ROBUST BLIND IDENTIFICATION OF SIMO CHANNELS: A SUPPORT VECTOR REGRESSION APPROACH

Ignacio Santamaría, Javier Vía*

Dept. Ingeniería de Comunicaciones,
Universidad de Cantabria, SPAIN
e-mail:{nacho,jvia}@gtas.dicom.unican.es

César C. Gaudes

Centro Politécnico Superior,
Universidad de Zaragoza, SPAIN
e-mail:ccaballe@posta.unizar.es

ABSTRACT

In this paper a novel technique for blind identification of multichannel FIR systems is derived from the learning paradigm of support vector machines (SVMs). Specifically, blind identification is formulated as a support vector regression problem and an iterative procedure, which avoids a trivial solution, is proposed to solve it. The SVM-based approach can be viewed as a regularized version of the least squares method for blind identification. In the paper we show that minimizing the complexity of the solution, as suggested by the structural risk minimization (SRM) principle, increases the robustness of the proposed SVM-based technique to channel order overestimation as well as to poor diversity channels (i.e., when a pair of subchannels have close zeros). The performance of the method is demonstrated through some simulation examples.

1. INTRODUCTION

Blind identification of single-input multiple-output (SIMO) channels is a common problem encountered in communications, sonar and seismic signal processing. SIMO channels appear either when the signal is oversampled at the receiver or from the use of an array of antennas. It is well known that, if the input signal is informative enough and the FIR channels are co-prime, second order statistics (SOS) are sufficient for blind identification [1].

The most popular SOS-based approaches for blind identification are the subspace methods (SS) [1, 2], the least squares (LS) subchannel matching technique [3], and the linear prediction (LP) methods [4, 5]. A good review of these methods can be found in [6].

A drawback of the subspace and least squares techniques is that they require an exact knowledge of the channel order. On the other hand, LP methods are robust when the channel order is slightly overestimated (one or two taps) [5], but their performance degrades when the channel order is highly overestimated or at low SNRs [7]. Other techniques that exploit additional statistical information about the problem, for instance the constant modulus property of the input signal, have recently been proposed for blind identification robust to channel order overestimation [8].

In this paper we consider the application of support vector machines (SVMs) [9, 10] for blind identification of SIMO channels. Recently, similar ideas have been successfully applied to blind equalization of constant modulus signals [11, 12]. For the problem at hand, the good generalization capabilities of SVMs (theoretically proven by results from statistical learning theory [9, 10]), translate into an increased robustness to channel order overestimation. Furthermore, as we will show in the paper, the proposed technique still works in poor diversity conditions (which happens, for instance, when some subchannels have close zeros).

*This work was supported by MCYT (Ministerio de Ciencia y Tecnología) under grant TIC2001-0751-C04-03 and by Telefónica Móviles under grant PPT UMTS.

2. BLIND SIMO IDENTIFICATION

Without loss of generality, in this paper we focus on the one-input/two-output channel setting. If the order of the FIR channels is M, the output of the ith channel for $i = 1, 2$ is

$$x_i[n] = \sum_{k=0}^{M} h_i[k]s[n-k] + r_i[n],$$

where $s[n]$ is an arbitrary input sequence and $r_i[n]$ is a zero-mean white Gaussian noise. The objective of blind channel identification is to identify the unknown channel responses $h_i[n]$ based on the channel output only.

As we will see later, the proposed SVM-based method can be viewed as a regularized version of the LS technique proposed in [3]. Therefore, we now briefly summarize the key ideas and properties of the LS subchannel matching method. Consider the system depicted in Fig. 1, where $\hat{h}_1$ and $\hat{h}_2$ are the estimates of order M (known) of both subchannels. In a noiseless situation, and if the subchannels are co-prime, the output $y[n] = y_1[n] - y_2[n]$ is zero iff the subchannels have been identified up to a constant arbitrary factor, i.e.,

$$\begin{aligned} \hat{H}_1(z) &= cH_1(z), \\ \hat{H}_2(z) &= cH_2(z). \end{aligned}$$

