



Letters

Robust γ -filter using support vector machines

G. Camps-Valls^{a,*}, M. Martínez-Ramón^b,
J.L. Rojo-Álvarez^b, E. Soria-Olivas^a

^a*Grup de Processament Digital de Senyals, Dept. Enginyeria Electrònica, Universitat de València,
C/ Dr. Moliner, 50, 46100 Burjassot, València, Spain*

^b*Dept. Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, Avda. Universidad, 30,
28911 Leganés, Madrid, Spain*

Communicated by R. Newcomb

Abstract

This Letter presents a new approach to time-series modelling using the support vector machines (SVM). Although the γ -filter can provide stability in several time-series models, the SVM is proposed here to provide robustness in the estimation of the γ -filter coefficients. Examples in chaotic time-series prediction and channel equalization show the advantages of the joint SVM γ -filter.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Support vector machines; γ -Filter; Iterated prediction; Channel equalization

1. Introduction

Support vector machines (SVM) are state-of-the-art tools for solving linear and non-linear machine learning problems [12]. In addition, several signal-processing problems have been specifically formulated from the SVM framework, such as regression [7], non-parametric spectral analysis [10], and auto-regressive moving average (ARMA) system identification [11]. The SVM allows us to control the robustness of time-series modelling when outliers are present, when few data samples

*Corresponding author. Tel.: +34-96-3160-197; fax: +34-96-3160-466.

E-mail address: gustavo.camps@uv.es (G. Camps-Valls).

are available, or when the assumed model does not accurately match the underlying system. However, a problem that can arise is that of ensuring that the obtained model is stable, a common requirement for AR time-series prediction and ARMA system identification.

A highly effective compromise between stability and simplicity of adaptation can be provided by the γ -filter, which was first proposed in [9]. The γ -filter can be regarded as a particular case of the generalized feed forward filter, an infinite impulse response (IIR) digital filter with restricted feedback architecture. The γ -structure results in a more parsimonious filter, and has been used for echo cancellation [8], time-series prediction [6], and system identification [9]. Two main advantages of the γ -filter are claimed: it provides stable models and it permits the study of the memory depth of a model.

We propose the use of SVM to incorporate robustness in the estimation of the γ -filter coefficients. The combined strategy of the SVM and the γ -filter structure is motivated by the robustness and stability of each method. The SVM γ -filter presented here uses the robust cost function previously presented in [10], which allows to deal with different kinds of noise simultaneously, and minimizes a constrained, regularized functional by means of the method of Lagrange multipliers to provide smoothness to the solution. Section 2 presents the SVM γ -filter formulation for linear parametric system identification and time-series modelling. Section 3 includes several application examples, showing the capabilities of our method. Finally, Section 4 provides some conclusions.

2. The SVM γ -filter

The standard γ -filter is defined by the following expressions:

$$y_n = \sum_{i=1}^P w_i x_n^i, \quad (1)$$

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

where y_n is the filter output signal, x_n is the filter input signal, x_n^i is the signal present at the input of the i th gamma tap, n is the time index, and μ is a free parameter (see Fig. 1). The error signal e_n is defined as the difference between the desired d_n and the output signal, $e_n = d_n - y_n$. For $\mu = 1$, this structure reduces to Widrow's adaline, whereas, for $\mu \neq 1$, it has an IIR transfer function due to the recursion in (2). In comparison to general IIR filters, the feedback structure in the γ -filter presents two complementary conditions: (a) *locality*, since the loops are kept local with respect to the taps, and (b) *globality*, since all the loops have the same loop gain $1 - \mu$. 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. A proposed

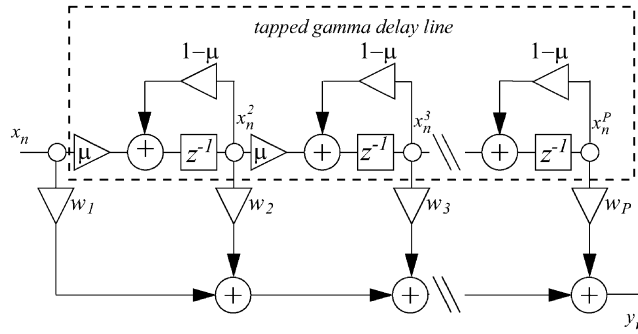


Fig. 1. The γ -filter structure.

measurement of the memory depth of a model, which allows us to quantify the past information retained, is $M = P/\mu$, and it has units of time samples [9].

