Blind Equalization of Constant Modulus Signals Using Support Vector Machines

Ignacio Santamaría, *Member, IEEE*, Carlos Pantaleón, *Member, IEEE*, Luis Vielva, and
Jesús Ibáñez, *Student Member, IEEE*

Abstract—In this paper, the problem of blind equalization of constant modulus (CM) signals is formulated within the support vector regression (SVR) framework. The quadratic inequalities derived from the CM property are transformed into linear ones, thus yielding a quadratic programming (QP) problem. Then, an iterative reweighted procedure is proposed to blindly restore the CM property. The technique is suitable for real and complex modulations, and it can also be generalized to nonlinear blind equalization using kernel functions. We present simulation examples showing that linear and nonlinear blind SV equalizers offer better performance than cumulant-based techniques, mainly in applications when only a small number of data samples is available, such as in packet-based transmission over fast fading channels.

Index Terms—Blind equalization, nonlinear blind equalization, support vector machines.

I. INTRODUCTION

IN many communication systems, digital signals are transmitted through an unknown bandlimited channel with severe intersymbol interference (ISI). When a training sequence is not available, blind equalization techniques must be used to recover the input signal. These techniques exploit the knowledge about the statistical properties of the input signal or the structure of the channel [1].

A number of blind algorithms are based on stochastic gradient descent (SGD) minimization (online techniques) of a specially designed non-MSE cost function (to this class belongs the widely used CMA [2]). Other algorithms collect a block of data (batch techniques) and iteratively maximize a cost function based on cumulants (for instance, the so-called "super-exponential" algorithm by Shalvi and Weinstein [3], [4]) or use cumulant matching methods, as in the works presented by Tugnait [5] and by Hatzinakos and Nikias [6].

In burst data transmission over fast fading channels, blind algorithms must be able to remove a sufficient level of ISI by using a short block of data. In this case, stochastic gradient descent algorithms, which typically suffer from slow convergence, cannot be employed. Similarly, batch cumulant-based algorithms also offer poor performance in this situation.

In this paper, we propose an alternative blind equalization technique for CM signals, which is expected to require fewer data samples than SGD and cumulant-based algorithms. Blind equalization is formulated as a support vector regression (SVR) problem [7], and an iterative procedure, which is denoted as iterative reweighted quadratic programming (IRWQP), is proposed to find the optimal regressor. Support vector machines (SVMs) have been successfully applied to linear and nonlinear supervised equalization problems [8]–[11]. In these works, the equalization problem is viewed as a supervised classification problem (with a training sequence), and the corresponding SV classifier is derived. In this paper, the problem is formulated as a nonsupervised regression problem: Only the knowledge about the CM property of the input signal is exploited.

Recently, some techniques have been proposed to formulate the blind equalization problem either as a quadratic problem with binary constraints [12] or as a convex optimization method subject to some linear and semidefiniteness constraints [13]. In both cases, the problem is solved via semidefinite programming (SDP) techniques. Similarly to these approaches, here, we formulate a convex problem that has a global optimal solution, but, in addition to this, the proposed solution has several attractive features: It is derived from the powerful theory of SV machines; efficient quadratic programming (QP) implementations can be used [14]; and, finally, it can readily be extended to nonlinear blind equalizers. Some simulation examples show the advantages of this technique in comparison to cumulant-based algorithms.

The rest of this paper is organized as follows: In Section II, the problem of blind equalization is briefly reviewed. In Section III, blind equalization of CM signals is formulated as an SVR problem, and the QP problems for real and complex modulations are derived. The technique is also extended to nonlinear blind equalizers. The iterative reweighted quadratic programming (IRWQP) technique is presented in Section IV, whereas Section V is devoted to discuss some implementation details as well as to the selection of the SVR parameters. Some simulation results are presented in Section VI, comparing the performance of linear and nonlinear blind SVM equalizers against cumulant-based techniques. Finally, Section VII presents the conclusions and points out some lines for further research.

II. PROBLEM FORMULATION

We consider a baud-rate sampled baseband representation of the digital communication system. A sequence of i.i.d. symbols belonging to a CM alphabet $\{|s_i| = 1\}$ is sent through a linear

time-invariant channel with coefficients h_i. The symbols and the corresponding channel can be real (a BPSK modulation, for instance) or complex (QPSK or M-PSK being the most widely used modulation formats). The resulting channel output can be expressed as

$$x_i = \sum_n h_n s_{i-n} + e_i$$

where e_i is a complex or real zero-mean white Gaussian noise.

The objective of a blind linear equalizer is to remove the ISI at its output without using any training sequence. Typically, the equalizer is designed as an FIR filter with M coefficients $\mathbf{w}$; then, its output is given by

$$y_i = \sum_{n=0}^{M-1} w_n x_{i-n} = \mathbf{w}^T \mathbf{x}_i. \tag{1}$$

The method proposed by Shalvi and Weinstein [3], which will be used for comparison purposes, maximizes $|K_y|$, subject to the constraint $E\left[|y_i|^2\right] = E\left[|s_i|^2\right]$, where K_y is the kurtosis of y_i, which is defined as

$$K_y = E\left[|y_i|^4\right] - 2\left(E\left[|y_i|^2\right]\right)^2 - \left|E\left[y_i^2\right]\right|^2.$$

III. SV REGRESSION FOR CM BLIND EQUALIZATION

A. Real Channel and Binary Input

In this section, the problem of blind equalization of a CM signal is formulated within the SVR framework. To this end, we first consider the case where the channel coefficients are real and the input signal is binary $\{s_i \in \pm 1\}$. This case was discussed by the authors in [15].

Suppose we are given a set of N observations at the channel output: $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$, where $\mathbf{x}_i = (x_i, x_{i-1}, \ldots, x_{i-M+1})^T$. Then, the goal of a linear blind equalizer is to restore at its output the CM property of the digital communications signal, i.e., $(\mathbf{w}^T \mathbf{x}_i)^2 = y_i^2 = 1$, for $i = 1, \ldots, N$.

According to the structural risk minimization (SRM) principle [7], to estimate a linear equalizer with precision ϵ, one minimizes

$$J(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}|1 - (\mathbf{w}^T\mathbf{x}_i)^2|_\epsilon \tag{2}$$

where $C > 0$ is a penalty value, and

$$|1 - (\mathbf{w}^T\mathbf{x}_i)^2|_\epsilon = \max\{0, |1 - (\mathbf{w}^T\mathbf{x}_i)^2| - \epsilon\}$$

the so-called Vapnik's ϵ-insensitive loss function.

This cost function establishes a tradeoff between the complexity of the equalizer and a term that penalizes those outputs of the equalizer with deviations over the ideal unit modulus output larger than ϵ. By introducing a set of positive slack variables ξ_i and $\tilde{\xi}_i$, the minimization of (2) can be rewritten as the following constrained optimization problem: To minimize

$$L(\mathbf{w}, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}(\xi_i + \tilde{\xi}_i)$$

subject to

$$(\mathbf{w}^T\mathbf{x}_i)^2 - 1 \le \epsilon + \xi_i \tag{3}$$
$$1 - (\mathbf{w}^T\mathbf{x}_i)^2 \le \epsilon + \tilde{\xi}_i$$
$$\xi_i, \tilde{\xi}_i \ge 0 \tag{4}$$

for all $i = 1, \ldots, N$.