Specifically, if the channel output $x_i[n]$ is available for $n = 0, \cdots, N-M-1$, the channel estimate can be obtained by solving the following set of equations

$$\underbrace{\begin{bmatrix} \mathbf{X}_2 & -\mathbf{X}_1 \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \hat{\mathbf{h}}_1 \\ \hat{\mathbf{h}}_2 \end{bmatrix}}_{\hat{\mathbf{h}}} = 0, \tag{1}$$

where

$$\mathbf{X}_i = \begin{bmatrix} x_i[M] & x_i[M-1] & \cdots & x_i[0] \\ \vdots & \ddots & \ddots & \vdots \\ x_i[N+M-1] & \cdots & \cdots & x_i[N-1] \end{bmatrix}.$$

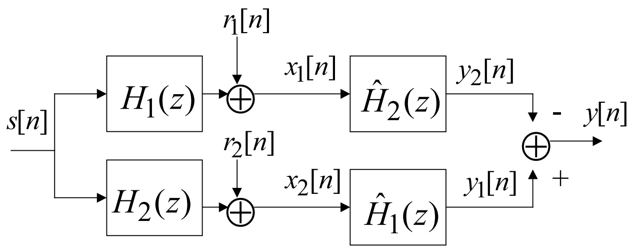

Fig. 1. Least squares blind identification of SIMO channels.

To avoid a trivial solution, the channel estimate is obtained by solving

$$\min_{||h_{LS}||=1} \mathbf{h}_{LS}^T \left(\mathbf{X}^T\mathbf{X}\right) \mathbf{h}_{LS},$$

and hence the solution is the eigenvector corresponding to the minimum eigenvalue of $\mathbf{X}^T\mathbf{X}$. In SIMO channels with p outputs, we can form $p(p-1)/2$ pairs of equations as in (1) and combine all of them into a larger linear system.

The LS method exhibits good performance but requires an accurate estimate of the channel order. This can be an unrealistic assumption in some applications; for instance in communications, where due to the transmitting and receiving shaping filters, the overall impulse response can have long tails of small amplitude terms. In the next section we propose an SVM-based alternative to mitigate this problem.

3. SVM-BASED ALGORITHM

3.1. SV regression

In this section we formulate the blind identification problem as a regression problem within the support vector machine (SVM) framework. From Fig. 1 we can write

$$\begin{aligned} \mathbf{X}_2\hat{\mathbf{h}}_1 &= \mathbf{y}_1, \\ \mathbf{X}_1\hat{\mathbf{h}}_2 &= \mathbf{y}_2; \end{aligned}$$

then, our aim is to make $y[n] = y_1[n] - y_2[n]$ as small as possible $\forall\ n$. According to the structural risk minimization (SRM) principle [9], one minimizes

$$J(\hat{\mathbf{h}}) = \frac{1}{2}\left\|\hat{\mathbf{h}}\right\|^2 + C\sum_{n=1}^{N} L\left(\xi_n, \tilde{\xi}_n\right), \quad (2)$$

subject to

$$\begin{aligned} y[n] &\leq \epsilon + \xi_n, && (3) \\ -y[n] &\leq \epsilon + \tilde{\xi}_n, && (4) \\ \xi_n, \tilde{\xi}_n &\geq 0, && (5) \end{aligned}$$

for $n = 1, \cdots, N$.

In (2) $\hat{\mathbf{h}}$ is the overall channel impulse response, $C > 0$ is a regularization parameter and $L(\xi_n, \tilde{\xi}_n)$ can be any monotonic convex function. Typically we use a linear loss function: $L = \xi_n + \tilde{\xi}_n$, or a quadratic loss function: $L = \xi_n^2 + \tilde{\xi}_n^2$. Finally, in (3) and (4) ϵ is a parameter that determines the precision of the regression: deviations over the desired output smaller than ϵ are not penalized.

The cost function (2) establishes a tradeoff between the complexity of the estimated channel impulse response (measured through its squared norm $||\hat{\mathbf{h}}||^2$) and a term that penalizes deviations over the ideal output $y[n] = 0$. The complexity term acts as a regularization term and it is the responsible for the increased robustness of the method. On the other hand, and from a Bayesian point of view, the loss function should be chosen according to the noise distribution [13]. In this paper we use the quadratic loss function $L = \xi_n^2 + \tilde{\xi}_n^2$, which seems the best choice under Gaussian noise.