The free parameters of the γ -filter (μ and $w_i, i = 1, \dots, P$) can be updated using the least mean squares (LMS) updating rules [9]. However, LMS algorithms exhibit some limitations in conditions such as small-sized data sets and the presence of outliers, which preclude the use of the γ -filter in many applications. These problems can be alleviated by using the SVM methodology and a robust cost function for estimating the filter coefficients [10]. In addition, the selection of the optimal μ parameter by means of LMS adaptation produces a hard to optimize surface containing numerous local minima. This is still an unsolved problem for the γ -filter [8,6,2], and by extension, to the SVM version. For the purpose of fair comparison, we perform an identical search of μ in all the experiments. The same procedure was followed to choose the best filter order P . Therefore, the minimization of the functional is only done with regard to weights and errors.

The SVM γ -filter algorithm minimizes the sum of two terms: the L_2 -norm of the model coefficients, and a robust cost function of the model errors, i.e.,

$$F(w_i, e_n) = \frac{1}{2} \sum_{i=1}^P w_i^2 + \sum_{n=P+1}^N L(e_n), \tag{3}$$

where

$$L(e_n) = \begin{cases} 0, & |e_n| \leq \varepsilon, \\ \frac{1}{2\delta} (|e_n| - \varepsilon)^2, & \varepsilon \leq |e_n| \leq e_C, \\ C(|e_n| - \varepsilon) - \frac{1}{2}\delta C^2, & |e_n| \geq e_C, \end{cases} \tag{4}$$

and where $e_C = \varepsilon + \delta C$, ε is the insensitive parameter, and δ and C control the trade-off between the L_2 -norm regularization of the coefficients and the losses. The proposed cost function adapts itself automatically to the noise nature in three different tracts: super-Gaussian, Gaussian, and sub-Gaussian [10]. Therefore, the estimation of the SVM γ -filter coefficients reduces to the minimization of the primal

(Lagrange) functional

$$\frac{1}{2} \sum_{i=1}^P w_i^2 + \frac{1}{2\delta} \sum_{n \in I_1} (\zeta_n^2 + \zeta_n^{*2}) + C \sum_{n \in I_2} (\zeta_n + \zeta_n^*) - \sum_{n \in I_2} \frac{\delta C^2}{2} \quad (5)$$

with respect to $\{w_i\}$ and $\{\zeta_n^{(*)}\}$,¹ constrained to

$$\begin{aligned} y_n - \sum_{i=1}^P w_i x_n^i &\leq \varepsilon + \zeta_n, \\ -y_n + \sum_{i=1}^P w_i x_n^i &\leq \varepsilon + \zeta_n^* \quad \text{and} \quad \{\zeta_n^{(*)}\} > 0, \quad n = P+1, \dots, N. \end{aligned} \quad (6)$$

Here, $\{\zeta_n^{(*)}\}$ are *slack variables* or *losses*, and are introduced to deal with committed errors. I_1, I_2 are the sets of samples for which losses are required to have a quadratic or a linear cost, respectively. Note that these sets are not static during the optimization procedure.

We can obtain the dual problem by including linear constraints (6) into (5) by means of Lagrange multipliers $\alpha_n^{(*)}$ associated to each constraint. The dual functional obtained has to be minimized with respect to primal variables (w_i and $\zeta_n^{(*)}$) and maximized with respect to dual variables ($\alpha_n^{(*)}$), which corresponds to the maximization of

$$-\frac{1}{2}(\alpha - \alpha^*)^T \mathbf{R}_x^P (\alpha - \alpha^*) - \frac{\delta}{2}(\alpha + \alpha^*)^T \mathbf{I}(\alpha + \alpha^*) + (\alpha - \alpha^*)^T \mathbf{y} - \varepsilon \mathbf{1}^T (\alpha + \alpha^*) \quad (7)$$

