, they are still suffering from one or more of the followings: convergence performance, structure complexity, and optimisation computational complexity. In addition, the support vector machines (SVMs) [11] and kernel based methods [7], associated with their modified least squares support vector machines (LS-SVMs) [6], have become considerably spectacular. Their promising performance can be perceived in applications such as classification, regression, and density function estimation [3]. In SVM algorithms family, the input data are non-linearly transformed to a higher dimensional separable feature space via kernel mapping. A linear decision surface (hyperplane) is then constructed in the feature space. The optimisation in SVM, however, is usually solved by the quadratic programming (QP), which is computationally costly.

In this paper, we propose a low complexity MCQP based approach for nonlinear channel equalisation in wireless communications. As an extension of the generalised learning theory presented in [12], MCQP enables a performance improvement. In addition, MCQP is more computationally efficient, compared to standard SVMs, because of the optimisation technique associated with MCQP only requires solving a set of linear equations. The preceding advantages raise the motivation to use MCQP for nonlinear equalisation. Simulation results confirm performance enhancement of the proposed MCQP over standard SVM based equaliser. It also provides a performance close to that of the optimal Bayesian detector. Furthermore, the MCQP based equaliser considerably demonstrates its robustness to the time variation effects of channel coefficients.

The rest of this paper is organised as follows. Section 2 describes the system model used including the channel model used in the experiments. Section 3 presents the optimal Bayesian symbol detector for binary decisions. Section 4 contains a brief description to the SVM based equaliser. In Section 5, the proposed MCQP based equaliser is presented. The computer simulations settings and results are discussed in Section 6. The conclusion is drawn in Section 7.

2. SYSTEM MODEL

The communication system used in this paper is shown in Figure 1. Assuming baseband transmission and perfect symbol matching filtering associated with real valued data, a discrete-time real channel can be considered to fit the learning-based equalisers adopted. The channel model, according to [11], consists of a deterministic term $y_p(k)$ and random process term $v(k)$ which represent additive Gaussian noise samples. The deterministic term (2) is a polynomial combination of order P_c of a linear, finite impulse response (FIR) filter with length L, that is defined in (1). Hence, for a transmit symbol $d(k) \in \{+1,-1\}$, the output of a general form of nonlinear channel can be modelled as follows

$$y_l(k) = \sum_{l=0}^{L-1} h_l(k) d(k-l) \tag{1}$$

and

$$y_p(k) = \sum_{i=0}^{P_c} c_i y_l^i(k) \tag{2}$$

so that the channel output is

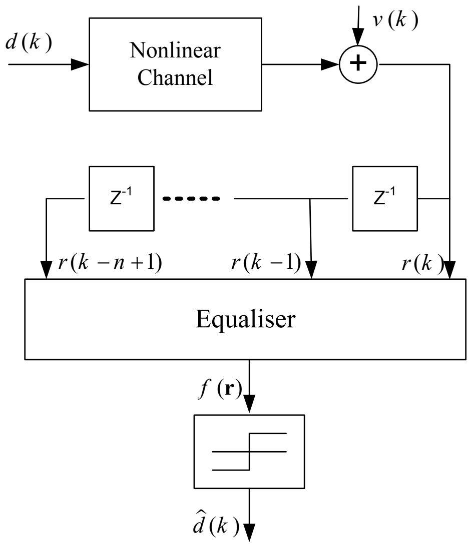


Figure 1: Discrete time system model

$$r(k) = y_p(k) + v(k) \tag{3}$$

The channel output in (3) can be grouped into vectors of length n as

$$\mathbf{r}(k) = [r(k),\, r(k-1), ..., r(k-n+1)] \tag{4}$$

where n is the dimension of received signal vectors that is chosen to match the length of the channel so that the equaliser output in Figure 1 is dependent on the length of the ISI channel (*i.e.* $n = L$). This means that the number of channel states (signal constellation) for the binary detection is 2^{n+1} if no white random noise is added.

3. OPTIMAL BAYESIAN DETECTOR

The utilised optimal detector in this study is the Bayesian or maximum a posteriori (MAP) equaliser. This section presents the binary decision role for the Bayesian detector. Bayesian equaliser works in a symbol-by-symbol manner, with the aim that is to maximise posteriori of a symbol $d(k)$ is being transmitted, given the liklihood and the priori of the observed signal [2].

