Support vector machines framework for linear signal processing

J.L. Rojo-Álvarez[a,*], G. Camps-Valls[b], M. Martínez-Ramón[a],
E. Soria-Olivas[b], A. Navia-Vázquez[a], A.R. Figueiras-Vidal[a]

[a]*Dept. Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, Av. Universidad 30, 28911 Leganés, Madrid, Spain*
[b]*Dept. Enginyeria Electrònica, Universitat de València, Spain*

Abstract

This paper presents a support vector machines (SVM) framework to deal with linear signal processing (LSP) problems. The approach relies on three basic steps for model building: (1) identifying the suitable base of the Hilbert signal space in the model, (2) using a robust cost function, and (3) minimizing a constrained, regularized functional by means of the method of Lagrange multipliers. Recently, autoregressive moving average (ARMA) system identification and non parametric spectral analysis have been formulated under this framework. The generalized, yet simple, formulation of SVM LSP problems is particularized here for three different issues: parametric spectral estimation, stability of Infinite Impulse Response filters using the gamma structure, and complex ARMA models for communication applications. The good performance shown on these different domains suggests that other signal processing problems can be stated from this SVM framework.

Keywords: Support vector machines; Spectral estimation; Linear processing; Gamma filter; Complex ARMA

1. Introduction

Support vector machines (SVM) were originally conceived as efficient methods for pattern recognition and classification [1]. The support vector regressor (SVR) was subsequently proposed as the SVM implementation for regression and function approximation [2]. Many other data analyses have been stated from SVM principles [3–5], all of them taking advantage of the rather old ''kernel trick'' [6] to develop non-linear versions of established linear methods. However, from our viewpoint, it is more convenient first stating the problem in the linear domain, and then using the kernel trick to provide non-linearity to the model. This way, one could take advantage of the formulation to

scrutinize the statistical properties of the data in the original domain, and then, to decide which is the most suitable transformation. This will be specially adequate in time-series problems, where knowledge of the statistical properties (autocorrelation, ergodicity, or nature of interferent noise) is fundamental for their processing. In this setting, we have recently formulated the autoregressive moving average (ARMA) system identification [7], and the non-parametric spectral analysis [8] according to SVM principles. This paper concentrates in generalizing this previous work to propose an SVM framework for linear signal processing (LSP) problems. Extensions of SVM LSP formulations to the non-linear case can be easily treated by using Mercer's kernels, as usual in the SVM literature.

The SVM LSP framework consists of three basic steps for model building: (1) identifying the suitable base of the signal space, implicit in the time series model at hand; (2) using the robust cost function previously presented in [7], which allows to deal with different kinds of noise simultaneously; and (3) minimizing a constrained, regularized functional by means of the method of Lagrange multipliers. We illustrate its use in three representative signal processing areas of growing sophistication: parametric power spectral density (PSD) estimation [9], model stability of infinite impulse response (IIR) filters by using the gamma structure [10], and complex ARMA modelling [11] for communication applications.

In Section 2, the SVM framework for LSP is presented. In Section 3, each of the LSP problem is formulated as a particular case study under this framework. Section 4 provides simulation results. In Section 5, conclusions are drawn.

2. SVM framework for LSP

Let $\{y_n\}$ be a discrete time series in a Hilbert space, from which a set of N consecutive samples are measured and grouped in a vector of observations, $\mathbf{y} = [y_1, y_2, \ldots, y_N]^{\mathrm{T}}$, and let the set of vectors $\{\mathbf{z}^p\}$ be a base expanding a P-dimensional subspace into an N-dimensional Hilbert subspace, given by vectors $\mathbf{z}^p = [z_1^p, z_2^p, \ldots, z_N^p]^{\mathrm{T}}$, with

$p = 1, \ldots, P$. Each observed signal vector $\mathbf{y}$ can be represented as a linear combination of elements of this base, plus an error term $\mathbf{e} = [e_1, \ldots, e_N]^{\mathrm{T}}$ modelling errors; $\mathbf{y} = \sum_{p=1}^{P} w^p \mathbf{z}^p + \mathbf{e}$. For a given time instant n, a linear time series model can be written as

$$y_n = \sum_{p=1}^{P} w^p z_n^p + e_n = \mathbf{w}^{\mathrm{T}} \mathbf{v}_n + e_n, \qquad (1)$$

where $\mathbf{w}^{\mathrm{T}} = [w^1, \ldots, w^P]$ is the model weight vector to be estimated, and $\mathbf{v}_n^{\mathrm{T}} = [z_n^1, \ldots, z_n^P]$ represents the *input space* at time instant n. Note that the input space is closely related to the Hilbert signal space. The input vector is given at time instant n by the nth element of the Hilbert signal space vectors. For instance, in non-parametric spectral estimation, the base contains sinusoidal harmonics [8], whereas in ARMA system identification and modelling, it is generated by delayed versions of input and output signals [7].

SVM algorithms for linear classification and regression problems minimize a cost function of the residuals. In [7], the following robust cost function of the residuals was proposed,

$$L(e_n) = \begin{cases} 0, & |e_n| \leqslant \varepsilon, \\ \dfrac{1}{2\delta}(|e_n| - \varepsilon)^2, & \varepsilon \leqslant |e_n| \leqslant e_C, \\ C(|e_n| - \varepsilon) - \frac{1}{2}\delta C^2, & |e_n| \geqslant e_C, \end{cases} \qquad (2)$$

where $e_C = \varepsilon + \delta C$; ε is the insensitive parameter, and δ and C control the trade-off between the regularization and the losses. The ε-insensitive zone ignores errors lower than ε; quadratic cost zone uses the L_2-norm of errors, which is appropriate for Gaussian noise; and linear cost zone limits the effect of outliers. Note that (2) is Huber's cost when $\varepsilon = 0$ [1]. Also, using ε-insensitive cost always requires a numerical regularization by adding a small diagonal matrix to the Hessian of the QP problem, whose value can be shown to be δ parameter of (2), as detailed in [7] for SVM–ARMA system identification. Therefore, Vapnik ε-insensitive cost is considered here a particular case of (2).