In the conventional SVM approach for regression and function approximation, the inequality constraints are linear in the unknown $\mathbf{w}$, thus yielding a quadratic programming (QP) problem that can be efficiently solved [14]. The proposed constraints (3) and (4) for blind equalization of CM signals are, however, quadratic with respect to the coefficients of the equalizer. For this reason, a direct introduction of the constraints into the cost function, by means of Lagrange multipliers, does not render a QP problem.

At this point, it is convenient to stress that in our formulation, the linear regressor $y_i = \mathbf{w}^T\mathbf{x}_i$ is in fact the equalizer's output (1). With the goal of transforming the quadratic inequalities into linear ones, let us rewrite the squared modulus of the output of the equalizer as

$$(\mathbf{w}^T\mathbf{x}_i)^2 = y_i(\mathbf{w}^T\mathbf{x}_i).$$

Now, considering y_i fixed, the inequalities (3) and (4) become linear in $\mathbf{w}$ and can be rewritten as

$$y_i(\mathbf{w}^T\mathbf{x}_i) - 1 \le \epsilon + \xi_i$$
$$1 - y_i(\mathbf{w}^T\mathbf{x}_i) \le \epsilon + \tilde{\xi}_i$$

where, in this way, the blind equalization problem can be formulated within the conventional support vector framework. In particular, the optimization problem reduces to the following: Given C and ϵ, to find a saddle point of the quadratic problem

$$L(\mathbf{w}, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}, \boldsymbol{\alpha}, \tilde{\boldsymbol{\alpha}}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}(\xi_i + \tilde{\xi}_i) - \sum_{i=1}^{N}\left(\tilde{\gamma}_i\tilde{\xi}_i + \gamma_i\xi_i\right)$$
$$- \sum_{i=1}^{N}\alpha_i[1 - y_i(\mathbf{w}^T\mathbf{x}_i) + \epsilon + \xi_i] - \sum_{i=1}^{N}\tilde{\alpha}_i[y_i(\mathbf{w}^T\mathbf{x}_i) - 1 + \epsilon + \tilde{\xi}_i]$$

minimum with respect to $\mathbf{w}$, ξ_i and $\tilde{\xi}_i$, and maximum with respect to Lagrange multipliers $\alpha_i \ge 0$, $\tilde{\alpha}_i \ge 0$, $\gamma_i \ge 0$, and $\tilde{\gamma}_i \ge 0$, for all $i = 1, \ldots, N$.

The solution for the linear equalizer can be expanded in terms of the outputs y_i, the input patterns $\mathbf{x}_i$, and the Lagrange multipliers α_i and $\tilde{\alpha}_i$ as

$$\mathbf{w}_{QP} = \sum_{i=1}^{N}(\tilde{\alpha}_i - \alpha_i)y_i\mathbf{x}_i. \tag{5}$$

The Lagrange multipliers are obtained by maximizing the following quadratic form:

$$W(\boldsymbol{\alpha}, \tilde{\boldsymbol{\alpha}}) = -\epsilon\sum_{i=1}^{N}(\tilde{\alpha}_i + \alpha_i) + \sum_{i=1}^{N}(\tilde{\alpha}_i - \alpha_i)$$
$$- \frac{1}{2}\sum_{i,j=1}^{N}(\tilde{\alpha}_i - \alpha_i)(\tilde{\alpha}_j - \alpha_j)(y_iy_j)\langle\mathbf{x}_i, \mathbf{x}_j\rangle \tag{6}$$

subject to $0 \leq \alpha_i, \tilde{\alpha}_i \leq C$ and where $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$ denotes the inner product between the input patterns.

According to the Karush–Kuhn–Tucker (KKT) conditions, the input patterns that appear in the expansion (5) are points where exactly one of the Lagrange multipliers is greater than zero: These input patterns are called support vectors.

The difference with the conventional SVR formulation is that the linear kernel $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$ in the quadratic form (6) is now weighted by the factor $y_i y_j$ and that the solution is expanded in terms of a set of weighted Lagrange multipliers

$$\mathbf{w}_{\mathrm{QP}} = \sum_{i=1}^{N} \beta_i \mathbf{x}_i \tag{7}$$

where we have defined the weighted Lagrange multipliers as

$$\beta_i = (\tilde{\alpha}_i - \alpha_i) y_i. \tag{8}$$

With this definition, the output of the equalizer can be written as

$$y_k = \sum_{i=1}^{N} \beta_i \langle \mathbf{x}_i, \mathbf{x}_k \rangle. \tag{9}$$

Note that the optimal equalizer (7) depends on its output through the weighted Lagrange multipliers (8); therefore, the solution cannot be obtained in a single step. An iterative procedure to solve this problem is described in Section IV.

Note also that in the conventional SVR framework, the linear regressor includes a bias term $y_k = \mathbf{w}^T \mathbf{x}_k + b$. However, to apply this technique to blind equalization of CM signals, it is necessary to force $b = 0$; otherwise, the QP problem will always give the trivial solution $\mathbf{w} = 0$ and $b = 1$.

B. Complex Channel and QPSK Input

The extension of the above procedure to complex modulations requires some further elaboration. In this case, taking into account that all $\mathbf{w}$, $\mathbf{x}_i$, and y_i are complex variables, we have

$$\|\mathbf{w}^T \mathbf{x}_i\|^2 = \mathrm{Re}(y_i)^2 + \mathrm{Im}(y_i)^2$$

which can be rewritten as

$$\|\mathbf{w}^T \mathbf{x}_i\|^2 = \mathrm{Re}(y_i) \begin{bmatrix} \mathrm{Re}(\mathbf{w}^T) & \mathrm{Im}(\mathbf{w}^T) \end{bmatrix} \begin{bmatrix} \mathrm{Re}(\mathbf{x}_i) \\ -\mathrm{Im}(\mathbf{x}_i) \end{bmatrix}$$
$$+ \mathrm{Im}(y_i) \begin{bmatrix} \mathrm{Re}(\mathbf{w}^T) & \mathrm{Im}(\mathbf{w}^T) \end{bmatrix} \begin{bmatrix} \mathrm{Im}(\mathbf{x}_i) \\ \mathrm{Re}(\mathbf{x}_i) \end{bmatrix}$$

or, in a more compact formulation, as

$$\|\mathbf{w}^T \mathbf{x}_i\|^2 = \tilde{\mathbf{w}}^T \tilde{\mathbf{g}}_i \tag{10}$$

where we have defined the following vectors:

$$\tilde{\mathbf{w}} = \begin{bmatrix} \mathrm{Re}(\mathbf{w}) \\ \mathrm{Im}(\mathbf{w}) \end{bmatrix}$$

and

$$\tilde{\mathbf{g}}_i = \mathrm{Re}(y_i) \begin{bmatrix} \mathrm{Re}(\mathbf{x}_i) \\ -\mathrm{Im}(\mathbf{x}_i) \end{bmatrix} + \mathrm{Im}(y_i) \begin{bmatrix} \mathrm{Im}(\mathbf{x}_i) \\ \mathrm{Re}(\mathbf{x}_i) \end{bmatrix} = \begin{bmatrix} \mathrm{Re}(y_i \mathbf{x}_i^*) \\ \mathrm{Im}(y_i \mathbf{x}_i^*) \end{bmatrix} \tag{11}$$

where the superscript $(\cdot)^*$ denotes complex conjugate. Let us remark that $\tilde{\mathbf{w}}$ and $\tilde{\mathbf{g}}_i$ are $2M \times 1$ vectors with real components (M being the length of the equalizer).