The minimization of (2) under the constraints (3), (4) and (5) can be transformed into a quadratic programming (QP) problem that can be efficiently solved [14]. A direct application of this approach, however, yields the solution $\hat{\mathbf{h}} = \mathbf{0}$. In the next section we discuss an iterative procedure to avoid a trivial solution.

3.2. Iterative SV regression

The proposed technique divides the QP problem (2) into two independent regression problems (one for each subchannel)

$$J(\hat{\mathbf{h}}_i) = \frac{1}{2}\left\|\hat{\mathbf{h}}_i\right\|^2 + C\sum_{n=1}^{N} L\left(\xi_n, \tilde{\xi}_n\right), \quad (6)$$

subject to

$$\begin{aligned} y_i[n] - y_d[n] &\leq \epsilon + \xi_n, && (7) \\ y_d[n] - y_i[n] &\leq \epsilon + \tilde{\xi}_n, && (8) \\ \xi_n, \tilde{\xi}_n &\geq 0, \end{aligned}$$

for all $n = 1, \cdots, N$; and for $i = 1, 2$.

In (7) and (8) $y_d[n]$ is the desired output, which is the same for both subchannels. Obviously, $y_d[n]$ is unknown; however, both QP problems can be simultaneously solved by using the following iterative procedure: let us suppose that at iteration k we have estimates $\hat{\mathbf{h}}_{1,k}$ and $\hat{\mathbf{h}}_{2,k}$, which produce at their output $\mathbf{y}_{1,k}$ and $\mathbf{y}_{2,k}$, respectively. Then, at iteration $k+1$ a new support vector regression problem is solved for both subchannels with a common target output $\mathbf{y}_d = (\mathbf{y}_{1,k} + \mathbf{y}_{2,k})/2$, i.e., the new desired output signal is the mean value of the outputs given by the subchannel estimates at the previous iteration.

To discuss the convergence of the proposed algorithm, let us consider a quadratic loss function with $\epsilon = 0$ (this is the case used in all the simulation examples). The key point is that the proposed iterative procedure defines a nonexpansive mapping [15], i.e.,

$$||\mathbf{y}_{1,k+1} - \mathbf{y}_{2,k+1}||^2 \leq ||\mathbf{y}_{1,k} - \mathbf{y}_{2,k}||^2, \quad (9)$$

otherwise $\mathbf{y}_{1,k}$ and $\mathbf{y}_{2,k}$ would be better solutions than $\mathbf{y}_{1,k+1}$ and $\mathbf{y}_{2,k+1}$ to the QP regression problems at iteration $k+1$. Furthermore the map is strictly nonexpansive[1]; then, there is a unique fixed point that must fulfill $\mathbf{y}_1 = \mathbf{y}_2$.

To avoid a trivial solution the estimates of both subchannels are normalized after each iteration as follows

$$\begin{aligned} \hat{\mathbf{h}}_{1,k} &= \hat{\mathbf{h}}_{1,k}/\sqrt{\hat{\mathbf{h}}_{1,k}^T\hat{\mathbf{R}}_{x2}\hat{\mathbf{h}}_{1,k}} && (10) \\ \hat{\mathbf{h}}_{2,k} &= \hat{\mathbf{h}}_{2,k}/\sqrt{\hat{\mathbf{h}}_{2,k}^T\hat{\mathbf{R}}_{x1}\hat{\mathbf{h}}_{2,k}} && (11) \end{aligned}$$

[1] A mapping F is nonexpansive if $d(Fx, Fy) \leq d(x, y)$, and is strictly nonexpansive if strict inequality holds whenever $x \neq y$.

where $\hat{\mathbf{R}}_{x1}$ and $\hat{\mathbf{R}}_{x2}$ are estimates of the input correlation matrices for both subchannels (see Fig.1). In this way, we force a unit variance at the output of each subchannel.