The coefficients are estimated by minimizing the previous loss function which is regularized with the

2

squared norm of model coefficients, as follows:

$$\frac{1}{2}\sum_{p=1}^{P}(w^p)^2 + \frac{1}{2\delta}\sum_{n\in I_1}(\xi_n^2 + \xi_n^{\star 2}) + C\sum_{n\in I_2}(\xi_n + \xi_n^\star)$$
$$-\sum_{n\in I_2}\frac{\delta C^2}{2} \tag{3}$$

with respect to w^p, $\{\xi_n^{(\star)}\}$ (notation for both $\{\xi_n\}$ and $\{\xi_n^\star\}$), constrained to

$$y_n - \mathbf{w}^T\mathbf{v}_n \leqslant \varepsilon + \xi_n, \tag{4}$$

$$-y_n + \mathbf{w}^T\mathbf{v}_n \leqslant \varepsilon + \xi_n^\star, \tag{5}$$

$$\xi_n, \quad \xi_n^\star \geqslant 0 \tag{6}$$

for $n = n_0,\ldots,N$, where n_0 is determined by the initial conditions required; $\{\xi_n^{(\star)}\}$ are *slack variables* or *losses*, which are introduced to handle the residuals according to the robust cost function; and I_1, I_2 are the sets of samples for which losses are required to have a quadratic or a linear cost, respectively, and these sets are dynamically updated during the optimization procedure (see [7]).

Derivations of the dual functional can be found in the literature for SVR [2], for linear SVM–ARMA [7], and for SVM non-parametric spectral analysis [8]. Here, by including linear constraints (4)–(6) into (3), the primal–dual functional (or Lagrange functional) is obtained for the general LP problem:

$$L_{\text{PD}} = \frac{1}{2}\sum_{p=1}^{P}(w^p)^2 + \frac{1}{2\delta}\sum_{n\in I_1}(\xi_n^2 + \xi_n^{\star 2})$$
$$+ C\sum_{n\in I_2}(\xi_n + \xi_n^\star) - \sum_{n\in I_2}\frac{\delta C^2}{2}$$
$$- \sum_n(\beta_n\xi_n + \beta_n^\star\xi_n^\star)$$
$$+ \sum_n(\alpha_n - \alpha_n^\star)(y_n - \mathbf{w}^T\mathbf{v}_n - \varepsilon - \xi_n) \tag{7}$$

constrained to $\alpha_n^{(\star)}, \beta_n^{(\star)}, \xi_n^{(\star)} \geqslant 0$. By making zero the gradient of L_{PD} with respect to the primal variables [7], we obtain $\alpha_n^{(\star)} = (1/\delta)\xi_n^{(\star)}$ ($n \in I_1$), $\alpha_n^{(\star)} = C - \beta_n^{(\star)}$ ($n \in I_2$), and $w^p = \sum_n(\alpha_n - \alpha_n^\star)z_n^p$, and if these constraints are included into (7), primal variables can be removed. The correlation matrix of input space vectors can be identified, and

denoted as $\mathbf{R}_\mathbf{v}(s,t) \equiv \mathbf{v}_s^T\mathbf{v}_t$. The dual problem can now be obtained and expressed in matrix form, and it corresponds to the maximization of

$$-\tfrac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^\star)^T[\mathbf{R}_\mathbf{v} + \delta\mathbf{I}](\boldsymbol{\alpha} - \boldsymbol{\alpha}^\star) + (\boldsymbol{\alpha} - \boldsymbol{\alpha}^\star)^T\mathbf{y}$$
$$- \varepsilon\mathbf{1}^T(\boldsymbol{\alpha} + \boldsymbol{\alpha}^\star) \tag{8}$$

constrained to $C \geqslant \alpha_n^{(\star)} \geqslant 0$, where $\boldsymbol{\alpha}^{(\star)} = [\alpha_1^{(\star)},\ldots,\alpha_N^{(\star)}]^T$, and $\mathbf{1}$ denotes a column vector of ones. After obtaining Lagrange multipliers $\boldsymbol{\alpha}^{(\star)}$, the time series model for a new sample at time instant m is $y_m = f(\mathbf{v}_m) = \sum_{n=1}^{N}(\alpha_n - \alpha_n^\star)\mathbf{v}_n^T\mathbf{v}_m$, which is a function of weights in the input space.

3. Applications of SVM framework for LSP

This section illustrates the use of the framework presented in the previous section in three relevant LSP applications: parametric spectral estimation, gamma filter, and complex ARMA modelling. Once we identify the base of the Hilbert signal space $\mathbf{z}^p$ in the model and weight vector $\mathbf{w}$, we can directly state the quadratic programming (QP) problem (8).

3.1. SVM-AR parametric spectral estimation

The most used linear system model in PSD estimation is the all-pole structure. For a white noise input e_n, the filter output y_n is an AR(P), given by

$$y_n = \sum_{p=1}^{P}w^p y_{n-p} + e_n, \tag{9}$$

where w^p are the AR parameters. The base of Hilbert space is here given by P delayed versions of signal observations $\mathbf{y}$. This is, $\mathbf{z}^p = [y_{n_0-p},\ldots,y_{N-p}]^T$ for $p = 1,\ldots,P$, where $n_0 = P + 1$ accounts for the initial conditions. In our framework, the classical all-pole PSD estimator [9] becomes

$$\Phi(f) = \frac{1}{f_s}$$
$$\times \frac{\sigma_P^2}{\left|1 - \sum_{p=1}^{P}\sum_{n=1}^{N}(\alpha_n - \alpha_n^\star)y_{n-p}e^{-j2\pi pf/f_s}\right|^2}, \tag{10}$$

where f_s is the sampling frequency, and σ_P^2 is the variance of the residuals. In (10), the AR spectrum is the inverse of the spectrum of samples with noise above ε. For $\varepsilon = 0$, all the noise will be considered in the model, and overfitting will occur. As ε increases, the included noise level will decrease and generalization will improve, but if the number of support vectors is too low, the model will have little information and error will increase. For white Gaussian noise, a good trade-off in the choice of ε is the noise standard deviation.