By using (10), the following quadratic inequalities involving complex variables, which impose the CM property

$$\|\mathbf{w}^T \mathbf{x}_i\|^2 - 1 \leq \epsilon + \xi_i$$
$$1 - \|\mathbf{w}^T \mathbf{x}_i\|^2 \leq \epsilon + \tilde{\xi}_i$$

can be rewritten as the following linear inequalities with real variables:

$$\tilde{\mathbf{w}}^T \tilde{\mathbf{g}}_i - 1 \leq \epsilon + \xi_i \tag{12}$$
$$1 - \tilde{\mathbf{w}}^T \tilde{\mathbf{g}}_i \leq \epsilon + \tilde{\xi}_i. \tag{13}$$

Therefore, to blindly restore the CM property for complex modulations, we minimize

$$L(\tilde{\mathbf{w}}, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}) = \frac{1}{2} \|\tilde{\mathbf{w}}\|^2 + C \sum_{i=1}^{N} (\xi_i + \tilde{\xi}_i)$$

subject to (12) and (13) for all $i = 1, \ldots, N$. This problem is equivalent to maximizing the following quadratic form:

$$W(\boldsymbol{\alpha}, \tilde{\boldsymbol{\alpha}}) = -\epsilon \sum_{i=1}^{N} (\tilde{\alpha}_i + \alpha_i) + \sum_{i=1}^{N} (\tilde{\alpha}_i - \alpha_i)$$
$$- \frac{1}{2} \sum_{i,j=1}^{N} (\tilde{\alpha}_i - \alpha_i)(\tilde{\alpha}_j - \alpha_j) \langle \tilde{\mathbf{g}}_i, \tilde{\mathbf{g}}_j \rangle \tag{14}$$

subject to $0 \leq \alpha_i, \tilde{\alpha}_i \leq C$, and where the input patterns $\tilde{\mathbf{g}}_i$ are now given by (11). In this way, we arrive again at a conventional QP problem with real variables, whose solution is given by

$$\tilde{\mathbf{w}} = \sum_{i=1}^{N} (\tilde{\alpha}_i - \alpha_i) \tilde{\mathbf{g}}_i = \sum_{i=1}^{N} \begin{bmatrix} \mathrm{Re}\left((\tilde{\alpha}_i - \alpha_i) y_i \mathbf{x}_i^*\right) \\ \mathrm{Im}\left((\tilde{\alpha}_i - \alpha_i) y_i \mathbf{x}_i^*\right) \end{bmatrix}.$$

Similarly to the real channel case, if we define the weighted Lagrange multipliers as $\beta_i = (\tilde{\alpha}_i - \alpha_i) y_i$ (note that now these multipliers are complex), the complex equalizer can be expressed as

$$\mathbf{w}_{\mathrm{QP}} = \sum_{i=1}^{N} \beta_i \mathbf{x}_i^* \tag{15}$$

and the output of the equalizer is finally given by

$$y_k = \sum_{i=1}^{N} \beta_i \langle \mathbf{x}_i^*, \mathbf{x}_k \rangle. \tag{16}$$

Therefore, complex and real modulations admit a common SVR formulation: Compare (7) and (9) for the linear case to (15) and (16) for the complex case, respectively.

C. Nonlinear Equalization

In this section, we consider the extension of the above procedure to nonlinear equalization. In particular and due to the lack of space, in this paper, we focus only on the case of real modulations since the nonlinear complex case requires further elab-

oration. To obtain a nonlinear equalizer, the linear regression is performed in another feature space

$$y_i = \mathbf{w}^T \Phi(\mathbf{x}_i)$$

where $\Phi(\mathbf{x}_i)$ is a nonlinear mapping that transforms the input vector into a higher dimension vector (probably infinity). The problem now reads as follows: To minimize

$$L(\mathbf{w}, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{N} (\xi_i + \tilde{\xi}_i)$$

subject to

$$y_i \mathbf{w}^T \Phi(\mathbf{x}_i) - 1 \leq \epsilon + \xi_i$$
$$1 - y_i \mathbf{w}^T \Phi(\mathbf{x}_i) \leq \epsilon + \tilde{\xi}_i$$
$$\xi_i, \tilde{\xi}_i \geq 0$$

for all $i = 1, \ldots, N$. Similarly to the linear case, we now get the following QP problem:

$$W(\boldsymbol{\alpha}, \tilde{\boldsymbol{\alpha}}) = -\epsilon \sum_{i=1}^{N} (\tilde{\alpha}_i + \alpha_i) + \sum_{i=1}^{N} (\tilde{\alpha}_i - \alpha_i)$$
$$- \frac{1}{2} \sum_{i,j=1}^{N} (\tilde{\alpha}_i - \alpha_i)(\tilde{\alpha}_j - \alpha_j)(y_i y_j) \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$$

subject to $0 \leq \alpha_i, \tilde{\alpha}_i \leq C$.

Fortunately, the transformation itself does not need to be known to solve the QP problem since only the inner product between transformed input vectors, which is given by a nonlinear kernel satisfying the Mercer condition [7], must be evaluated: $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$. Typical examples are the radial basis function kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{(2\sigma^2)}\right) \qquad (17)$$

or the polynomial kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\langle \mathbf{x}_i, \mathbf{x}_j \rangle + \tau)^p \qquad (18)$$

where p is the polynomial order, and τ is a positive real constant that plays a similar role as σ for the RBF kernel [16]. To summarize, and from a practical point of view, the extension of our procedure to nonlinear blind equalization reduces to replacing, in (6), the linear dot product by a nonlinear kernel.

The solution for the optimal lineal regressor in the feature space is expanded again in terms of the weighted Lagrange multipliers and the transformed input vectors

$$\mathbf{w}_{\text{QP}} = \sum_{i=1}^{N} \beta_i \Phi(\mathbf{x}_i) \qquad (19)$$

whereas the output of the equalizer is given by

$$y_k = \mathbf{w}^T \Phi(\mathbf{x}_k) = \sum_{i=1}^{N} \beta_i K(\mathbf{x}_i, \mathbf{x}_k). \qquad (20)$$

Traditionally, the problem of blind nonlinear equalization has been tackled by considering nonlinear structures such as multi-layer perceptrons [17], recurrent neural networks [18], or piecewise linear networks [19] among others, trained to minimize some cost function (Godard cost functions, kurtosis, minimum entropy, etc.). In all these approaches, the structure and complexity of the nonlinear equalizer must be specified in advance. The proposed nonlinear blind SVM, however, obtains simultaneously the topology of the nonlinear equalizer, as well as its parameters: We feel that this is one of the main advantages of the proposed procedure. In addition, another important feature of the proposed blind SVM approach is that it establishes a general framework that is suitable for both linear and nonlinear blind equalizers.

IV. ITERATIVE REWEIGHTED QP

The optimal blind regressor cannot be found in a single step because the weighted Lagrange multipliers depend on the solution i.e., $\beta_i = \beta_i(\mathbf{w})$ [see (8)]. Therefore, we need to apply an iterative procedure, which, due to its similarity with the iterative reweighted least squares (IRWLS) technique used in some approximation and regression problems [20], is called iterative reweighted quadratic programming (IRWQP). The procedure, which is particularized here for real modulations, is as follows.

```
1) Solve the QP problem (6) considering y_i
fixed.
2) Obtain the new equalizer as (7), and
compute the corresponding new output y_i.
3) Repeat until convergence.
```

In order to complete the algorithm, it is necessary to smooth the equalizer coefficients from iteration to iteration. The reason for this smoothing is the following: Suppose that the initial output of the equalizer is y_i. Then, in the first step, we perform a linear (or nonlinear) regression, trying to fit $1/y_i$ as desired output, forcing, in this way, a constant modulus signal. If we apply a new iteration of the IRWQP procedure, the new weights applied to the Lagrange multipliers will be close to $1/y_i$, whereas the new desired output will again be close to y_i. Then, to avoid a limit cycle oscillation between these two outputs, some type of smoothing must be introduced.