Given a set of noise-free received vectors (*i.e.* channel states) $\{\mathbf{r}_i^+, \mathbf{r}_i^-\}$, the decision rule is to choose the optimal Bayesian symbol ($\hat{d}(k)$) for a noisy received vector ($\mathbf{r}(k)$, *or* $\mathbf{r}$ for simple notation). $\hat{d}(k)$ is estimated by

$$\hat{d}(k) = \text{sign}\{f_{BAYES}(\mathbf{r})\} = \begin{cases} +1, & f_{BAYES}(\mathbf{r}) \geq 0 \\ -1, & f_{BAYES}(\mathbf{r}) < 0 \end{cases} \tag{5}$$

where *sign{.}* denotes the decision function, and the optimal Bayesian function is given by

$$\begin{aligned} f_{BAYES}(\mathbf{r}) = \;& \textstyle\sum_{i=1}^{N^+} \exp\left(-\left\|\mathbf{r} - \mathbf{r}_i^+\right\|^2 / 2\sigma^2\right) \\ & -\textstyle\sum_{i=1}^{N^-} \exp\left(-\left\|\mathbf{r} - \mathbf{r}_i^-\right\|^2 / 2\sigma^2\right) \end{aligned} \tag{6}$$

where $\mathbf{r}_i^{\pm} = [y_p(k),\, y_p(k-1), ..., y_p(k-n+1)]$ for $d(k) = \pm 1$, and $1 \leq i \leq N^{\pm}$ respectively. N^+ and N^-in (6) refer to the number of channel states for +1, -1 symbols (in this application, $N^+ = N^- = 2^n$). σ^2 denotes the additive white noise power. The Bayesian decision function in (6) assumes equiprobable a priori probabilities and a binary decision solution.

4. SVM BASED EQUALISATION

In this section, we briefly present the support vector machine (SVM) for binary classification problem. More details of SVMs, and their application in digital communications equalisation, can be found in [1, 3, 11, 12]. The fundamental principle of SVC is to find a linear hyperplane ($\mathbf{w}$) , in higher dimensional space, that maximize the distance (margin) between two different patterns. Hence, the optimisation problem in its primal form is defined to

$$\begin{aligned} & \textit{minimise} && \tfrac{1}{2}\left\|\mathbf{w}\right\|^2 + K\textstyle\sum_{i=1}^{P} \xi_i \\ & \textit{subject to} && \bar{d}_i[\phi(\bar{\mathbf{r}}(i))^T \mathbf{w} + b] \geq 1 - \xi_i,\; (i = 1, 2, ..., P) \end{aligned} \tag{7}$$

where b is the bias term representing distance from origin and $\phi(.)$ is nonlinear mapping that will be discussed in Subsection 5.2. $\bar{\mathbf{r}}(i) = [r(i),\, r(i-1), ..., r(i-n+1)]^T$is the observed data from the transmit training sequence $\bar{d}(i) \in \{-1, +1\}$ for $1 \leq i \leq P$, where P is the size of the training set (the pilot size in our application). For simple notation, we will use $\bar{\mathbf{r}}_i$, $\bar{d}_i$ instead of $\bar{\mathbf{r}}(i)$, $\bar{d}(i)$ for the rest of this paper. ξ_i in (7) are slack variables for misclassification tolerance. By introducing Lagrange multipliers (γ's) and applying Karush-Kuhn-Tucker (KKT) conditions [3] for the optimisation of a constrained function, the primal objective function with its constrains in model (7) is converted to dual formulation. The optimisation process, according to [12], is to find the values of $\gamma' s$ of SVC that

$$\begin{aligned} & \textit{maximise} && \textstyle\sum_{i=1}^{P} \gamma_i - \tfrac{1}{2}\sum_{i=1}^{P}\sum_{j=1}^{P} \gamma_i\gamma_j\bar{d}_i\bar{d}_j K(\bar{\mathbf{r}}_i, \bar{\mathbf{r}}_j) \\ & \textit{subject to} && \textstyle\sum_{i=1}^{P} \gamma_i\bar{d}_i = 0 \\ & && 0 \leq \gamma_i \leq K \end{aligned} \tag{8}$$

where P is the size of training set and K is a controlling parameter for the optimisation stability. The dual form facilitates the nonlinear separable data patterns to depend only on the size of training set not the dimension of high feature space. The quadratic programming (QP) [11] is usually used to minimise the objective function in (8). The nonzero values of the optimisation solution are referred to support vectors (SV) set that are used to construct the classifier in (10).