Regarding the selection of C and ϵ, we propose to use $\epsilon = 0$ and update C according to [16]

$$C = 5\max(|\overline{y}_1| + 3\sigma_{y_1}, |\overline{y}_2| + 3\sigma_{y_2}), \qquad (12)$$

where $\overline{y}_i$ and σ_{y_i} denote the mean value and standard deviation of y_i, respectively. Finally, the iterative SV regression procedure can be summarized as follows:

Initialize $\epsilon = 0$ and $\hat{\mathbf{h}}_1 = \hat{\mathbf{h}}_2 = \delta[n-d]$.
while Convergence criterion not true **do**
  Obtain the current outputs: $\mathbf{X}_2\hat{\mathbf{h}}_1 = \mathbf{y}_1$ and $\mathbf{X}_1\hat{\mathbf{h}}_2 = \mathbf{y}_2$.
  Update C according to (12).
  Solve (6) for i=1,2 using $\mathbf{y}^d = (\mathbf{y}_1 + \mathbf{y}_2)/2$ as target output.
  Normalize the channel estimates according to (10) and (11).
end while

Algorithm 1: Summary of the SVM blind identification algorithm.

4. SIMULATION RESULTS

In the first simulation we consider a raised-cosine pulse limited in $3T$ (T is the symbol period) with roll-off factor 0.1 and the following multipath channel: $h(t) = \delta(t) - 0.7\delta(t - T/4)$. The input signal type is i.i.d. 4PAM and the received data were sampled at twice the symbol rate. The algorithm's performance is measured through the normalized mean squared error (NMSE) defined as in [8]

$$NMSE = \frac{1}{||\mathbf{h}||^2} \min_{\alpha, k \geq 0} \left|\left| \alpha\hat{\mathbf{h}} - \begin{bmatrix} \mathbf{0}_{k,1} \\ \mathbf{h} \\ \mathbf{0}_{M'-M-k} \end{bmatrix} \right|\right|, \qquad (13)$$

where $M' \geq M$ is the estimated channel order.

For the blind SVM procedure we use a quadratic loss function with $\epsilon = 0$ and update the regularization parameter C according to (12). In addition, to speed up the procedure the iterations are carried out until $||\mathbf{y}_{1,k} - \mathbf{y}_{2,k}||^2 < 10^{-3}$ or a maximum number of 100 iterations is reached. The SNR for this example is 30 dB. Figure 2 shows the estimated NMSE for the proposed procedure and the LS technique [3] when the number of input symbols varies from N=5 to N=100, and the channel order is either correctly estimated $M' = M$ or highly overestimated $M' = M + 8$. We can see that the proposed blind SVM technique is robust to channel order overestimation. Also, the NMSE curves for the blind SVM tend to flatten when the number of input symbols increases: this is due to a bias in the SVM-based channel estimate. This bias (provoked by the regularization term) can be clearly seen in Figs. 3 and 4, which show the result of 50 independent trials using $N = 100$ symbols for the cases $M' = M$ and $M' = M + 8$, respectively. The increased robustness to channel order overestimation over the LS technique is evident.

In the second example we compare the performance of the proposed algorithm when the channel provides poor diversity: an issue that has not been satisfactorily solved yet. We consider the following subchannels: $H_1(z) = 0.5 - 0.6136z^{-1} + 0.3088z^{-2} - 0.065z^{-3}$, $H_2(z) = 2 - 1.52z^{-1} + 0.38z^{-2} - 0.0638z^{-3}$ and

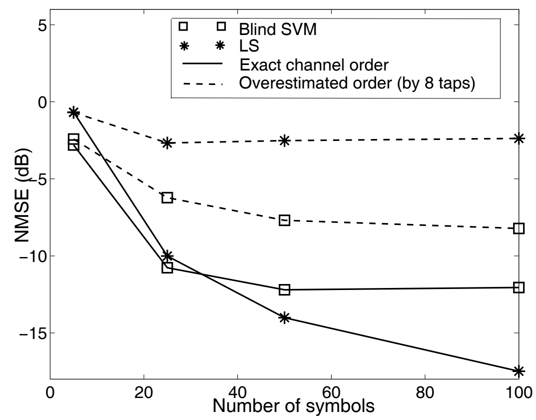

Fig. 2. Comparison between the LS and blind SVM techniques when the channel order is exact (solid line) or overestimated by 8 taps (dashed line).