3.2. SVM-gamma filter

An important issue in time series parametric modelling is how to ensure that the obtained model is stable. A remarkable trade-off between stability and simplicity of adaptation can be provided by the gamma filter [10], which is defined by

$$y_n = \sum_{p=1}^{P} w^p x_n^p + e_n, \qquad (11)$$

$$x_n^p = \begin{cases} x_n, & p = 1, \\ (1-\mu)x_{n\ 1}^p + \mu x_{n\ 1}^{p\ 1}, & p = 2, \dots, P, \end{cases} \qquad (12)$$

where μ is a free parameter. For $\mu = 1$, the structure reduces to Widrow's adaline, whereas for $\mu \neq 1$ it has an IIR transfer function due to the recursion in (12). The stability is trivially obtained with $0 < \mu < 1$ for a low-pass transfer function, and with $1 < \mu < 2$ for a high-pass transfer function. By taking into account initial conditions for $n_0 = \max(P+1, Q)$, the base of the Hilbert space is given now by $\mathbf{z}^p = [x_{n_0}^p, x_2^p, \dots, x_N^p]^T$, for $p = 1, \dots, P$, i.e., by the input vector signal after each gamma unit loops. Note that, for a previously fixed value of μ, the Hilbert vector space base is fully determined. For details on the formulation, see [12].

3.3. SVM ARMA complex formulation

An important field of application of LSP can be found in Digital Communications, where complex envelope of modulated signals is commonly used. SVM has been proposed for channel equalization using independent and nested, real, binary classifiers [13]. We adapt the SVM–ARMA model, successfully tested for channel equalization in [7], to multilevel, complex environments, using a regressor in the complex-plane. The cost function is applied to the error between desired and obtained outputs.

Let us consider two complex, discrete time processes $\{x_n\}$ and $\{y_n\}$, representing the input and the output, respectively, of a linear, time invariant system

$$y_n = \sum_{i=1}^{M} a^i y_{n\ i} + \sum_{j=1}^{Q} b^j x_{n\ j+1} + e_n, \qquad (13)$$

where $\{a^i\}$, $\{b^j\}$ are M and Q complex coefficients standing for the AR and the MA parts of the system, respectively, and e_n is complex residual noise. By denoting $\mathbf{z}^p = [y_{1\ p}, \dots, y_{N\ p}, x_p, \dots, x_{N\ p+1}]^T$ and $\mathbf{w} = [\mathbf{a}^T, \mathbf{b}^T]^T$, model (13) can be rewritten as (1). Therefore, Hilbert space base vectors are formed by delayed versions of both the input and output time series. We must include real and imaginary slack variables, and their corresponding constraints, in the cost function. Hence, for complex e_n we define $L(e_n) = L(\Re\{e_n\}) + L(\Im\{e_n\})$. Introducing the constraints into the primal functional by means of Lagrange multipliers $\{\alpha_{R,n}\}$, $\{\alpha_{R,n}^\star\}$, $\{\alpha_{I,n}\}$, $\{\alpha_{I,n}^\star\}$, for the real (subscript R) and imaginary (subscript I) parts, we obtain the primal–dual functional. By making zero its gradient with respect to w^p, we obtain

$$\mathbf{a} = \sum_{n=n_0}^{N} \psi_n y_{n\ 1}^*, \qquad \mathbf{b} = \sum_{n=n_0}^{N} \psi_n \mathbf{x}_n^*, \qquad (14)$$

where $n_0 = \max(P+1, Q)$, and $\psi_n = (\alpha_{R,n} - \alpha_{R,n}^\star) + j(\alpha_{I,n} - \alpha_{I,n}^\star)$. By placing (14) in the primal–dual functional and grouping terms, the dual functional consists of maximizing

$$-\tfrac{1}{2}\psi^H(\mathbf{R_v} + \delta\mathbf{I})\psi + \mathrm{Re}(\psi^H\mathbf{y})$$
$$- (\alpha_R + \alpha_R^\star + \alpha_I + \alpha_I^\star)\mathbf{1}\varepsilon \qquad (15)$$

constrained to Lagrange multipliers being non-negative. This is a real-valued QP problem, formally similar to (8). However, the joint information between the real and the imaginary components of the residuals (by means of the

Lagrange multipliers) is taken into account by the quadratic term, so that this formulation is different from considering just two separate one-dimensional problems, one for real and the other for imaginary observations. In that sense, this can be viewed as an optimal problem statement.

4. Experiments

Once the functional to be minimized is defined, free parameters must be tuned, both the general ones (C, δ, ε) and the model-specific ones (P in the case of SVM-AR; P and μ for SVM-gamma filter; P and Q for complex SVM–ARMA model). When the free parameters could not be chosen a priori, their optimal values were always obtained with cross-validation: for a training set of observations, 80% were used for building the SVM models with fixed free parameters, and 20% for independently evaluating the performance. A range of each free parameter was explored in a uniform or logarithmic grid of values, with the remaining free parameters fixed. The search was sequential, in the order above described, and repeated twice (details in Refs. [20,21]). The optimization procedure was based on a modification of the iterative re-weighted least-squares (IRWLS) algorithm used in [8].