In the detection mode, the estimated symbols from the SVC are expressed as

$$\hat{d}(k) = \text{sign}\{f_{SVC}(\mathbf{r})\} \tag{9}$$

where $f_{SVC}(.)$ is the classification function of the SVC, which is defined as

$$f_{SVC}(\mathbf{r}) = \sum_{i \in SV} \gamma_i \bar{d}_i K(\mathbf{r}, \bar{\mathbf{r}}_i) + b \tag{10}$$

where b is a threshold term that indicates how far the origin is from the hyperplane. *SV* is a set of support vectors which can be partial or full of the training vectors. The powerful advantage of *SV*s in SVMs is that only some of the training vectors are used in the classification stage, hence, a huge saving in detection complexity can be obtained.

5. MCQP BASED EQUALISATION

In this section, we present the multi-criteria convex quadratic programming (MCQP) model [9] that is used to construct the proposed equaliser. The MCQP uses training data, similar to SVM, to estimate the decision function for the detection stage. The estimated function is, then, used to perform the classification to the testing transmission data.

For a binary classification problem, the idea of MCQP model is based on maximising the external distance between the two classes' groups and minimising the internal distance within the same class group. This model has two significant advantages; the first is its relatively low complexity since it only needs to solve a linear set of equations. The second advantage is the performance enhancement due to the introduction of internal distance to the optimisation object function. Furthermore, kernel functions can also be used to solve nonlinear patterns. The following subsections develop the model formulation for linearly and non-linearly separable patterns.

5.1 MCQP for Linearly Separable Patterns

Same as that in the SVC, the data patterns are separated by a hyperplane of direction ($\mathbf{w} = [w_1, w_2, ..., w_n]^T$, where n is the data pattern dimension) and a scalar distance b from the origin. The MCQP model is formulated as to

$$\begin{aligned} &minimise \quad \tfrac{1}{2}\|\mathbf{w}\|^2 + A\textstyle\sum_{i=1}^{P}\alpha_i^2 - B\sum_{i=1}^{P}\beta_i \\ &subject\,to \quad \bar{d}_i(\bar{\mathbf{r}}_i^T\mathbf{w} - b) = -\alpha_i + \beta_i, \ (i = 1, 2, ..., P) \end{aligned} \tag{11}$$

where A, B are arbitrary pre-defined model parameters that control the optimisation objectives. $\alpha_i, \beta_i \geq 0$ represent the slack distances for misclassification errors and the distances of correctly classified points from the hyperplane respectively.

Assuming $\alpha_i = 0$ for correctly classified points and $\beta_i = 0$ for misclassified points, and by introducing $\eta_i = \alpha_i - \beta_i$, model (11) can be rewritten as

$$\begin{aligned} &minimise \quad \tfrac{1}{2}\|\mathbf{w}\|^2 + \tfrac{1}{2}A\textstyle\sum_{i=1}^{P}\eta_i^2 - B\sum_{i=1}^{P}\eta_i + \tfrac{1}{2}Cb^2 \\ &subject\,to \quad \bar{d}_i(\bar{\mathbf{r}}_i^T\mathbf{w} - b) = -\eta_i, \ (i = 1, 2, ..., P) \end{aligned} \tag{12}$$

the new term $\left(\frac{1}{2}Cb^2\right)$ in (12) is introduced to add strong convexity to the objective function [9]. The weight C is an arbitrary positive number. By introducing Lagrange multipliers (θ_i), The optimisation model in (12) including the Lagrange function (13) can be obtained as follows

$$\begin{aligned} L_1(\mathbf{w}, b, \eta, \theta) = \; & \tfrac{1}{2}\|\mathbf{w}\|^2 + \tfrac{1}{2}A\textstyle\sum_{i=1}^{P}\eta_i^2 - B\sum_{i=1}^{P}\eta_i \\ & + \tfrac{1}{2}Cb^2 - \textstyle\sum_{i=1}^{P}\theta_i[\bar{d}_i(\bar{\mathbf{r}}_i^T\mathbf{w} - b) + \eta_i]. \end{aligned} \tag{13}$$

