

# Support Vector Method for ARMA System Identification: A Robust Cost Interpretation

José Luis Rojo-Álvarez, Manel Martínez-Ramón, Aníbal R. Figueiras-Vidal,  
Mario de Prado-Cumplido, and Antonio Artés-Rodríguez

Dept. Signal Theory and Communications, Universidad Carlos III de Madrid,  
28911 Leganés, Madrid, Spain

{j1rojo,manel,arfv,mprado,antonio}@tsc.uc3m.es

<http://www.tsc.uc3m.es>

**Abstract.** This paper deals with the application of the Support Vector Method (SVM) methodology to the Auto Regressive and Moving Average (ARMA) linear-system identification problem. The SVM-ARMA algorithm for a single-input single-output transfer function is formulated. The relationship between the SVM coefficients and the residuals, together with the embedded estimation of the autocorrelation function, are presented. Also, the effect of the numerical regularization is used to highlight the robust cost character of this approach. A clinical example is presented for qualitative comparison with the classical Least Squares (LS) methods.

## 1 Introduction

The ARMA modeling of a linear time invariant (LTI) system relating two simultaneously observed discrete-time processes is a wide framework, its applications ranging from digital communications (channel identification) to plant identification and biological signal analysis [1]. The two families of ARMA techniques, the Prediction Error Methods (such as Least Squares, LS) and the Correlation Methods, still present a number of limitations such as sensitivity to outliers, noise and order selection [2].

A robust solution can be drawn from the SVM framework [3,4,5]. Among the potential advantages of the SVM are: first, its single-minimum solution; second, it is an strongly regularized method, appropriate to ill-posed problems; and finally, it extracts the maximum information from the available samples whenever the statistical distribution is unknown. Therefore, the SVM appears as an attractive robust approach to the system identification problem.

The SVM has already been used for AR modeling in time series prediction [6], by the straightforward approach of using the embedded time series as input for the common SVM regression algorithm. Nevertheless, ARMA system identification is a troublesome framework which makes it worth to extend the regression algorithm to this particular problem structure. Moreover, little attention has been paid to the role of the numerical regularization on the cost function, and

in fact it allows a very useful statistical interpretation of the different terms of the SVM algorithm for the identification problem.

In the next section, the formulation of the SVM-ARMA algorithm is presented. In Section 3, the effect of the numerical regularization on the empirical cost function is analyzed. A clinical application example is introduced in Section 4. Finally, conclusions are drawn in Section 5.

## 2 The SVM-ARMA Formulation

A LTI system relating  $N$  observations of two discrete-time sequences,  $\{x_n\}$  and  $\{y_n\}$ , can be approximated by an ARMA system, which is completely defined by the following difference equation [2],

$$y_n = \sum_{i=1}^p a_i y_{n-i} + \sum_{j=1}^q b_j x_{n-j+1} + e_n \tag{1}$$

where  $a_n$  ( $b_n$ ) is the  $p$ -length ( $q$ -length) sequence determining the AR (MA) coefficients, and  $\{e_n\}$  stands for measurement errors. This equation is used to obtain a constraint for the output samples observed in the time-lags  $n = k_o, \dots, N$  (where  $k_o = \max(p + 1, q)$  in order to take into account initial conditions).

The SVM approach for the linear regression problem uses the  $L_2$  norm for the structural cost and the  $L_\epsilon$ -insensitive function for the empirical cost [8], the later given by

$$L_\epsilon(v) = \begin{cases} |v| - \epsilon, & \text{if } |v| \geq \epsilon, \\ 0, & \text{if } |v| < \epsilon. \end{cases} \tag{2}$$

Then, by following the usual methodology for deriving SVM algorithms [3], the SVM-ARMA problem can be stated as minimizing

$$L_P(\xi_k, \xi_k^*, a_i, b_j) = \frac{1}{2} \sum_{i=1}^p a_i^2 + \frac{1}{2} \sum_{j=1}^q b_j^2 + C \sum_{k=k_o}^N (\xi_k + \xi_k^*) \tag{3}$$

constrained to

$$y_k - \sum_{i=1}^p a_i y_{k-i} - \sum_{j=1}^q b_j x_{k-j+1} \leq \epsilon + \xi_k \tag{4}$$

$$-y_k + \sum_{i=1}^p a_i y_{k-i} + \sum_{j=1}^q b_j x_{k-j+1} \leq \epsilon + \xi_k^* \tag{5}$$