4.1. SVM-AR parametric spectral estimation

Three examples illustrate the capabilities of the SVM-AR for parametric spectral estimation. The first two experiments examine the effect of outliers and of the number of available observations. In a third experiment, we test the SVM-AR for sinusoid detection in noise. We illustrate the convenience of choosing a good value for ε and its effect on the presence of artifacts in the estimated pseudospectrum.

(a) *Insensitivity to outliers.* Real-world measurements of a time series are often corrupted by outliers, i.e., atypical samples in the distribution, which may occur due to erroneous measurements or noisy phenomenon appearing in the tail portion of some noise distribution functions. In those

situations, the model may fit those unwanted data, thus leading to high-frequency distortions.

We used data generated as an ARMA-process $A(z^{-1})y_n = C(z^{-1})e_n$, where z^{-1} denotes the unit delay operator, $A(z^{-1})$ and $C(z^{-1})$ are polynomials in z^{-1}, and e_n is white Gaussian noise with unit variance, $N(0, 1)$. Two systems [14] were analyzed: an AR(3)-process given by $A(z^{-1}) = 1 + 0.9816z^{-1} + 0.9400z^{-2} + 0.7799z^{-3}$, $C(z^{-1}) = 1$, and a narrow-band ARMA(4,4)-process given by $A(z^{-1}) = 1 - 1.02z^{-1} + 2.0902z^{-2} - 0.9808z^{-3} + 0.9275z^{-4}$, $C(z^{-1}) = 1 + 0.48z^{-1} + 0.6876z^{-2} + 0.4476z^{-3} + 0.3538z^{-4}$. The input discrete process to these systems was a $N(0, 1)$ sequence with $L = 128$ samples length. The corresponding output signal was corrupted by an additive small variance random process $N(0, 0.1)$. In addition, a randomly placed 20% of samples were affected by impulsive noise from a zero mean (z.m.) and unit variance (u.v.) uniform distribution. These L samples were used for training the model and 1000 samples more were used for validation. For all simulations, we varied factors C and δ in the range $[10^{-5}, 10^5]$, and set $\varepsilon = 0$ since the insensitivity to outliers was expected to be reduced by the linear zone. A natural performance criterion is the integrated mean-square error (IMSE) [9], given by

$$\text{IMSE (dB)} = \frac{1}{F} \sum_{f=1}^{F} |\Phi(f) - \hat{\Phi}(f)|^2, \quad (16)$$

where F is the number of estimated frequencies in the spectrum. The experiment was repeated 100 times and the best model was selected according to the IMSE in the validation set.

Fig. 1 illustrates the effect of different power of outliers (P_o) on the accuracy of the PSD estimation. Results are worse with outlier amplitude increasing, but SVM-AR outperforms the standard methods in all situations, with an average gain of 1.5–2 dB. As the noise level increases, differences between the methods reduce, specially significant in ARMA(4,4) system.

(b) *Size of the dataset*. We studied the influence of the length of the signal on the accuracy of the PSD estimate. This is an important issue since in many real-world applications the whole data is segmented, and dedicated AR models are adjusted for each low-sized subset.
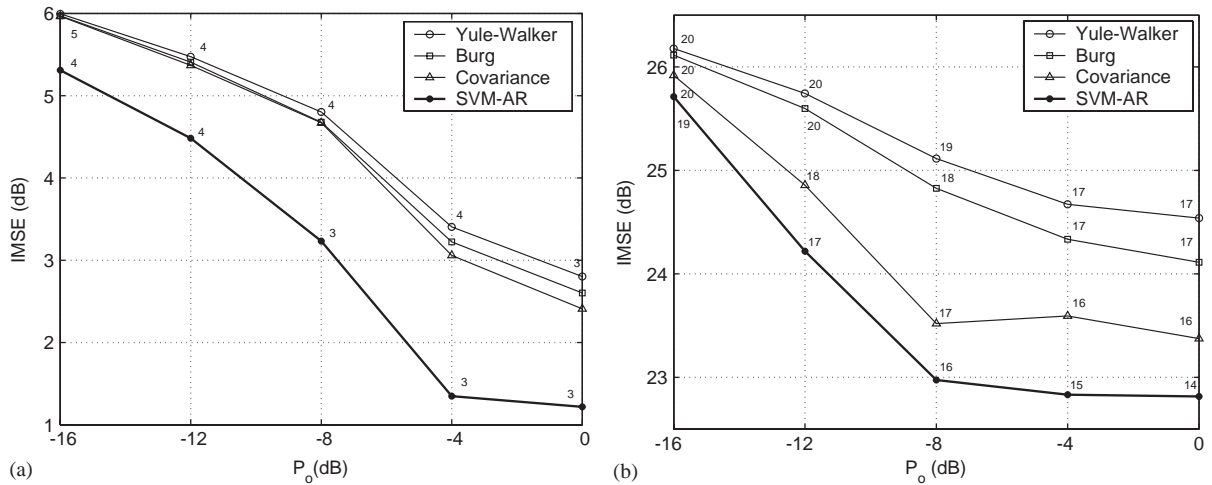
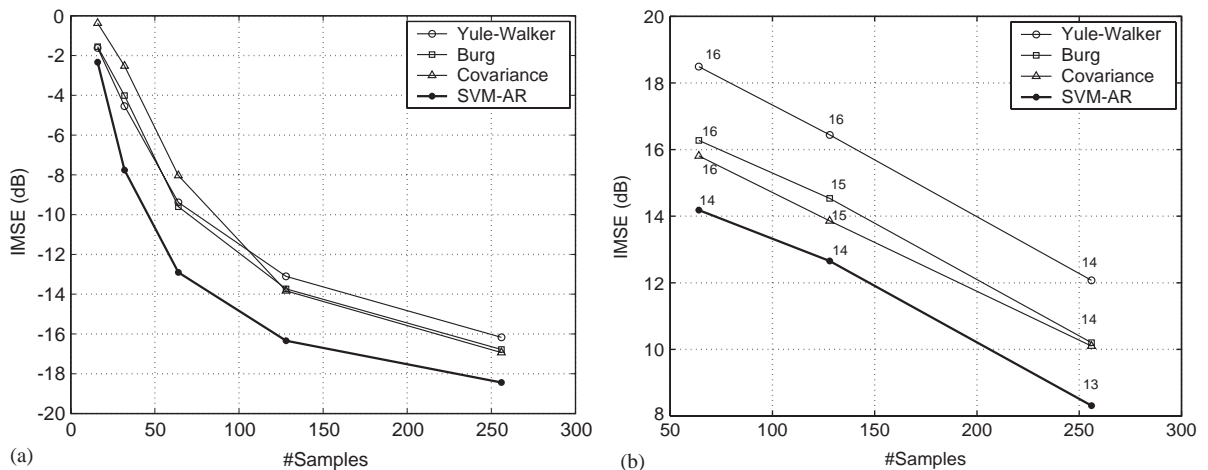Fig. 1. Evolution of IMSE for different P_o, for AR(3) (a) and ARMA(4,4) (b).