A matrix-vector notation is adopted for simple notation conventions. Let $\theta = [\theta_1, \theta_2, ..., \theta_P]^T$, $\eta = [\eta_1, \eta_2, ..., \eta_P]^T$ and $\mathbf{e} = [1, 1, ..., 1]^T$ be column vectors of P dimension. $\mathbf{R} = [\bar{\mathbf{r}}_1, \bar{\mathbf{r}}_2, ..., \bar{\mathbf{r}}_P]^T$ is a $P \times n$ matrix, and $\bar{\mathbf{D}}$ is a diagonal matrix of $diag(\bar{d}_1, \bar{d}_2, ..., \bar{d}_P)$. The optimal solution of (13) can be obtained from the first derivatives so that

$$\begin{aligned} &\tfrac{\delta}{\delta \mathbf{w}}L_1 = \mathbf{w} - \mathbf{R}^T\bar{\mathbf{D}}\theta = 0, \\ &\tfrac{\delta}{\delta b}L_1 = Cb - \mathbf{e}^T\bar{\mathbf{D}}\theta = 0, \\ &\tfrac{\delta}{\delta \eta}L_1 = A\eta + B\mathbf{e} - \theta = 0, \\ &\tfrac{\delta}{\delta \theta}L_1 = \bar{\mathbf{D}}(\mathbf{R}\mathbf{w} - b\mathbf{e}) + \eta = 0, \end{aligned} \tag{14}$$

thus, by simple manipulation to the equations in (14), the optimum solution can be expressed as

$$\theta = \left[\frac{1}{A}\mathbf{I} + \bar{\mathbf{D}}\left(\mathbf{R}\mathbf{R}^T + \frac{1}{B}\mathbf{e}\mathbf{e}^T\right)\bar{\mathbf{D}}\right]^{-1}\left[\frac{B}{A}\mathbf{e}\right] \tag{15}$$

5.2 MCQP for Non-linearly Separable Patterns

For nonlinear separable clouds of data, the kernel mapping [7] is utilised. By applying nonlinear mapping, through the transformation function $\phi(.)$, the original nonlinear input data space is transformed to a high dimension linearly separable feature space. The kernel mapping, however, can transform the inner product of the higher data vectors without explicitly knowing $\phi(.)$. Therefore, the kernel mapping is defined as

$$K(\bar{\mathbf{r}}_i, \bar{\mathbf{r}}_j) = \phi(\bar{\mathbf{r}}_i)^T\phi(\bar{\mathbf{r}}_j) \tag{16}$$

and

$$\mathbf{K}(\mathbf{R}, \mathbf{R}^T) = \begin{bmatrix} K(\bar{\mathbf{r}}_1, \bar{\mathbf{r}}_1) & K(\bar{\mathbf{r}}_1, \bar{\mathbf{r}}_2) & \cdots & K(\bar{\mathbf{r}}_1, \bar{\mathbf{r}}_P) \\ K(\bar{\mathbf{r}}_2, \bar{\mathbf{r}}_1) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ K(\bar{\mathbf{r}}_P, \bar{\mathbf{r}}_1) & & \cdots & K(\bar{\mathbf{r}}_P, \bar{\mathbf{r}}_P) \end{bmatrix} \tag{17}$$

From model (12) and its optimal conditions (14), and by substituting $\mathbf{w} = \mathbf{R}^T\bar{\mathbf{D}}\theta$, and replacing $\mathbf{R}\mathbf{R}^T$ by $\mathbf{K}(\mathbf{R}, \mathbf{R}^T)$, we can reformulate the model as in (18) below

$$\begin{aligned} &minimise \quad \tfrac{1}{2}\|\theta\|^2 + \tfrac{1}{2}A\textstyle\sum_{i=1}^{P}\eta_i^2 - B\sum_{i=1}^{P}\eta_i + \tfrac{1}{2}Cb^2 \\ &subject\,to \quad \bar{\mathbf{D}}\left(\mathbf{K}(\mathbf{R}, \mathbf{R}^T)\bar{\mathbf{D}}\theta - b\mathbf{e}\right) = -\eta \end{aligned} \tag{18}$$