and to  $\xi_k, \xi_k^* \geq 0$ , for  $k = k_o, \dots, N$ , where  $\xi_k, \xi_k^*$  are the *slack variables* or losses. The Lagrange functional,  $L_{PD}$ , has to be minimized with respect to the primal variables  $a_i, b_i, \xi_k, \xi_k^*$  and maximized with respect to the dual variables  $\alpha_k, \alpha_k^*, \beta_k, \beta_k^*$ . By deriving  $L_{PD}$  with respect to the primal variables and equating

to zero, two main consequences are drawn. First, dual variables are shown to be constrained to an upper and a lower bound,

$$0 \leq \alpha_k, \alpha_k^* \leq C \tag{6}$$

for  $k = k_0, \dots, N$ . Second, the following conditions can be traced,

$$a_l = \sum_{k=k_0}^N (\alpha_k - \alpha_k^*) y_{k-l} \tag{7}$$

$$b_l = \sum_{k=k_0}^N (\alpha_k - \alpha_k^*) x_{k-l+1} \tag{8}$$

showing the relationship between model coefficients, dual coefficients and the observations. These conditions are introduced into the  $L_{PD}$  functional, thus maximizing the dual counterpart,

$$L_D = -\frac{1}{2} (\bar{\alpha} - \bar{\alpha}^*)^T [\mathbf{R}_x^q + \mathbf{R}_y^p] (\bar{\alpha} - \bar{\alpha}^*) + (\bar{\alpha} - \bar{\alpha}^*)^T \bar{y} - \varepsilon \bar{\mathbf{1}}^T (\bar{\alpha} + \bar{\alpha}^*) \tag{9}$$

where  $\bar{\alpha}^{(*)} = \{\alpha_{k_0}^{(*)}, \dots, \alpha_N^{(*)}\}^T$ ,  $\bar{y} = \{y_{k_0}, \dots, y_N\}^T$ , and

$$R_x^q(m, k) = \sum_{j=1}^q x_{m-j+1} x_{k-j+1} \tag{10}$$

$$R_y^p(m, k) = \sum_{j=1}^p y_{m-j} y_{k-j}. \tag{11}$$

These equalities can be seen as the local-lime  $p^{th}$  and  $q^{th}$  order estimators of the  $m, k$  lags of the data autocorrelation functions.

### 3 On the Numerical Regularization

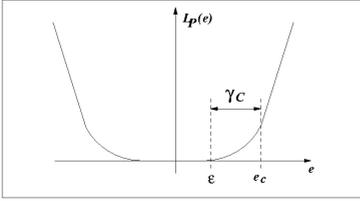
The numerical procedure used for solving (9) is the maximization of a quadratic problem of the form  $\mathbf{x}^T \mathbf{H} \mathbf{x} + \mathbf{b}^T \mathbf{x}$  with respect to  $\mathbf{x}$  and with some linear constraints. In the framework of the SV regression, the matrix  $\mathbf{H}$  is not invertible and the problem is usually *ad hoc* regularized by adding a small-element diagonal matrix, this is, substituting the square matrix by  $\mathbf{H}' = \mathbf{H} + \gamma \mathbf{I}$  (for instance, see [8] in the SVM regression problem). The numerical regularization is in fact a  $L_2$  norm regularization on the dual coefficients, and it can be included into the dual problem in 9 by an additional term, thus leading to the maximization of

$$L_D = -\frac{1}{2} (\bar{\alpha} - \bar{\alpha}^*)^T [\mathbf{R}_x^q + \mathbf{R}_y^p] (\bar{\alpha} - \bar{\alpha}^*) + (\bar{\alpha} - \bar{\alpha}^*)^T \bar{y} - \varepsilon \bar{\mathbf{1}}^T (\bar{\alpha} + \bar{\alpha}^*) - \frac{\gamma}{2} (\bar{\alpha}^T \mathbf{I} \bar{\alpha} + \bar{\alpha}^{*T} \mathbf{I} \bar{\alpha}^*) \tag{12}$$

constrained to (6). The inclusion of this term raises the following result. Let  $\bar{\alpha}^o, \bar{\alpha}^{*o}$  be the solution of the problem defined by (12). Then, the relationship

between the estimated measurement error  $e_m$  and the coefficients at the saddle point  $\alpha_m^o, \alpha_m^{o*}$  is given by

$$(\alpha_m^o - \alpha_m^{o*}) = f(e_m) = \begin{cases} -C, & -e_C \geq e_m \\ \frac{1}{\gamma}(e_m + \varepsilon), & -\varepsilon \geq e_m \geq -C \\ 0, & -\varepsilon \leq e_m \leq \varepsilon \\ \frac{1}{\gamma}(e_m - \varepsilon), & \varepsilon \leq e_m \leq e_C \\ C, & e_C \leq e_m. \end{cases} \quad (13)$$