constrained to $0 \leq \alpha_n^{(*)} \leq C$, where $\alpha^{(*)} = [\alpha_{P+1}^{(*)}, \dots, \alpha_N^{(*)}]^T$, $\mathbf{y} = [y_{P+1}, \dots, y_N]^T$. The term \mathbf{R}_x^P is an $(N - P) \times (N - P)$ matrix whose (s, t) -element is $R_x^P(s, t) \equiv \sum_{i=1}^P x_s^i x_t^i$, with $t \geq N$ and $s \geq 1$, and denotes the time-local P th order sample estimator of the autocorrelation function of the gamma-filtered versions of the input signal. Derivations of the dual functional in similar problems can be found in [10,11]. The optimization of the dual problem (7) can be solved through quadratic programming (QP), but substantial computational advantage is obtained by using the iterative re-weighted least squares (IRWLS) procedure [10]. Finally, the SVM γ -filter is expressed as

$$y_n = f(x_n) = \sum_{s=P+1}^N (\alpha_s - \alpha_s^*) \sum_{i=1}^P x_n^i x_s^i, \quad (8)$$

where only samples with non-zero $\alpha_s^{(*)}$ count in the solution and are called *support vectors*. Moreover, note that the second summation can be conveniently expressed as a dot product involving the delayed and gamma-filtered versions of the input time-series, which allows the non-linear extension of the model in a natural way by means of Mercer kernels [3]. In this paper, nevertheless, we focus on analysing the linear SVM γ -filter.

¹Hereafter, $\{\xi_n\}$ and $\{\zeta_n^{(*)}\}$ will be denoted with $\{\xi_n^{(*)}\}$ for notation simplicity.

3. Experiments

The SVM γ -filter is benchmarked to the original γ -filter and other related methods in time-series prediction and channel equalization problems, which illustrate robustness capabilities to model mismatch and to low-sized data sets, respectively.

3.1. Iterated prediction

In this experiment, we evaluated the SVM γ -filter for one-step ahead prediction of the classical high-dimensional chaotic system generated by the Mackey–Glass differential equation $x'(t) = -0.1x(t) + (0.2x(t - t_d))/(1 + x^{10}(t - t_d))$, with delay $t_d = 17$ [5]. We considered 1000 training samples and used the next 500 for free parameter selection with cross-validation. We used an embedding dimension $d = 6$ and a step size $\tau = 6$, as proposed in the literature [1]. We compared the performance of γ -filters to the best linear model (AR) reported in [1]. The best model parameters were: AR ($P = 66$), LS γ -filter ($P = 78$, $\mu = 0.65$), and SVM γ -filter ($P = 73$, $\mu = 0.55$, $\varepsilon = 0$, $\delta C = 145.20$). Therefore, a higher memory depth (M) was obtained by our method.

Model robustness was assessed by examining iterated prediction performance, i.e., models only receive predicted values for posterior prediction from a certain time instant [5]. Iterated predictions were made from 1000 different starting points. The log root normalized MSE is shown in Fig. 2(a). The linear slope illustrates the exponential divergence of errors. After 80 iterations, the n MSE for the AR method became greater than 0, thus indicating a prediction worse than that given by just the constant mean value. The γ -filters diverged similarly, and their iterated predictions deteriorated slower than the AR model. The mean and standard deviation of the iterated log average errors demonstrated that the SVM implementation of the γ -filter performs better than the LS-based one.

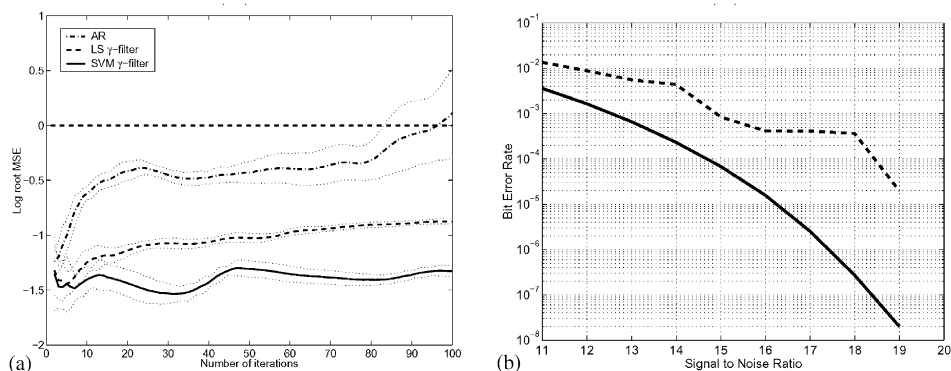


Fig. 2. Experiments. (a) Log-averaged root n MSE vs. number of iterations (solid) and standard deviations (dotted). (b) BER vs. averaged SNR for LS (dotted) and SVM (solid) γ -filter equalizers.