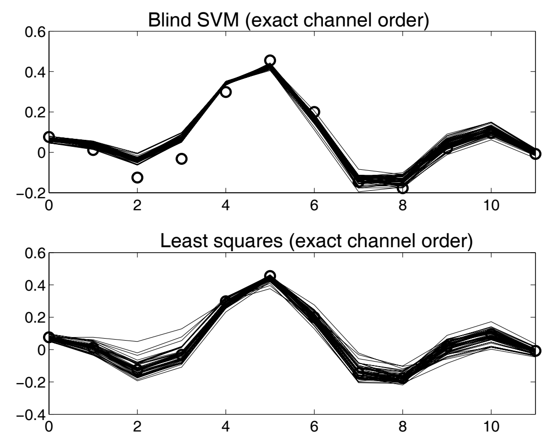

Fig. 3. 50 trials of the LS and blind SVM techniques when $M' = M$ for $N = 100$ symbols and $SNR = 30$ dB.

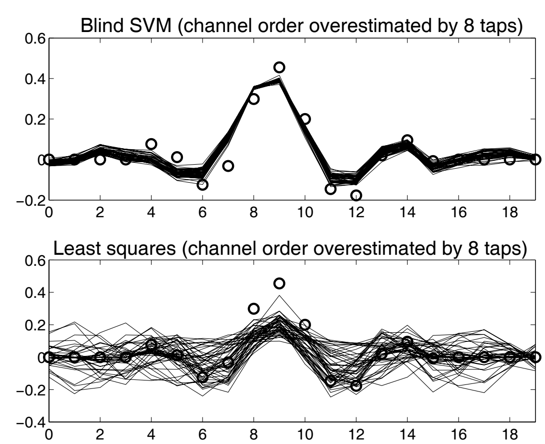

Fig. 4. 50 trials of the LS and blind SVM techniques when $M' = M + 8$ for $N = 100$ symbols and $SNR = 30$ dB.

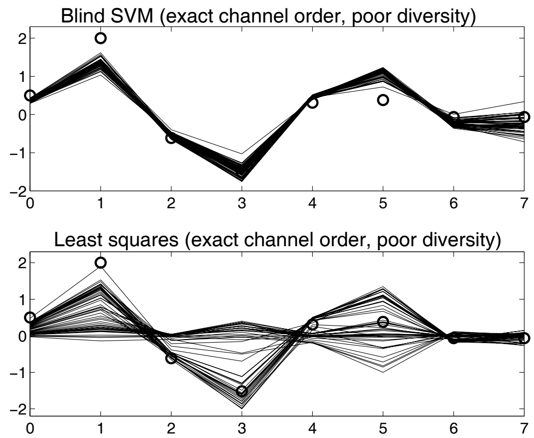

Fig. 5. 50 trials of the LS and blind SVM techniques in a situation of poor diversity.

an i.i.d. binary input signal. The first subchannel has a real zero at $z = 0.52$, whereas the second has a zero at $z = 0.51$: this pair of close zeros provoke a badly conditioned input correlation matrix that impairs subspace-based techniques. We consider a SNR= 20 dB and $N = 100$ input symbols. In Fig. 5 we see that the proposed technique again outperforms the LS subchannel matching algorithm. Obviously, the price to be paid for this robustness to channel order overestimation and badly conditioned channels is a notable increase in computational cost. Specifically, the LS technique computes an eigenvalue decomposition of an $2(M+1)\times 2(M+1)$ autocorrelation matrix, whereas the proposed technique solves at each iteration a QP problem of size N. Therefore, in its current form the use of the proposed technique is only advisable for small datasets (e.g. $N \leq 100$ symbols).

5. CONCLUSIONS

In this paper we have developed a new SVM-based algorithm for blind identification of SIMO channels. According to the SRM principle the proposed cost function establishes a tradeoff between a loss function and the complexity of the solution: this regularized functional leads to a robust solution to either channel order overestimation or badly conditioned channels (two problems that can appear in a realistic scenario). The method can be used with any persistently exciting input signal and can be easily generalized to an arbitrary number of FIR channels. Due to the high computational cost of the proposed iterative algorithm, the use of the SVM-based algorithm is advisable in applications when only a small number of data samples is available ($N \leq 100$ symbols). Nevertheless, some techniques to reduce the computational burden of QP problems that have recently appeared [17], could be used to alleviate this problem.