**Fig. 1.** Robust cost function for the SVM-ARMA algorithm.

The result can be easily shown by analyzing the Karush-Kuhn-Tucker (KKT) conditions in the solution [9]. This functional relationship is a straightforward nexus between the Lagrange multipliers and the residuals.

As the numerical regularization modifies the dual formulation, it necessarily modifies the primal formulation. Specifically, the consideration of the following cost function

$$L^C(e) = \begin{cases} 0, & |e| \leq \varepsilon \\ \frac{1}{2\gamma}(|e| - \varepsilon)^2, & \varepsilon \leq |e| \leq \gamma C + \varepsilon \\ C(|e| - \varepsilon) - \frac{1}{2}\gamma C^2, & |e| \geq \gamma C + \varepsilon \end{cases} \quad (14)$$

leads to the following function to be minimized:

$$L_P^C(\xi_k, \xi_k^*, a_i, b_j) = \frac{1}{2} \sum_{i=1}^p a_i^2 + \frac{1}{2} \sum_{j=1}^q b_j^2 + \frac{1}{2\gamma} \sum_{k \in I_1} (\xi_k^2 + \xi_k^{*2}) + C \sum_{k \in I_2} (\xi_k + \xi_k^*) \quad (15)$$

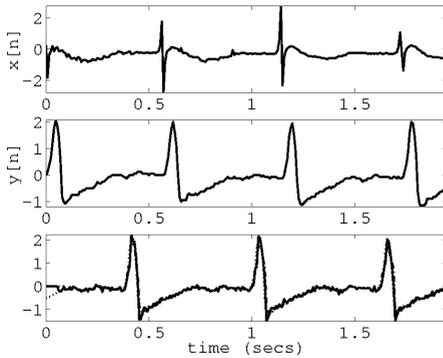
constrained to (4) and (5), where  $I_1$  is the set of samples for which  $\varepsilon \leq |\xi_k| \leq \varepsilon_C$ , and  $I_2$  is the set of samples for which  $|\xi_k| > e_C$  (Figure 1). The derivation of the dual problem leads to a functional to be maximized which is slightly different from (12), but it has the same solution.

Therefore, the primal function we are actually using is not the one suggested in (3), but rather (15). It can be seen as a robust cost function that allows different error penalties (insensitive, quadratic or linear cost) which can be straightly related to the residuals due to the analytical relationship (13), and the free parameters of the SVM can be tuned according to the statistical nature of the problem. A number of results can be derived from the signal analysis point of view, which are being currently developed.

## 4 Clinical Application Example

A kind of application where the characteristics of SVM-ARMA are appropriate and potentially useful is the digital, signal-based clinical featuring. If clinical

features are to be derived after a signal processing procedure, this method should be as robust as possible. In the present section, an example of clinical featuring is presented for qualitative comparison of the SVM-ARMA with the LS solution.



**Fig. 2.** Robust cost function for the SVM-ARMA algorithm.

implantable defibrillator in a patient during SR (Figure 2). The sample rate was 128 Hz for both channels.

**(a) The LS criterion.** Samples were divided into two non overlapping subsets for fitting the model and testing, with  $N = 250$  samples (three cardiac cycles) each. The Akaike Information Criterion was used to select the best model order, which was ARMA(1,8) [2](Model 1).