Specifically, we choose to smooth the coefficients of the linear regressor, which is obtained in the input space for a linear equalizer or in the feature space if a nonlinear equalizer is sought, i.e.,

$$\mathbf{w}_k = \lambda \mathbf{w}_{k-1} + (1 - \lambda)\mathbf{w}_{\text{QP}} \qquad (21)$$

where λ is a constant parameter close to one, and $\mathbf{w}_{\text{QP}}$ is the linear regressor, which is given either by (7) or (15) for a linear equalizer (for real or complex modulations, respectively), or by (19) for a nonlinear equalizer. Taking into account that the input patterns do not change from iteration to iteration of the IRWQP procedure, it is easy to realize that the same result can be obtained by smoothing the weighted Lagrange multipliers instead of the regressor, that is

$$\overline{\beta}_k = \lambda \overline{\beta}_{k-1} + (1 - \lambda)\overline{\beta}_{\text{QP}} \qquad (22)$$

where $\overline{\beta}_{QP}$ is a column vector containing the N weighted Lagrange multipliers obtained by solving the QP problem at the kth iteration, and $\overline{\beta}_k$ is a column vector containing the smoothed Lagrange multipliers. The only difference between (21) and (22) is that the former procedure cannot be directly applied for a nonlinear equalizer because in this case, the unknown nonlinear mapping appears in the expansion of the optimal regressor. For this reason and in order to get a common algorithm for linear and nonlinear equalization, we decided to use (22).

The initial $\overline{\beta}_0$ can be obtained by solving the following linear problem:

$$\mathbf{K}\overline{\beta}_0 = \mathbf{y}_d \tag{23}$$

where $\mathbf{K}$ is an $N \times N$ kernel matrix with elements given either by $K_{ij} = \langle \mathbf{x}_i, \mathbf{x}_j \rangle$ or $K_{ij} = \langle \mathbf{x}_i^*, \mathbf{x}_j \rangle$ for linear equalizers, or by $K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$ for nonlinear equalizers; in addition, $\mathbf{y}_d = (x_{1-d}, \ldots, x_{N-d})^T$. In this way, the initial equalizer delays the input d samples. This delay is typically chosen at the center of the equalizer coefficient vector.

To check the convergence of the iterative algorithm, we compute the average modulus error (AME), which is defined as the mean deviation of the squared modulus over the desired CM signal, i.e.,

$$\text{AME}(k) = \frac{1}{N} \sum_{i=1}^{N} \left(|y_{i,k}|^2 - 1 \right) \tag{24}$$

where $y_{i,k}$ denotes the output of the equalizer for the ith input pattern at the kth iteration. The difference between the AME after two consecutive iterations of the IRWQP algorithm was used to stop the algorithm, in particular, the convergence criterion is

$$\text{AME}(k-1) - \text{AME}(k) \leq 10^{-3}. \tag{25}$$

Although we have not yet been able to theoretically prove the convergence and stability of the proposed procedure, in all the examples, the algorithm always converged to a solution for which the AME was a minimum (assuming a smoothing parameter close to one). Nevertheless, a minimum of the AME does not necessarily mean a minimum of the ISI: This point will be clarified in the simulations section.

Finally, the proposed can be summarized as follows.

Algorithm 1: Summary of the SVM-based blind equalization algorithm
Initialize C, ϵ, λ, $\overline{\beta}_0$, and $k = 0$.
while Convergence criterion not true **do**
 Obtain the output of the equalizer, y_i, using $\overline{\beta}_k$.
 Solve QP problem (6), and obtain $\overline{\beta}_{QP}$.
 $\overline{\beta}_{k+1} = \lambda\overline{\beta}_k + (1 - \lambda)\overline{\beta}_{QP}$.
 Calculate new value for the AME, and check convergence.
end while

V. SELECTION OF SVR PARAMETERS

The SVR parameters ϵ (for the Vapnik's loss function) and C (regularization parameter), in addition to the smoothing parameter λ, must be selected by the user. The latter can be fixed in advance to a constant value close to unity ($\lambda = 0.9$ has been used in all the examples described in the paper), whereas the selection of C and ϵ requires more understanding of the SVR technique.

Parameter ϵ determines the width of the ϵ-insensitive zone, thus controlling the value of the number of SVs and, consequently, the generalization performance of the SVR [7]: A large value of ϵ yields a SVR with few SVs. Most of the techniques proposed to estimate this parameter suggest that ϵ should be selected proportionally to the noise variance [21], [22], which is assumed to be estimated from data. From a different point of view, Schölkopf *et al.* [23] proposed to fix the percentage of points outside the ϵ-tube instead of estimating ϵ. Here, we use the technique proposed by Cherkassky in [21] that estimates ϵ as

$$\epsilon = \tau \sqrt{\sigma_n^2 \frac{\ln N}{N}}$$

where σ_n^2 is the noise variance, N is the number of samples in the input data, and τ is an empirically selected constant that we have chosen as $\tau = 3$.

On the other hand, the regularization parameter C determines a tradeoff between model complexity (minimizing $\|\mathbf{w}\|^2$) and the penalty given to deviations larger than ϵ. This value is sometimes chosen as the range of the training output values [24]. Here, we use a modification of this idea that has been shown to be more robust to outliers [21]. Specifically, we use

$$C = \overline{g_k} + 3\sigma_g$$

where $g_k = |x_k|^2$, x_k is the channel output, and $\overline{g_k}$ denotes mean value.

If a nonlinear equalizer is sought, in addition to C and ϵ, its is necessary to select the kernel and the parameters defining the kernel (σ for an RBF kernel or τ for a polynomial kernel). The general question of how to select the ideal kernel for a given task is still an open problem. Without any *a priori* knowledge of the type of nonlinearity that distorts the CM input signal, there is no reason to choose a particular kernel. Once the kernel has been chosen, its parameter must be tuned to the particular data set under consideration: This can be done using a validation set.

VI. SIMULATION RESULTS

In this section, we compare the performance of the proposed blind (linear or nonlinear) SVM and the batch super-exponential algorithm proposed by Shalvi and Weinstein (denoted as the SW algorithm) [3], [4], which is based on fourth-order cumulants. The QP problem at each step of the IRWQP procedure has been solved using the Matlab SVM toolbox available in [14]. The Matlab programs that implement the SVM-based blind equalization algorithms can be obtained at http://gtas.dicom.unican.es/comp/nacho.html.
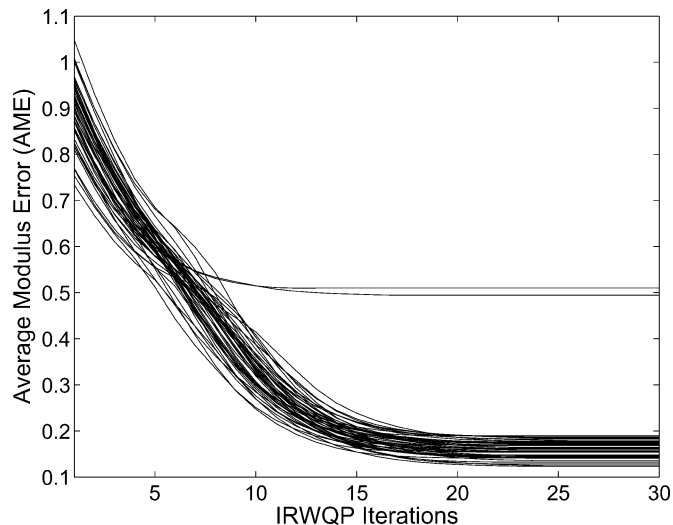
Fig. 1.   Average modulus error (AME) versus the number of iterations using a data block size of $N = 100$ and an SNR $= 30$ dB. Real channel $H_1(z)$ and binary input.