The Lagrange function to this model is

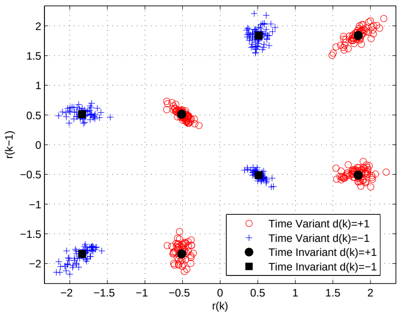
$$\begin{aligned} L_2(\theta, b, \eta, \rho) = \; & \tfrac{1}{2}\|\theta\|^2 + \tfrac{1}{2}A\textstyle\sum_{i=1}^{P}\eta_i^2 - B\sum_{i=1}^{P}\eta_i + \tfrac{1}{2}Cb^2 \\ & -\rho^T\left[\bar{\mathbf{D}}\left(\mathbf{K}(\mathbf{R}, \mathbf{R}^T)\bar{\mathbf{D}}\theta - b\mathbf{e}\right) + \eta\right] \end{aligned} \tag{19}$$

where $\rho = [\rho_1, \rho_2, ..., \rho_P]^T$ are Lagrange multipliers for nonlinear scenario. The optimality conditions of model (19) are then expressed by

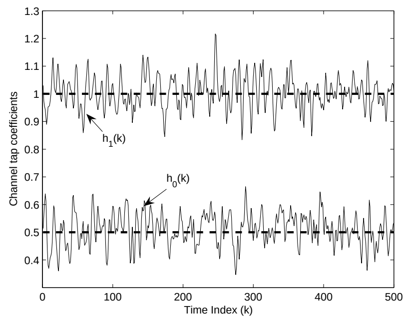
$$\begin{aligned} &\tfrac{\delta}{\delta \theta}L_2 = \theta - \bar{\mathbf{D}}\left(\mathbf{K}(\mathbf{R}, \mathbf{R}^T)\right)^T\bar{\mathbf{D}}\rho = 0, \\ &\tfrac{\delta}{\delta b}L_2 = Cb - \mathbf{e}^T\bar{\mathbf{D}}\rho = 0, \\ &\tfrac{\delta}{\delta \eta}L_2 = A\eta + B\mathbf{e} - \rho = 0, \\ &\tfrac{\delta}{\delta \rho}L_2 = \bar{\mathbf{D}}\left(\mathbf{K}(\mathbf{R}, \mathbf{R}^T)\bar{\mathbf{D}}\theta - b\mathbf{e}\right) + \eta = 0 \end{aligned} \tag{20}$$

Hence, the optimal solution is given by

$$\rho = \left[\frac{1}{A}\mathbf{I} + \bar{\mathbf{D}}\left(\mathbf{K}(\mathbf{R}, \mathbf{R}^T)\left(\mathbf{K}(\mathbf{R}, \mathbf{R}^T)\right)^T + \frac{1}{B}\mathbf{e}\mathbf{e}^T\right)\bar{\mathbf{D}}\right]^{-1}\left[\frac{B}{A}\mathbf{e}\right] \tag{21}$$



(a)



(b)

Figure 2: Channel models; (a) Signal constellation for noise free channel states. (b) Channel taps coefficients.

5.3 Symbol Detection by MCQP Equaliser

For digital communication channel equalisation, the training stage is accomplished through transmitting pilot symbols $(\bar{\mathbf{D}})$, then receiving their corresponding channel outputs $(\mathbf{R})$. Thus, by applying (21), the equaliser parameters are evaluated. Then, the symbol estimation, for an observed received channel output $(\mathbf{r})$, is estimated by the function (22) for detection process.

$$f_{MCQP}(\mathbf{r}) = \left(\mathbf{K}(\mathbf{r},\mathbf{R}^T)\left(\mathbf{K}(\mathbf{r},\mathbf{R}^T)\right)^T + \frac{1}{C}\mathbf{e}^T\right)\bar{\mathbf{D}}\rho \qquad (22)$$

where, for binary signalling, the estimate of $d(k)$ decision is given by

$$\hat{d}(k) = \text{sign}\{f_{MCQP}(\mathbf{r})\} = \begin{cases} +1, & f_{MCQP}(\mathbf{r}) \geq 0 \\ -1, & f_{MCQP}(\mathbf{r}) < 0 \end{cases} \qquad (23)$$