3.2. Channel equalization

This experiment consisted of equalizing a binary pulse amplitude modulation signal at the output of a dispersive channel, whose low-pass model was a tapped-delay line with $h_n = \delta_n + 0.6\delta_{n-1} + 0.2\delta_{n-2} - 0.1\delta_{n-3} + 0.1\delta_{n-4}$. This impulse response can represent a minimum-phase dispersive channel, which is common in suburban and hilly terrain environments [4]. A set of 128 randomly generated samples was transmitted; 64 samples were used to find the coefficients (training data set) and the remaining samples (validation set) were used to choose the value of the μ parameter. In order to measure the bit error rate (BER), an independent test burst of 10^5 samples was also used, which provides a reasonable confidence margin for the least measured BER. Gaussian noise was added. The experiment was done 1000 times for each SNR from 11 to 18 dB (1 dB steps), which represents a reasonable confidence margin for the least measured BER. We used $\delta C = 10^4$ and hence $C = 10^5$, although lower values produced similar results. In this sense, note that the linear section of the cost function was not used with these values. For both LS and SVM γ -filters, $P = 15$ assured that the filter order was high enough, and $\varepsilon = 0$ allowed us not to have to discard any of the training samples.

BER is depicted in Fig. 2(b). The LS γ -filter performance was poorer than the SVM-based one. The low number of samples used to train the filter generate a channel output which does not have enough information to estimate the statistical distribution of the process, and thus, was not enough to train the γ -filter with the LS criterion. The SVM γ -filter procedure appears to be a better choice for dealing with small training data sets, as experiments show a difference of at least 1.5 dB.

4. Conclusions and future work

A new approach to time series modelling and system identification using the γ -filter has been introduced. More confident models can be obtained whenever either low-sized training data sets are available or there exists disagreement between the true system and the assumed model. Further work will attempt to extend the SVM γ -filter to the non-linear case by using non-linear kernels. Details on the optimization issues, source code, and more application examples can be found at <http://gpds.uv.es/svmgamma/>.

References

- [1] M. Casdagli, S. Eubank, Nonlinear modeling and forecasting, in: Proceedings of Santa Fe Institute Studies in the Science of Complexity, vol. XII, Addison-Wesley, Reading, MA, November 1992.
- [2] T.W. Chiang, The analytical solution of gamma filter model, in: International Conference on Neural Networks, vol. 2, Houston, TX, USA, June 1997, pp. 685–689.
- [3] R. Courant, D. Hilbert, Methods of Mathematical Physics, Interscience Publications, Wiley, New York, 1953.

- [4] M.C. Jeruchim, P. Balaban, K.S. Shanmugan, *Simulation of Communication Systems*, Plenum Press, New York, 1992.
- [5] H. Kantz, T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, 1997.
- [6] J.M. Kuo, S. Celebi, J. Principe, Adaptation of memory depth in the gamma filter, in: *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP94*, vol. 4, Adelaide, Australia, April 1994, pp. 373–376.
- [7] K.-R. Müller, A. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, V. Vapnik, Predicting time series with support vector machines, in: B. Schölkopf, C.J.C. Burges, A.J. Smola (Eds.), *Advances in Kernel Methods—Support Vector Learning*, MIT Press, Cambridge, MA, 1999, pp. 243–254.
- [8] M. Palkar, J. Principe, Echo cancellation with the gamma filter, in: *Proceedings of the International Conference on Acoustics, Speech and Signal Processing, ICASSP94*, vol. 3, Adelaide, Australia, April 1994, pp. 369–372.
- [9] J.C. Principe, B. deVries, P.G. deOliveira, The gamma filter—A new class of adaptive IIR filters with restricted feedback, *IEEE Trans. Signal Proc.* 41 (2) (1993) 649–656.
- [10] 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 Proc. Lett.* 10 (11) (2003) 320–323.
- [11] J.L. Rojo-Álvarez, M. Martínez-Ramón, M. dePrado Cumplido, A. Artés-Rodríguez, A.R. Figueiras-Vidal, Support vector method for robust ARMA system identification, *IEEE Trans. Signal Proc.* 52 (1) (2004) 155–164.
- [12] V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, 1995.