6. REFERENCES

[1] L. Tong, G. Xu, T. Kailath, “Blind identification and equalization based on second-order statistics: A time domain approach”, *IEEE Trans. Inform. Theory*, vol. 40, pp. 340-349, Mar. 1994.

[2] E. Moulines, P. Duhamel, J.F. Cardoso, S. Mayrargue, “Subspace methods for the blind identification of multichannel FIR filters”, *IEEE Trans. Signal Processing*, vol. 43, pp. 516-525, Feb. 1995.

[3] G. Xu, H. Liu, L. Tong, T. Kailath, “A least-squares approach to blind channel equalization”, *IEEE Trans. Signal Processing*, vol. 43, pp. 2982-2993, Dec. 1995.

[4] D. Slock, “Blind fractionally-spaced equalization, perfect reconstruction filter banks and multichannel linear prediction”, *Proc. ICASSP’94*, vol. IV, pp. 585-588, Adelaide, Australia, May 1994.

[5] K. Abed-Meraim, E. Moulines, P. Loubaton, “Prediction error method for second-order blind identification”, *IEEE Trans. Signal Processing*, vol. 45, pp. 694-705, Mar. 1997.

[6] Z. Ding, Y. Li, *Blind Equalization and Identification*, New York: Marcel Dekker, 2001.

[7] A. P. Liavas, P. A. Regalia, J. P. Delmas, “On the robustness of the linear prediction method for blind channel identification with respect to effective undermodeling/overmodeling”, *IEEE Trans. Signal Processing*, vol. 48, pp. 1477-1481, May 2001.

[8] A. Safavi, K. Abed-Meraim, “Blind channel identification robust to order overestimation: A constant modulus approach”, *Proc. ICASSP 2003*, vol. IV, pp. 313-316, Hong Kong, China, Apr. 2003.

[9] V. Vapnik, *The Nature of Statistical Learning Theory*, New York: Springer Verlag, 1995.

[10] B. Schölkopf, A. Smola, *Learning with kernels*, Cambridge, MA: The MIT Press, 2002.

[11] I. Santamaría, J. Ibáñez, L. Vielva, C. Pantaleón, “Blind equalization of constant modulus signals via support vector regression”, *Proc. ICASSP 2003*, vol. II, pp. 737-740, Hong Kong, China, Apr. 2003.

[12] I. Santamaría, C. Pantaleón, L. Vielva, J. Ibáñez, “Blind equalization of constant modulus signals using support vector machines”, *IEEE Trans. Signal Processing*, to be published.

[13] J. T. Kwok, I. W. Tsang, “Linear dependency between ϵ and the input noise in ϵ-support vector regression”, *IEEE Trans. Neural Networks*, vol. 14, pp. 544-553, May 2003.

[14] S. R. Gunn, *MATLAB Support Vector Machine Toolbox*, available at `http://www.isis.ecs.soton.ac.uk/isystems/kernel/`, University of Southampton, Image Speech and Intelligent Systems Research Group, UK, 1998.

[15] V. T. Tom, T. F. Quatieri, M. H. Hayes, J. H. Mclellan, “Convergence of iterative nonexpansive signal reconstruction algorithms”, *IEEE Trans. on Acoust. Speech and Signal Processing*, vol. 29, no. 5, Oct. 1981.

[16] V. Cherkassky and Y. Ma, “Selection of meta-parameters for support vector regression”, in J. R. Dorronsoro (Ed.), *Proc. of ICANN’2002*, pp. 687-693, Berlin, Springer-Verlag, 2002.

[17] J. Platt, “Fast training of support vector machines using sequential minimal optimization,” in B. Schölkopf, C. J. C. Burges, and A. J. Smola, (Eds.), *Advances in Kernel Methods: Support Vector Learning*, pp. 185-208, Cambridge, MA, The MIT Press, 1999.