Fig. 2. Evolution of the IMSE as a function of the length of the dataset for AR(3) (a) and for ARMA(4,4) (b). Since higher orders are necessary to attain significant results for ARMA(4,4), only sequences with more than 64 samples were used.

Fig. 2 shows the results obtained by considering different lengths of the acquired signal ($L = \{16, 32, 64, 128, 256\}$ samples). No outliers were introduced in this experiment. In both cases, the SVM-AR method yields more accurate estimates of the spectrum. For AR(3) system, an important decrease of the IMSE with the number of available samples is observed, with a maximum difference between the SVM-AR and the other methods of 3.25–5 dB ($L = 64$). Similar deterioration for the IMSE is observed for ARMA(4,4) process, and

a lower difference ranging from 1.5 to 4 dB ($L = 128$).

(c) *Sinusoid detection.* The pseudospectrum of two sinusoids of unit amplitude in white Gaussian noise of power $\sigma^2 = 0.16$ was estimated by using the SVM-AR and the minimum variance (MV) methods [9]. The number of training samples was 30 and model order was $P = 10$. A value of $C = 0.1$ was set to assure smoothness of the solution, and $\delta = 10^{-4}$ was used, though good results were observed for a wide range of this parameter.
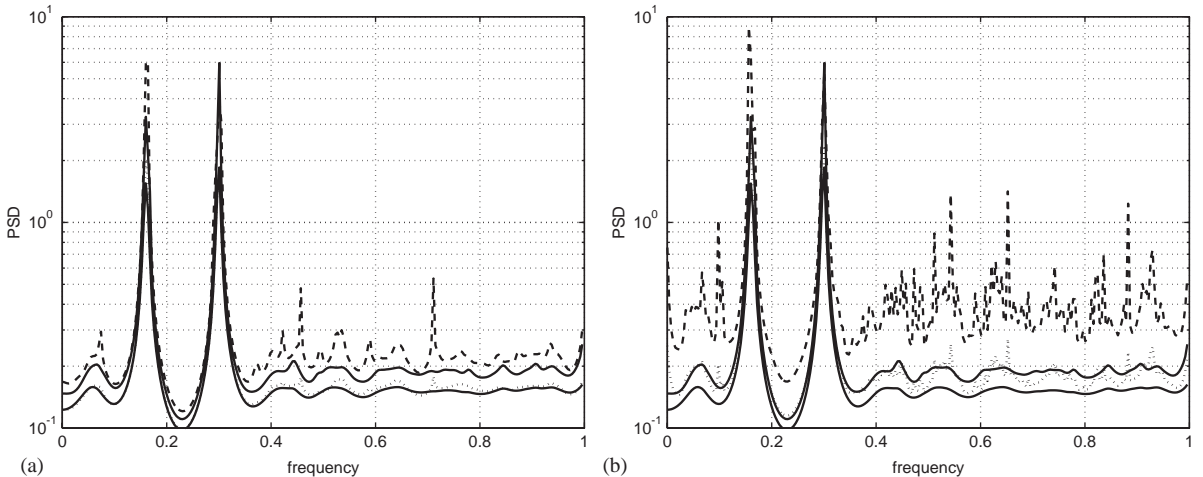
6

Fig. 3. (a) SVM pseudospectrum using $\varepsilon = \sigma$ (lower solid, mean; upper solid, variance) and $\varepsilon = 0$ (dotted, mean; dashed, variance). (b) SVM pseudospectrum using $\varepsilon = \sigma$ (lower solid, mean; upper solid, variance) and MV (dotted, mean; dashed, variance).

Fig. 3(a) shows the different performance of SVM-AR as a function of ε. Two-hundred trials with $\varepsilon = 0$ and $\varepsilon = \sigma$ were run, and mean and variance of resulting spectra were computed. Experiments with $\varepsilon = \sigma$ showed much lower variance than those with $\varepsilon = 0$, and a smaller number of artifacts. Fig. 3(b) shows the outperformance of SVM-AR when $\varepsilon = \sigma$ over MV.

4.2. SVM-gamma filter

The performance of the SVM-gamma filter is shown in terms of accuracy, robustness to outliers, memory depth, and regularization.