6. SIMULATION RESULTS

In this section, we present the computer simulation configurations and settings of the proposed system followed by a discussion to the obtained results. To simplify visualisation, a channel model of 2 FIR taps were used (*i.e.* $L = n = 2$), and the general channel output is defined as [5]

$$r(k) = y(k) + \mu y^3(k) + v(k) \qquad (24)$$

where

$$y(k) = h_0(k)d(k) + h_1(k)d(k-1) \qquad (25)$$

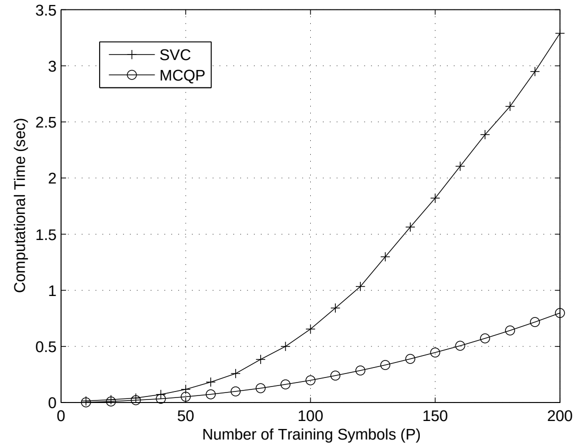


Figure 3: Training computational complexity comparison.

For all simulations, μ is chosen to be 0.1 and $[h_0(k),\, h_1(k)] = [0.5,\, 1]$ for time invariant model. For time variant scenario, according to [4], $h_0(k)$, $h_1(k)$ are two time-varying coefficients. These coefficients were generated by passing white Gaussian noise of power $\beta^2 = 0.01$, and centred around $[0.5,\, 1]$, through a Butterworth low pass filter. The normalised cuttoff frequency (f_D) is 0.15 representing a Doppler shift relative to symbol rate. The noise-free channel states, for both time invariant and time variant, are shown in Figure 2. The FIR channel coefficients variations are plotted in Figure 2(b).

The MCQP parameters settings for simulations were as follows; $A = 0.9$, $B = 0.1$and $C = 0.5$ based on empirical tests. The training symbols (pilot) is set to $P = 100$. For nonlinear transformation by kernel function, the Gaussian radial basis function (GRBF) is used as a preferred kernel for communication applications [1]. The GRBF is defined as

$$K(\mathbf{r}_a, \mathbf{r}_b) = \exp\left(\frac{-\|\mathbf{r}_a - \mathbf{r}_b\|^2}{2\sigma^2}\right) \qquad (26)$$

where σ is the kernel width parameter that is proportional to the additive noise. The SVC is used for comparison. The controlling parameter for the SVC (K) is set to 10. The kernel function is chosen similar to that in the proposed MCQP for fair comparison.

The computational complexity of the proposed equaliser for training is tested in terms of computer execution time. Figure 3 shows a comparison between the proposed equaliser and SVC equaliser training time for different pilot sizes. Results confirm the massive reduction of the proposed equaliser complexity where the computational complexity increase is almost linear (*i.e.* $\approx O(P)$) with respect to pilot size. This is a massive reduction compared to that of SVC where the trend follows a quadratic order (*i.e.* $O(P^2)$).

The resulting BER curves of the Bayesian, SVC and MCQP based equalisers are depicted in Figure 4. The results show the superb performance of the proposed equaliser and its convergence to the optimal Bayesian detector, especially for SNR levels over 10 dB. It is also shown that the high capability of tracking the time variation with the proposed equaliser, where the MCQP performance for time vari-