Fig. 2.   ISI versus the number of iterations using a data block size of $N = 100$ and a SNR $= 30$ dB. Real channel $H_1(z)$ and binary input.

A. Linear Equalization

In the first example, we consider a linear blind SVM and a real channel. A binary signal is sent through the channel $H_1(z) = (0.4 + z^{-1} - 0.7z^{-2} + 0.6z^{-3} + 0.3z^{-4} - 0.4z^{-5} + 0.1z^{-6})$ (used in [3]), and at the channel output, white Gaussian noise is added. We have used an equalizer of length $M = 17$, which was initialized as $\mathbf{w} = \delta[n - 8]$. As a measure of equalization performance, we use the ISI defined as

$$\text{ISI} = 10 \log_{10} \frac{\sum_n |\theta_n|^2 - \max_n |\theta_n|^2}{\max_n |\theta_n|^2}$$

where $\theta = \mathbf{h} * \mathbf{w}$ is the combined channel-equalizer impulse response, which is a delta function for a zero-forcing equalizer. The initial ISI for the selected channel and for an equalizer initialized with a centered spike is of 1.03 dB.

Similarly to other cumulant-based algorithms, the SW algorithm provides poor results with very short blocks of data [3]. This could happen, for instance, in burst time division multiple access (TDMA) transmissions (without any training sequence) over fast fading channels. It is in this situation when the proposed blind SVM technique is expected to offer some advantages. To corroborate this point, we have tested the SW and blind SVM algorithms using blocks of input data ranging from $N = 25$ to 500 samples in a low noise situation SNR $= 30$ dB and in a high noise situation SNR $= 10$ dB. For the blind SVM, we used Vapnik's ϵ-insensitive loss function (the results were similar for the quadratic loss function) with C, ϵ, and λ parameters selected as described in Section V.

First, to illustrate the convergence of the algorithm, Fig. 1 shows the evolution of the average modulus error (AME) for an example that uses an input block size of $N = 100$ samples in a low noise situation: SNR $= 30$ dB. The results of 50 independent simulations are depicted in Fig. 1: The algorithm always converges in the direction of the decreasing AME until a minimum is found. Most of the trials converge to a final AME lower than 0.2; however, we can see that two of them yield a final AME
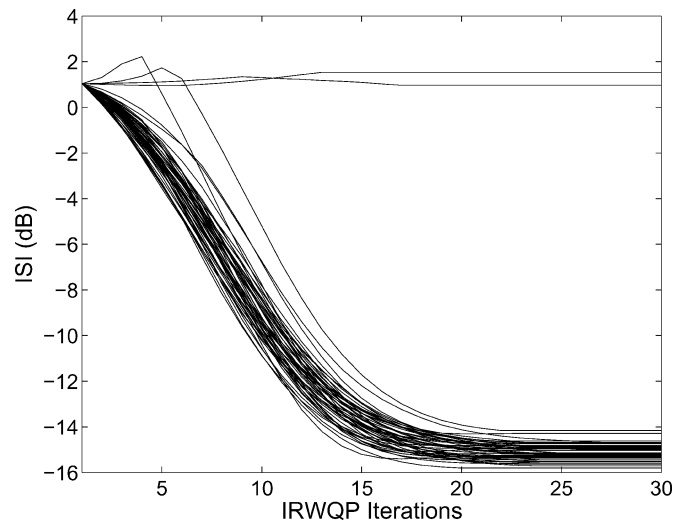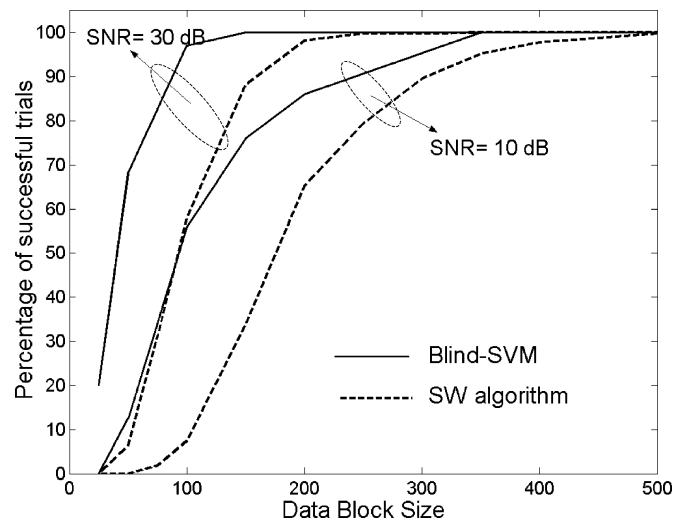

Fig. 3.   Percentage of convergence (final ISI below -5 dB) versus data block size for the proposed blind-SVM and the Shalvi–Weinstein algorithm. Real channel $H_1(z)$ and binary input.

around 0.5. Fig. 2 shows the evolution of the ISI for the same scenario: The two trials that gave a final AME around 0.5 correspond now to the curves that fail to decrease the ISI. As we already pointed out, a decrease of the AME does not necessarily imply a decrease of the ISI. Moreover, this result suggests the possibility of using the AME not only for testing the convergence of the algorithm but also for knowing in advance if the channel has been successfully equalized.

Second, we have compared the performance of our SVM-based blind equalizer and the SW algorithm. For both methods, if the final ISI after a trial was below -5 dB, we considered that the channel was successfully equalized, since with this level of ISI, it is already possible to switch to a decision-directed mode. For each data block size and noise level, both algorithms were tested in 200 Monte Carlo trials. The results are depicted in Fig. 3, which shows the percentage of trials in which each algorithm successfully converged, and in Fig. 4, which shows the mean ISI level after convergence for the successful trials.
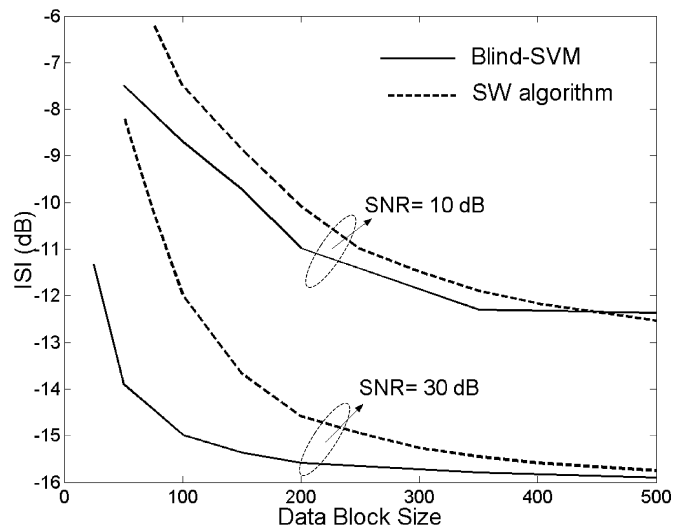
Fig. 4. Mean ISI level after convergence for the successful trials versus data block size for the proposed blind-SVM and the Shalvi–Weinstein algorithm. Real channel $H_1(z)$ and binary input.