(a) *Insensitivity to outliers*. The system to be identified was $y_n = 1.7901 y_{n\,1} - 0.8100 y_{n\,2} + 0.0068 x_n + 0.0088 x_{n\,1} + 0.0043 x_{n\,2}$. The gamma filter with $0 < \mu < 1$ is appropriate to model this low-pass system, whose impulse response length (memory depth) is much greater than the number of coefficients. Input signal $\{x_n\}$ was a 200-sample, white, Gaussian noise sequence of z.m and u.v., and output signal $\{y_n\}$ was z.m., reduced variance $\sigma_y^2 = 0.012$, and it was corrupted by additive, white, Gaussian random process $\{z_n\}$ with z.m. and small variance $\sigma_z^2 = 0.0025$. Impulsive noise $\{j_n\}$ was generated as a sparse sequence for which 5% of the samples were randomly placed, being
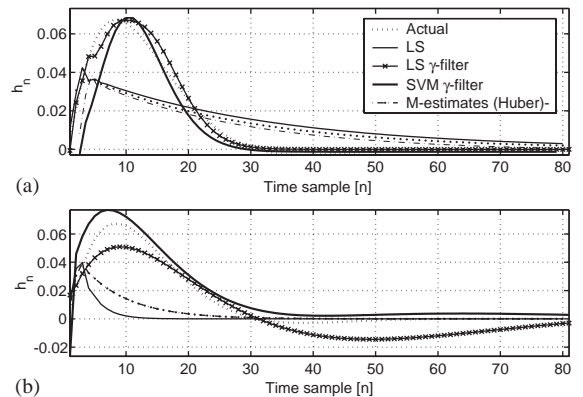

Fig. 4. Effect of outliers in the estimation of the impulse response. (a) Additive Gaussian noise and (b) additive Gaussian noise plus outliers.

Gaussian, high-amplitude samples with z.m. and u.v. ($\sigma_j^2 = 1$).

Impulse response was estimated from the observations with and without impulsive noise, $\{y_n\} + \{z_n\} + \{j_n\}$ and $\{y_n\} + \{z_n\}$, respectively. In both cases, several estimation procedures were used. The least-squares (LS) solution of an ARMA model was obtained using all the samples and the true model order, and robust M-estimate using Huber's cost [15] was obtained.

Fig. 4 shows the estimated impulse responses for the three methods. In the presence of a relatively
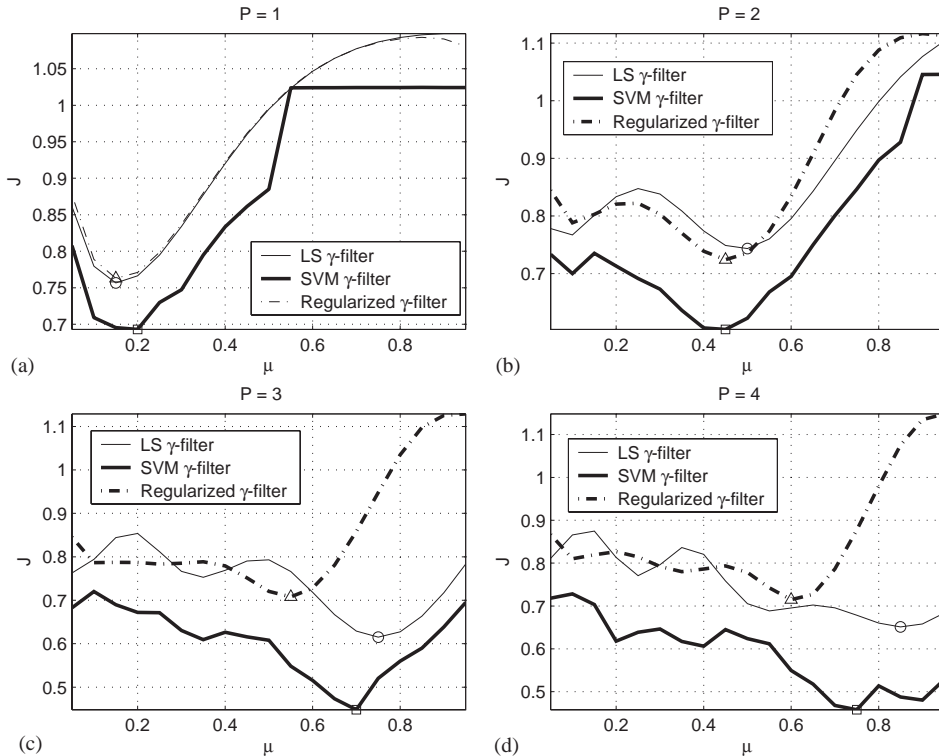
Fig. 5. Performance criterion for the identification of an elliptic filter by the LS gamma filter (thin) and the SVM gamma filter (thick) for different orders (P between 1 and 4). The optimal μ parameter is also indicated for the LS ('$\circ$'), the SVM ('$\square$') and the regularized ('$\triangle$') gamma filter.

high noise power in the output sequence, the LS solution deteriorates, while both gamma filter solutions still provide a good estimation of the system. In the presence of outliers, the robustness of the SVM allows us to obtain a more accurate system estimation, avoiding the oscillations present in the LS gamma filter. The M-estimates solution is not capable of totally compensating the effect of outliers. We obtained an optimum at $\mu = 0.2$ and $P = 7$, and thus, the memory depth of the estimated model was $M = P/\mu = 35$ samples, which contains most of the power of the true impulse response. An interesting result is that the product $\delta C = 0.78$ is near $2\sigma_y$, which is automatically found by the cross-validation in order to limit the effect of outliers in the model.

(b) *Memory depth and regularization*. In this experiment, we focused on the main advantages of both the gamma filter structure (stability and memory depth) and the SVM (regularization). We

compared the memory parameter μ in the LS, the SVM and the regularized gamma filters. The last one used the *leaky* LS algorithm described in [16]. We identified the third-order elliptic low pass filter given by $C(z^{-1}) = 0.0563 - 0.0009z^{-1} - 0.0009z^{-2} + 0.0563z^{-3}$ and $A(z^{-1}) = 1 - 2.1291z^{-1} + 1.7834z^{-2} - 0.5435z^{-3}$, which was previously analyzed in [10] because of its longer impulse response. A 100-sample input discrete process $\{x_n\}$ is a white, Gaussian noise sequence with z.m. and u.v. The output signal $\{y_n\}$ was corrupted by additive, small variance ($\sigma_e^2 = 0.1$) random process. An independent set of 100 samples was used for testing. The experiment was repeated 100 times.