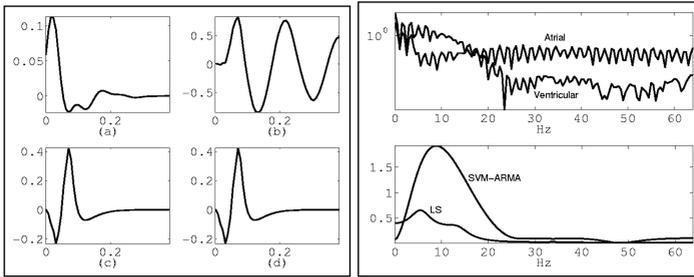
**(b) The SVM-ARMA algorithm.** The following procedure was followed for estimating the free parameters. First,  $\gamma = 1$  was fixed. An initial SVM-ARMA(40,40) was previously obtained, as this order was a 25% of the cardiac cycle, and it was considered high enough to embed all the coefficients. With  $\varepsilon = 0$  and  $C = +\infty$ , the coefficients were estimated (Model 2). The wide order was shortened by determining the first coefficients containing the 95% of the coefficients energy, leading to an ARMA(2,13) (Model 3). Then, the optimal value for the free parameters was found by sweeping a range for each of them according to the prediction error on the training samples. A range of values was found proper for  $\varepsilon$  and  $C$ , and a possible set of values were  $\varepsilon = 0.2$  and  $C = 2$ .

Figure 3 shows the impulse responses obtained for the models. The estimated system in (b) exhibits oscillation, due to the extremely high number of coefficients. This situation improves after the moderation of the order in Model 3. The agreement between the SVM-ARMA(2,13) before and after parameter tuning (Model 3 and 4) is evident.

The spectral analysis of atrial and ventricular channels, together with the spectral representation of systems in Model 1 and Model 4, are shown in Figure 3. Note that: (a) the SVM spectral response provides an enhanced energy compensation in the band where it is required (4 to 18 Hz); (b) the SVM-ARMA does not distort the low-frequency components (0 to 4 Hz), while LS does.

The modeling problem is to establish an LTI system representation for the atrial and the ventricular cardiac electrical functions, from signals that are recorded in implantable defibrillators. The usual healthy cardiac rhythm is known as Sinus Rhythm (SR), and it is represented by an almost periodic cardiac activation in both atrial and ventricular channels (electrograms). If the model is shown to be robust enough, it might be used to establish indexes of cardiac alterations from the SR, although this topic is not studied here.

The digitized atrial and ventricular signals were obtained from an im-



**Fig. 3. Left:** Impulse responses: (a) Model 1; (b) Model 2; (c) Model 3; (d) Model 4. **Right:** Frequency domain representation. Up: power spectral densities of atrial and ventricular channels. Down: modulus for Model 1 and Model 4.

Therefore, a first approach to this problem shows the SVM-ARMA exhibiting higher reproducibility and robustness than classical LS estimators. Further, specific analysis is being done on the basis of these findings.

## 5 Conclusions

A first approach to the SVM-ARMA system identification has been suggested. The role of the numerical regularization has been examined leading to a robust cost functional interpretation. A clinical application example has been developed in order to test the algorithm in comparison with the LS approach. Future work includes the extension to the non-linear system identification problem by using Mercer's kernels and non-uniform sampling. Also, robust modeling of cardiovascular and hemodynamic biomedical signals is being currently developed.

## References

1. Proakis, J.G., Rader, C.M., Ling, F., Nikias, C.L.: Advanced Digital Signal Processing. Macmillan Publishing Company, NY, US, 1992.
2. Ljung, L.: System Identification. Theory for the User. Prentice Hall, NJ, US, 1987.
3. Vapnik, V.: The Nature of Statistical Learning Theory Springer-Verlag, NY, 1995.
4. Schölkopf, B., Sung, K.: Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers IEEE Trans. on Signal Proc. Vol 45, n 11. 1997.
5. Pontil, M., Verri, A.: Support Vector Machines for 3D Object Recognition. IEEE Trans. on Pattern Anal. and Mach. Intell. Vol. 20, n 6, 1998.
6. Müller, K.R., Smola, A., Rätsch, G.R., Schölkopf, B., Kohlmorgen, J., Vapnik, V.: Predicting Time Series with Support Vector Machines. In Advances in Kernel Methods. Support Vector Learning. MIT Press, MA, USA. 1999.
7. Tikhonov, A.N., Arsenen, V.Y.: Solution to Ill-Posed Problems. V.H. Winston & Sons. Washington, US, 1977.
8. Smola, A.J., Schölkopf, B.: A Tutorial on Support Vector Regression. NeuroCOLT2 NC2-TR-1998-030,1998.
9. Luenberguer, D.G.: Linear and Nonlinear Programming. Addison-Wesley Pub Co, Reading, MA, 1984