Fig. 5. Percentage of convergence (final ISI below -5 dB) versus data block size for the proposed blind-SVM and the Shalvi–Weinstein algorithm. Complex channel $H_2(z)$ and QPSK input.

The blind SVM obtains better results than the SW method for both the percentage of convergence and the final level of ISI. The improvement is more important when the input data block size is very short (less than 150 samples). The price to be paid is obviously an increase in the required computational burden, in comparison with the SW algorithm.

In the second example, we test the proposed algorithm with a QPSK input signal $s_k = \{\pm 1 \pm j\}/\sqrt{(2)}$ and the following allpass channel with phase error:

$$H_2(z) = \frac{0.7 - z^{-1}}{1 - 0.7z^{-1}} e^{j\pi/4}.$$

We used again an equalizer of length $M = 17$ initialized with a centered spike so that the initial value of the ISI is of 0.174 dB for this example. The probability of convergence and final ISI results for different data block sizes and noise levels are shown in Figs. 5 and 6, respectively. Again, the blind SVM provides better results than the SW algorithm mainly for short data registers, but for this example, the improvement in the final level of ISI over the SW algorithm is more remarkable than in the real channel case.

In the third example, we consider the real channel $H_3(z) = 0.407 + 0.815z^{-1} + 0.407z^{-2}$ that has been taken from [25] and a binary input signal. We used an input data block size of $N = 150$ samples and an equalizer of length $M = 12$. After convergence, the bit error rate (BER) was evaluated for both the blind SVM and the SW algorithm by counting errors after transmitting 10^4 or 10^5 symbols, depending on the SNR. We run 50 independent simulations. Fig. 7 shows the BER curves for this example.

B. Nonlinear Equalization

In our last example, we test the nonlinear blind SVM. We consider a nonlinear channel composed of a linear channel followed by a memoryless nonlinearity. Such a nonlinear channel can be encountered in digital satellite communications [18] and in digital magnetic recording [26]. The linear channel considered is $H_4(z) = 1 - 0.5z^{-1}$, and the nonlinear function applied
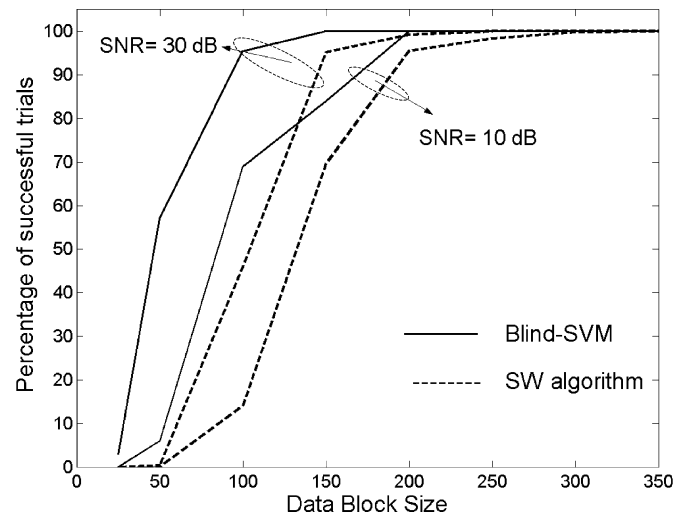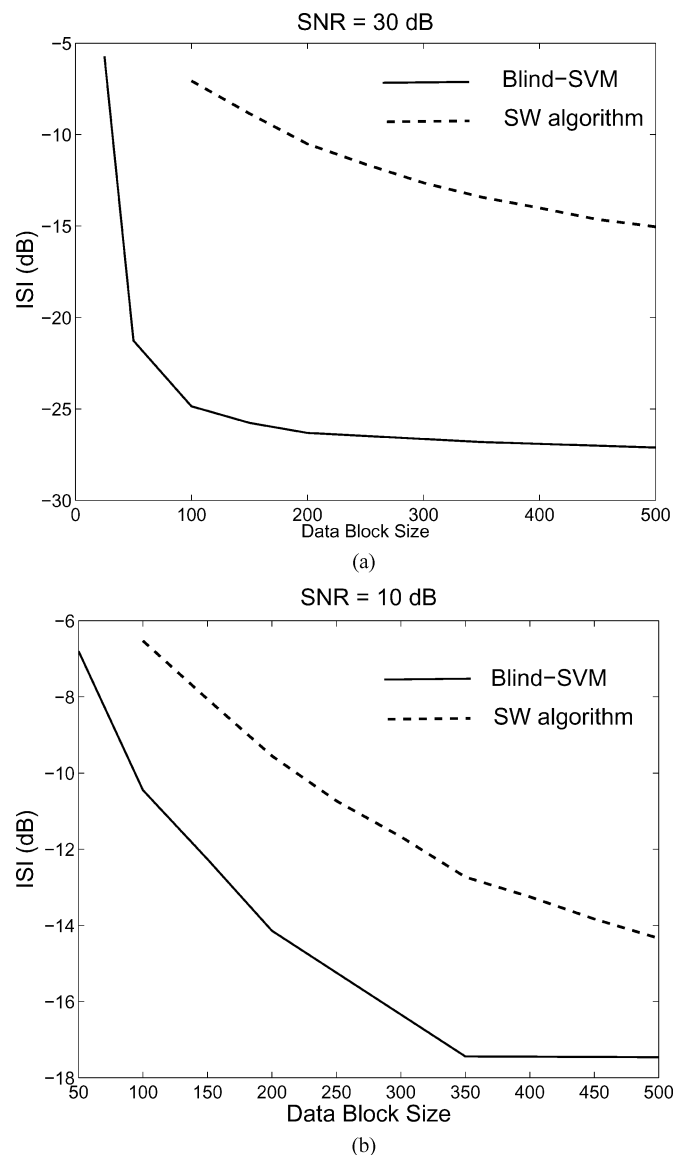

(a)

(b)