Fig. 5 shows the chosen performance criterion ($J_{\min} = \text{var}(e_n)/\text{var}(y_n)$) [10] as a function of μ and P in the test set. It is noteworthy that, in all cases, the adaline structure ($\mu = 1$) performs worse than the gamma structures. In addition, the SVM-gamma filter clearly improves the results of
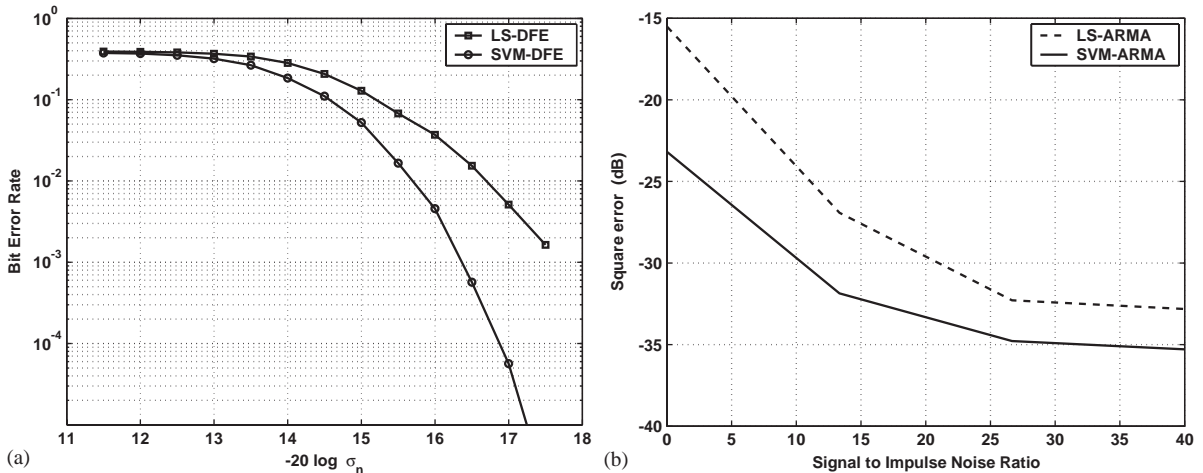
8

Fig. 6. Complex SVM ARMA experiments. (a) BER comparison between both equalizers, Least Squares DFE and complex SVM ARMA model acting as DFE (SVM DFE). (b) Squared error between ARMA truncated impulse response and the true channel, for complex LS ARMA and SVM ARMA approaches.

the LS and the regularized versions in terms of $J_{\min}$. The memory depth M for a fixed P increases with lower values of μ. This trend (observed for the three methods) is especially significant for the regularized gamma filter, but it occurs at the expense of poor performance of the criterion. The SVM-gamma filter still presents a better trade-off between memory depth and performance.

4.3. Complex SVM ARMA

Two simulations are provided to compare the complex SVM–ARMA model to standard equalizers and channel estimators.

(a) *Decision feedback equalizer*. DFE [17] is a suboptimal but low computational burden solution to reduce the impairments produced by the dispersive impulse responses of most communication channels. In a DFE, a linear feedforward filter is used to maximize the signal to noise ratio (SNR), and the feedback filter uses previous decisions to estimate and substract any trailing inter-symbols interference (ISI) caused by previous symbols.[1]

---

[1]DFE has non linear properties due to the threshold decisor. However, under the assumption that all signal values previously detected by the decisor are correct which is exact during the supervised learning phase its behavior is linear.

In this experiment, we tested a 5th order DFE trained using the complex SVM–ARMA algorithm in a channel with impulse response transfer function $H(z) = (0.98 - 0.1j) + (-0.27 - 1.38j)z^{-1} + (-0.55 - 0.73j)z^{-2}$ and 64-QAM transmission. Channel output is corrupted with Gaussian noise. The training burst consisted of 50 random symbols and we chose $\delta = 10^{-3}$, $\delta C = \sigma_n$, σ_n^2 being the thermal noise power, which was assumed to be known. We compared the SVM-DFE to the LS DFE one. Equalizers were trained with the training burst and then BER was measured with 10^5 set of independent samples.

Fig. 6 shows the result of averaging 1000 trials of the experiment. It can be observed how the SVM-DFE shows improved generalization capabilities with respect to the LS-DFE: in the hypothesis of small training data set the gain of SVM-DFE over LS-DFE exceeds 1.5 dB for a BER of 10^{-3}.

(b) *Channel estimation*. Channel estimation is a needed subprocess to optimally detect signals in dispersive channels through the Viterbi algorithm. In channel estimation, an IIR filter is commonly used, as the number of parameters of this kind of filters is much lower than those of FIR [11]. Once the model is adjusted, a truncated FIR response from those of IIR filter is used as an estimation.