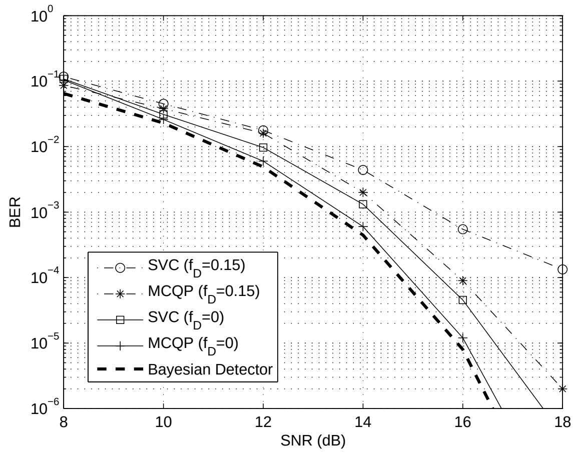


Figure 4: BER performance of MCQP equaliser

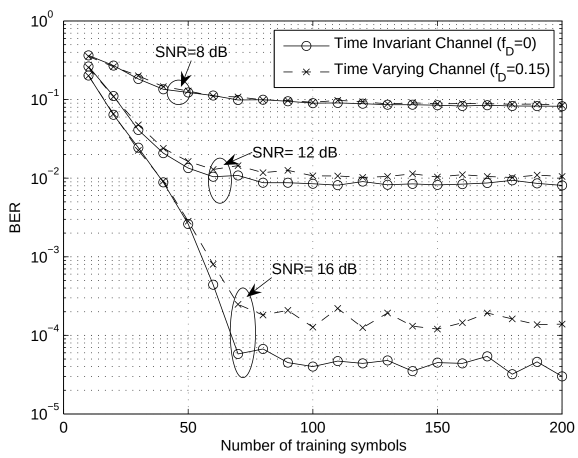


Figure 5: MCQP learning curves

ant channel, with fast channel variation, is almost identical to the SVC performance for corresponding static channel.

Figure 5 shows the learning curves of the proposed equaliser. By stating learning curve, it is meant to describe finding the sufficient number of training data that guarantees best performance for a particular conditions. The results in Figure 5 consider three levels of SNR $(8, 12 \text{ and } 16\,dB)$ for both time invariant and time variant scenarios. Curves show steadiness after approximately 80 symbols of pilot, hence a pilot of size 100 were chosen in the simulations. Moreover, curves confirm the close performance between time invariant and time variant scenarios.

7. CONCLUSION

The MCQP method of classification has been applied for nonlinear channel equalisation, which demonstrates a performance close to that of the optimal Bayesian detector. Compared to the SVM based equaliser, the proposed MCQP based approach has two advantages. First, it introduces the internal distance criteria to the objective function which improves the equalisation performance. Second, the optimisation in MCQP requires solving a linear set of equations, hence, a considerable reduction in the training computational complexity can be attained. Simulation results show that the proposed MCQP based equaliser outperforms, in terms of BER, the standard SVM based equaliser as well as reducing the training computational complexity significantly. Furthermore, the proposed equaliser has shown superb robustness to the time variation of the communication channels.

REFERENCES

[1] S. Chen, S. Gunn, and C. Harris. Decision feedback equalizer design using support vector machines. In *IEE Proc. Vision, Image and Signal Processing*, 2000.

[2] S. Chen, B. Mulgrew, and P. Grant. A clustering technique for digital communications channel equalization using radial basis function networks. *Neural Networks, IEEE Transactions on*, 4(4):570–590, Jul 1993.

[3] N. Christiani and Shawe-Taylor. *An introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.

[4] C. F. N. Cowan and S. Semnani. Time-variant equalization using a novel nonlinear adaptive structure. *International Journal of Adaptive Control and Signal Processing*, 12(2):195–206, 1998.

[5] Q. Liang and J. M. Mendel. Equalization of nonlinear time-varying channels using type-2 fuzzy adaptive filters. *IEEE Transactions on Fuzzy Systems*, 8(5):551–563, October 2000.

[6] C.-J. Lin, S.-J. Hong, and C.-Y. Lee. Using least squares support vector machines for adaptive communication channel equalization. *International Journal of Applied Sciences and Engineering*, 3(1):51–59, April 2005.

[7] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf. An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, 12(2):181–201, March 2001.

[8] J. C. Patra and R. N. Pal. A functional link artificial neural network for adaptive channel equalization. *Signal Process.*, 43(2):181–195, 1995.

[9] Y. Peng, G. Kou, Y. Shi, and Z. Chen. A multi-criteria convex quadratic programming model for credit data analysis. *Decis. Support Syst.*, 44(4):1016–1030, 2008.

[10] J. Proakis. *Digital communications*. McGroaw Hill, New York, NY, 4th edition, 2000.

[11] D. J. Sebald and J. A. Buklew. Support vector machine technique for nonlinear equalization. *IEEE Transactions on Signal Processing*, 48(11):3217–3226, Jun. 2000.

[12] V. Vapnik. *The nature of statistical learning theory*. Springer, 2001.

[13] J. Zhang, G. Walter, Y. Miao, and W. N. W. Lee. Wavelet neural networks for function learning. *Signal Processing, IEEE Transactions on*, 43(6):1485–1497, Jun 1995.