Fig. 6. Mean ISI level after convergence for the successful trials versus data block size for the proposed blind-SVM and the Shalvi–Weinstein algorithm. Complex channel $H_2(z)$ and QPSK input. (a) SNR $= 30$ dB. (b) SNR $= 10$ dB.
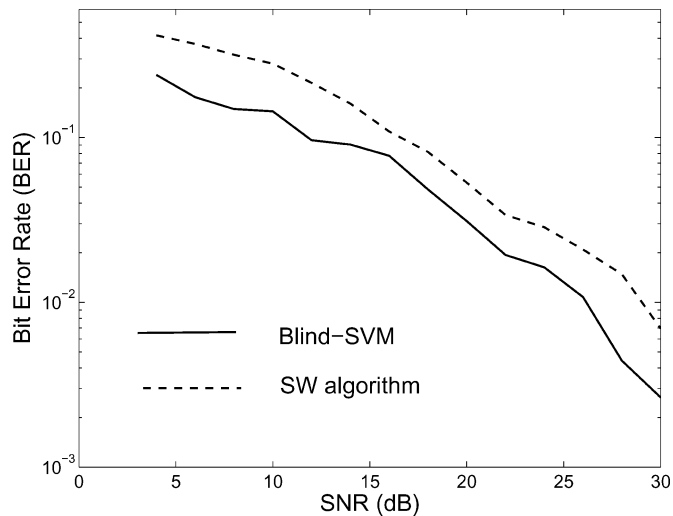
Fig. 7. BER comparison for the blind SVM (solid) and the SW algorithm (dashed) using a data block size of $N = 150$. Real channel $H_3(z)$ and binary input.
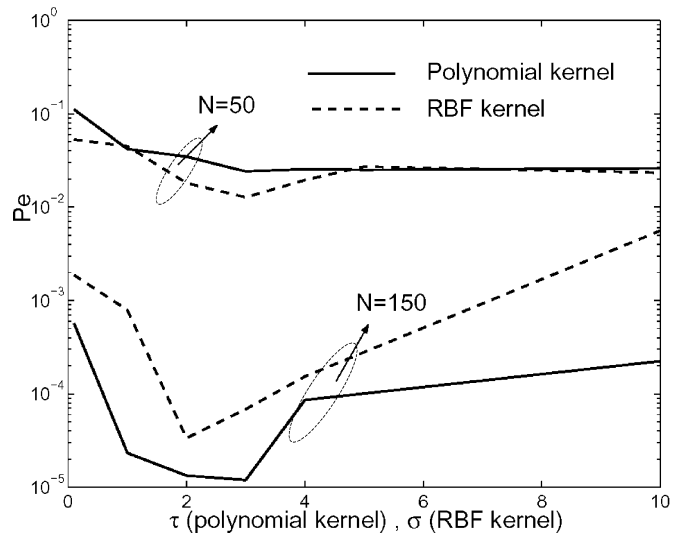

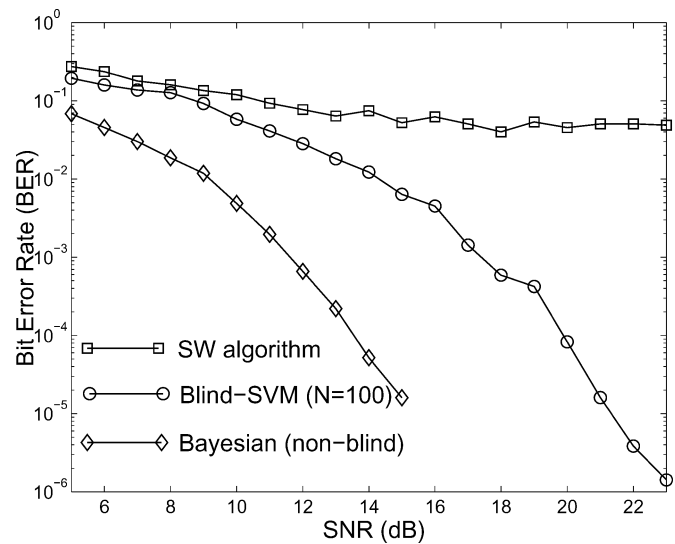Fig. 8. BER for polynomial and RBF kernels.

Fig. 9. BER comparison for the blind nonlinear SVM, the optimum nonblind Bayesian equalizer, and a linear equalizer trained with the SW algorithm. The data block size is $N = 100$ samples. Real channel $H_4(z)$ followed by a memoryless nonlinearity and binary input.
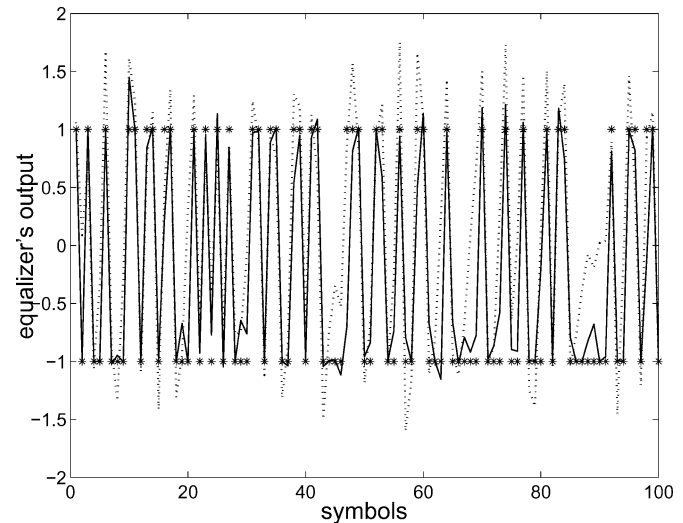

Fig. 10. Output of a nonlinear blind SVM equalizer (solid line), a linear equalizer trained with the SW algorithm (dotted line), and the true bits ('*'). The data block size is $N = 100$ samples, and the SNR is 10 dB.

is $z = x + 0.2x^2 - 0.9x^3$, where x is the linear channel output. Finally, white Gaussian noise was added.

We have compared the performance of the proposed procedure for polynomial and RBF kernels. For the polynomial kernel [see (18)], we choose a fixed degree $p = 3$, and then, we study the influence of the constant term τ, whereas for the RBF, the kernel variance was changed. We have used input data blocks of $N = 50$ and $N = 150$ samples. Fig. 8 shows the BER (estimated using a different validation set) for both kernels as a function of τ and σ. The curves show the averaged results of 25 independent simulations. For $N = 50$, the results are very similar for both kernels; however, for $N = 150$, the polynomial kernel performs slightly better than the RBF kernel. We attribute this improvement to the fact that the linear channel output is distorted, for this particular example, through a memoryless polynomial nonlinearity. It seems that a polynomial nonlinear distortion can be compensated more effectively by a polynomial kernel. This result, however, cannot be considered conclusive: The selection of the proper kernel is a topic that deserves further study.

According to these results, from now on, we use a polynomial kernel with $p = 3$ and $\tau = 3$. The rest of the SVR parameters are selected again according to the procedure described in Section V. The data block size is now $N = 100$. Fig. 9 shows the BER for the proposed nonlinear blind SVM equalizer, a linear equalizer trained with the SW algorithm, and the optimum nonlinear Bayesian equalizer [27]. The Bayesian equalizer assumes that the channel is known and is the minimum BER solution in a nonblind situation. On the other hand, as expected, the linear equalizer is not able to remove the nonlinear ISI introduced by the memoryless nonlinearity. Finally, the blind nonlinear SVM is able to restore the binary signal. The final number of support vectors for the nonlinear blind SVM ranges between 70 and 90, depending on the SNR. This is an important point since the computational cost depends directly on the number of SVs (note that this is not the case for a linear blind SVM). Finally, Fig. 10 shows an example of the solution obtained by the non-

linear SVM (in this example, the SNR was 10 dB, and 89 patterns became SVs) depicted by a solid line and the true bits depicted with asterisks. We have also included for comparison the solution provided by the SW algorithm by a dotted line. We can see that the blind SVM is able to restore the CM property of the input signal.

VII. CONCLUSIONS

In this paper, blind equalization of CM signals has been formulated as a regression problem, and the powerful SVM technique has been applied to solve it. An iterative reweighted quadratic programming (IRWQP) procedure has been proposed to find the optimal regressor. It is shown in the paper that blind SVM equalization has several attractive features: The quadratic programming problem solved at each iteration is convex and has a globally optimal solution; it simultaneously exploits all the information in the given block of data, thus requiring fewer data samples than other standard blind algorithms; finally, linear and nonlinear equalizers can be treated in a unified manner and within a powerful machine-learning framework.

It has been shown that the proposed technique offers better performance than cumulant-based methods, mainly for very short data registers. Its main drawback is the high computational cost of the IRWQP technique; nevertheless, some techniques to reduce the computational burden of QP problems that have recently appeared [28], [29] could be used to alleviate this problem and are currently under study. Another further line of research is the generalization of the proposed procedure to multilevel modulations.