9

In this experiment, we tested the robustness of complex SVM–ARMA against impulsive noise in channel estimation. We trained an ARMA(3,3) model to approximate the channel impulse response transfer function given by $H(z) = (1 + 0.5j) + (1.6 + 0.5j)z^{-1} + (0.78 - 0.05j)z^{-2} + (0.56 + 0.1j)z^{-3} + (0.37 + 0.02j)z^{-4} + (0.24 + 0.01j)z^{-5} + (0.15 + 0.01j)z^{-6} + (0.1 + 0.01j)z^{-7} + (0.06 + 0.003j)z^{-8} + (0.04 + 0.001j)z^{-9}$. This channel is a random realization of the typical urban (TU) 3GPP standard test channel [18]. The synthesis method is described in [19], and we truncated the impulse response to order 9. In order to train the model, 50 binary data were used and Signal to Thermal Noise Ratio was fixed to 30 dB. Channel impulse noise consisted of impulses appearing with a probability of 0.2. Experiments were done with an impulse variance from 0 to -40 dB under signal power. We compared the complex LS-ARMA and SVM–ARMA models. In order to adjust C and δ, we first chose $C = 10$ and swept from $\delta = 10^{-5}$ to $\delta = 10^{-3}$ with 10 exponentially distributed values. The optimal δ value was chosen to minimize the output error power. Such an adjustment becomes strictly necessary due to the presence of impulse noise. Then, the value of δC needs to be adjusted to the best possible compromise between L_2 cost function for Gaussian noise and L_1 for impulse noise.

Fig. 6 shows the squared error between the truncated impulse response of the ARMA and the true channel for the complex LS-ARMA and SVM–ARMA algorithms. It is observed that an improvement of up to 8 dB in the estimation error of the truncated impulse is obtained.

5. Conclusions

A generalized SVM framework to formulate digital LSP problems has been proposed. The framework has been applied in three relevant signal processing domains with simple considerations. All these applications have been shown to take advantage of the SVM approach. In particular, the SVM-AR method for PSD estimation provides more robust and stable results than standard approximations; the SVM-gamma filter allows a good trade-off between memory depth and performance; and the complex SVM–ARMA formulation yields significant gain over traditional equalizers and channel estimation schemes. Results suggest that other signal processing problems could be successfully solved under this SVM framework for LSP, and a wide field is open for the formulation and application of other classical signal processing tools.

References

[1] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 1995.

[2] A.J. Smola, B. Scholkopf, A tutorial on support vector regression, Statist. Comput. 14 (2004) 199 222.

[3] A. Ben Hur, D. Horn, H. Siegelmann, V. Vapnik, Support vector clustering, J. Mach. Learn. Res. 2 (2001) 125 137.

[4] F.R. Bach, M.I. Jordan, Kernel independent component analysis, J. Mach. Learn. Res. 3 (2002) 1 48.

[5] A. Gretton, R. Herbrich, A. Smola, The Kernel Mutual Information, Technical Report, Max Planck Institute for Biological Cybernetics, April 2003.

[6] A. Aizerman, E.M. Braverman, L.I. Rozoner, Theoretical foundations of the potential function method in pattern recognition learning, Autom. Remote Control 25 (1964) 821 837.

[7] J.L. Rojo Álvarez, M. Martínez Ramón, M. dePrado Cumplido, A. Artés Rodríguez, A.R. Figueiras Vidal, Support vector method for robust ARMA system identi fication, IEEE Trans. Signal Process. 52 (1) (2004) 155 164.

[8] J.L. Rojo Álvarez, M. Martínez Ramón, A.R. Figueiras Vidal, A. García Armada, A. Artés Rodríguez, A robust support vector algorithm for non parametric spectral analysis, IEEE Signal Process. Lett. 10 (11) (2003) 320 323.

[9] S.M.J. Marple, Digital Spectral Analysis with Applica tions, Prentice Hall, Englewood Cliffs, NJ, 1987.

[10] J.C. Principe, B. deVries, P.G. deOliveira, The gamma filter a new class of adaptive IIR filters with restricted feedback, IEEE Trans. Signal Process. 41 (2) (1993) 649 656.

[11] H.V. Trees, Detection, Estimation, and Modulation Theory, Wiley, New York, 2002.

[12] G. Camps Valls, M. Martínez Ramón, J. Rojo Álvarez, E. Soria Olivas, Robust gamma filter using support vector machines, Neurocomputing J. 62 (2004) 493 499.

[13] D.J. Sebald, J.A. Bucklew, Support vector machine techniques for nonlinear equalization, IEEE Trans. Signal Process. 48 (11) (2000) 3217 3266.

[14] P. Stoica, T. Sundin, Optimally smoothed periodogram, Signal Process. 78 (1999) 253 264.

[15] A. Basu, K.K. Paliwal, Robust M estimates and general ized M estimates for autoregressive parameter estimation,

in: Fourth IEEE Reg. 10th International Conference TENCON'89, Bombay, India, 1989, pp. 355 358.

[16] J.G. Harris, J. K. Juan, J.C. Principe, Analog hardware implementation of continuous time adaptive filter struc tures, J. Analog Integ. Circ. Signal Process. 18 (2) (1997) 227 229.

[17] J.G. Proakis, Digital Communications, fourth ed., McGraw Hill, New York, 2000.

[18] T.S.G. Working Group 4, Radio Access Network, UTRA (UE) TTD Radio Transmission and Reception. TS 25.102 v3.0.0, Technical Report, 3rd Generation Partnership Project 3GPP, 1999.

[19] M.C. Jeruchim, P. Balaban, K.S. Shanmugan, Simulation of Communication Channels, Plenum Press, New York, 1994.

[20] G. Camps Valls, J.L. Rojo Álvarez, M. Martínez Ramón, E. Soria Olivas, Robust γ filter using the support vector method, Technical Report No. TR DIE TSC 07/07/2003, Dept. Enginyeria Electrónica and Dept. Teoría de la Señal y las Comunicaciones, Universitat de València and Universidad Carlos III de Madrid, Spain. Available at http://www.uv.es/~gcamps.

[21] G. Camps Valls, J.L. Rojo Álvarez, M. Martínez Ramón, E. Soria Olivas, Improving γ filter performance with the support vector method, Technical Report No. TR DIE TSC 09/09/2003, Dept. Enginyeria Electrònica and Dept. Teoría de la Señal y las Comunicaciones, Universitat de València and Universidad Carlos III de Madrid, Spain. Avialable at http://www.uv.es/~gcamps.