REFERENCES

[1] Z. Ding and Y. Li, *Blind Equalization and Identification*. New York: Marcel Dekker, 2001.
[2] C. R. Johnson, Jr., P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas, "Blind equalization using the constant modulus criterion: A review," *Proc. IEEE*, vol. 86, pp. 1927–1950, Oct. 1998.
[3] O. Shalvi and E. Weinstein, "Universal methods for blind deconvolution," in *Blind Deconvolution*, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[4] ——, "Super-exponential methods for blind deconvolution," *IEEE Trans. Inform. Theory*, vol. 39, pp. 504–519, Mar. 1993.
[5] J. K. Tugnait, "Blind estimation and equalization of digital communication FIR channels using cumulant matching," *IEEE Trans. Commun.*, vol. 43, pp. 1240–1245, Feb./Mar./Apr. 1995.
[6] D. Hatzinakos and C. L. Nikias, "Blind equalization based on higher-order statistics (HOS)," in *Blind Deconvolution*, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[7] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
[8] S. Chen, S. Gunn, and C. J. Harris, "Decision feedback equaliser design using support vector machines," *Proc. Inst. Elect. Eng. Vis. Image Signal Process.*, vol. 147, no. 3, pp. 213–219, June 2000.
[9] D. J. Sebald and J. A. Bucklew, "Support vector machine techniques for nonlinear equalization," *IEEE Trans. Signal Processing*, vol. 48, pp. 3217–3226, Nov. 2000.
[10] F. Pérez-Cruz, A. Navia-Vázquez, P. L. Alarcón-Diana, and A. Artés-Rodríguez, "SVC-based equalizer for burst TDMA transmissions," *Signal Process.*, vol. 81, pp. 1681–1693, 2001.
[11] I. Santamaría, R. González, C. Pantaleón, and J. C. Príncipe, "Maximum margin equalizers trained with the Adatron algorithm," *Signal Process.*, vol. 83, pp. 593–602, 2003.
[12] Q. Li, E.-W. Bai, and Y. Ye, "Blind channel equalization and ϵ-approximation algorithms," *IEEE Trans. Signal Processing*, vol. 49, pp. 2823–2831, Nov. 2001.

[13] B. Maricic, Z.-Q. Luo, and T. N. Davidson, "Blind equalization of constant modulus signals via restricted convex optimization," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Salt Lake City, UT, May 2001.
[14] S. R. Gunn. (1998) MATLAB Support Vector Machine Toolbox. Image Speech and Intelligent Systems Research Group, Univ. Southampton, Southampton, U.K.. [Online]. Available: http://www.isis.ecs.soton.ac.uk/isystems/kernel/
[15] I. Santamaría, J. Ibáñez, L. Vielva, and C. Pantaleón, "Blind equalization of constant modulus signals via support vector regression," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, vol. II, Hong Kong, Apr. 2003, pp. 737–740.
[16] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*, Singapore: World Scientific, 2002.
[17] D. Erdogmus, D. Rende, J. C. Principe, and T. F. Wong, "Nonlinear channel equalization using multilayer perceptrons with information-theoretic criterion," in *Proc. IEEE Workshop Neural Networks Signal Processing XI*, Falmouth, MA, Sept. 2001, pp. 443–451.
[18] G. Kechriotis, E. Zervas, and E. S. Manolakos, "Using recurrent neural network for adaptive communication channel equalization," *IEEE Trans. Neural Networks*, vol. 5, pp. 267–278, Mar. 1994.
[19] X. Liu and T. Adaly, "Canonical piecewise linear network for nonlinear filtering and its application to blind equalization," *Signal Process.*, vol. 61, no. 2, pp. 145–155, Sept. 1997.
[20] P. W. Holland and R. E. Welch, "Robust regression using iterative reweighted least squares," *Commun. Statist. Theory Methods*, vol. A 6, no. 9, pp. 813–827, 1997.
[21] V. Cherkassky and Y. Ma, "Selection of meta-parameters for support vector regression," in *Proc. ICANN*, J. R. Dorronsoro, Ed., Berlin, Germany, 2002, pp. 687–693.
[22] J. T. Kwok, "Linear dependency between ϵ and the input noise in ϵ-support vector regression," in *Proc. ICANN*, G. Dorffner, H. Bischof, and K. Hornik, Eds., Berlin, Germany, 2001, pp. 405–410.
[23] B. Schölkopf, P. Bartlett, A. Smola, and R. Williamson, "Support vector regression with automatic accuracy control," in *Proc. ICANN*, L. Niklasson, M. Bodén, and T. Ziemke, Eds., Berlin, Germany, 1998, pp. 111–116.
[24] D. Mattera and S. Haykin, "Support vector regression for dynamic reconstruction of a chaotic system," 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. 211–241.
[25] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 2000.
[26] N. P. Sands and J. M. Cioffi, "Nonlinear channel models for digital magnetic recording," *IEEE Trans. Magn.*, vol. 29, pp. 3996–3998, Nov. 1993.
[27] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," *IEEE Trans. Neural Networks*, vol. 4, pp. 570–578, July 1993.
[28] 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.
[29] 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.

Ignacio Santamaría (M'96) was born in Vitoria, Spain, in 1967. He received the Telecommunication Engineer and the Ph.D. degree in electrical engineering from the Polytechnic University of Madrid, Madrid, Spain, in 1991 and 1995, respectively.

In 1992, he joined the Departamento de Ingeniería de Comunicaciones, Universidad de Cantabria, Santander, Spain, where he is currently an Associate Professor. In 2000, he spent a visiting period at the Computational NeuroEngineering Laboratory (CNEL), University of Florida, Gainesville. He has more than 60 publications in refereed journals and international conference papers. His current research interests include nonlinear modeling techniques, adaptive systems, and machine learning theories and their application to digital communication systems.
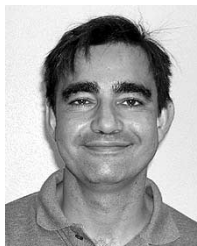
Carlos Pantaleón (M'95) was born in Badajoz, Spain, in 1966. He received the Telecommunication Engineer and the Doctor degrees from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 1990 and 1994, respectively.

In 1990, he joined the Departamento de Ingeniería de Comunicaciones, Universidad de Cantabria, Santander, Spain, where he is currently an Associate Professor. His research interests include digital signal processing and nonlinear and chaotic systems .

Jesús Ibáñez (S'00) was born in Santander, Spain, in 1971. He received the Radiocommunication Bachelor Engineer and Telecomm Engineer degrees from the Universidad de Cantabria, Santander, in 1992 and 1995, respectively.

In 1995, he joined the Departamento de Ingeniería de Comunicaciones, Universidad de Cantabria, where he is currently an Associate Professor. His research interests include digital signal processing, digital communication systems, and nonlinear systems.

Luis Vielva was born in Santander, Spain, in 1966. He received the Licenciado and Ph.D. degrees, both in physics, from the University of Cantabria, Santander, in 1989 and 1997, respectively.

In 1989, he joined the Departamento de Ingeniería de Comunicaciones, Universidad de Cantabria, where he is currently an Associate Professor. In 2001, he spent a visiting period at the Computational NeuroEngineering Laboratory (CNEL), University of Florida, Gainesville. He has more than 50 publications in refereed journals and international conference papers. His current research interests include blind source separation and bioinformatics.